PLOS Biology

RESEARCH ARTICLE

The human hippocampus supports auditory coding and multisensory integration

Omri Raccah[1,2*], **Aryan Agarwal**[1,2], **Yannan Zhu**[3], **Nicholas B. Turk-Browne**[1,2]

1 Department of Psychology, Yale University, New Haven, Connecticut, United States of America, **2** Wu Tsai Institute, Yale University, New Haven, Connecticut, United States of America, **3** Department of Biomedical Engineering, Duke University, Durham, North Carolina, United States of America

*omri.raccah@yale.edu

Check for updates

OPEN ACCESS

Citation: Raccah O, Agarwal A, Zhu Y, Turk-Browne NB (2026) The human hippocampus supports auditory coding and multisensory integration. PLoS Biol 24(11): e3004012. https://doi.org/10.1371/journal.pbio.3004012

Academic Editor: Raphael Kaplan, Universitat Jaume 1, SPAIN

Received: March 1, 2026

Accepted: September 9, 2026

Published: November 3, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pbio.3004012

Copyright: © 2026 Raccah et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution,

Abstract

The hippocampus receives convergent input from multiple sensory systems, yet in humans it has been studied almost exclusively through vision. Here we examine how the hippocampus contributes to sensory processing beyond the visual modality and to the multisensory integration of visual information with these other modalities. Participants were exposed to multiple repetitions of short naturalistic movie clips, each presented in four formats: auditory-only, visual-only, congruent audiovisual, and incongruent audiovisual (audio and video from different movies). Using high-resolution fMRI, we measured univariate activation and multivariate representations across the longitudinal axis and subfields of the human hippocampus. Whereas univariate analyses detected only visual responses across hippocampal subregions, with no activation for auditory stimuli and no benefit of congruent stimuli, multivariate analyses revealed robust representations of both auditory and visual scenes. The posterior hippocampus showed enhanced pattern similarity for congruent stimuli relative to unisensory stimuli, especially in the left hemisphere, demonstrating multisensory facilitation. The anterior hippocampus showed crossmodal decoding between auditory and visual versions of the same clip, driven by the right hemisphere, suggesting a more abstract representation. Finally, whole-brain searchlight analyses revealed parallel effects in cortical regions known to support multisensory integration. These findings advance understanding of auditory and multisensory coding in the human hippocampus, revealing specialization and lateralization along its longitudinal axis.

Introduction

The hippocampus and surrounding medial temporal lobe (MTL) cortex receive anatomical inputs from virtually all sensory systems [1]. In nonhuman animals, hippocampal neurons encode auditory, tactile, olfactory, and visual information, and represent crossmodal associations [2–4]. For example, macaque hippocampal neurons exhibit selectivity for congruent face and voice pairings of individual animals [4]. Yet, in

and reproduction in any medium, provided the original author and source are credited.

Data availability statement: The raw and preprocessed fMRI data are available on OpenNeuro (https://doi.org/10.18112/openneuro.ds008849.v1.0.0). The task code is publicly available at https://github.com/omriraccah/multimem_fMRI_expt and the analysis code at https://github.com/omriraccah/multimem_fMRI_analysis and via Zenodo at https://doi.org/10.5281/zenodo.22150287. The source data underlying the figures are available on OSF (https://doi.org/10.17605/OSF.IO/JCG8A).

Funding: O.R. was supported by an NIH Ruth L. Kirschstein Postdoctoral Individual National Research Service Award from the National Eye Institute (F32 EY035941; https://www.nei.nih.gov). N.B.T.-B. was supported by a National Institutes of Health grant (R01 MH069456; https://www.nih.gov) and the Canadian Institute for Advanced Research (https://cifar.ca). This study was supported by BrainWorks (RRID: SCR_024556) at the Center for Neurocognition and Behavior in the Wu Tsai Institute, Yale University. None of the sponsors or funders played any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: CSF, cerebrospinal fluid; DG, dentate gyrus; EC, entorhinal cortex; EPI, echo-planar imaging; GLMs, general linear models; GM, gray matter; HG, Heschl's gyrus; HRF, hemodynamic response function; LOC, lateral occipital cortex; MPI, message passing interface; MSPS, movie-specific pattern similarity; MTL, medial temporal lobe; OP, occipital pole; PHC, parahippocampal cortex; PRC, perirhinal cortex; pSTG, posterior superior temporal gyrus; ROIs, regions of interest; RSA, representational similarity analysis; RTs, response times; STS, superior temporal sulcus; TFCE, threshold-free cluster enhancement; WM, white matter.

humans, the hippocampus has been studied almost exclusively in the visual modality [5]. This leaves a fundamental gap: how does the human hippocampus represent sensory modalities beyond vision, and how does it integrate sense information across modalities? Understanding this integration is fundamental to explaining how the brain constructs coherent experiences and memories from fragmented inputs.

Multisensory learning enhances memory across species, from Drosophila to humans [6–9]. However, whether hippocampal mechanisms support this mnemonic benefit remains unknown. Determining the contribution of the hippocampus to multisensory processing requires considering its broader representational capacity. Beyond its traditional role in memory, an emerging literature suggests that the human hippocampus contributes to central aspects of visual perception [10,11], including eye movements [12,13], imagery [14,15], scene discrimination [16,17], and visual attention [18,19]. Given that the hippocampus represents complex conjunctions of visual features, it may additionally be critical for binding information across sensory modalities.

We tested this possibility for auditory and visual modalities. In line with recent approaches [11], we focused on characterizing representational coding schemes in the hippocampus, as opposed to testing a specific role in perception or memory behavior. Using high-resolution fMRI in 30 participants, we examined unisensory and multisensory processing in the whole hippocampus, along its longitudinal axis, across hemispheres, and within its subfields. Participants were exposed to multiple repetitions of short naturalistic movie clips presented in four formats: auditory-only, visual-only, congruent audiovisual, and incongruent audiovisual (audio and video from different clips). We first conducted a general linear model (GLM) analysis to estimate univariate activation in each condition across hippocampal regions of interest (ROIs). Next, we performed representational similarity analysis (RSA) to investigate multivariate representations of unimodal and multimodal stimuli in these regions. We repeated these analyses across the whole brain to juxtapose effects in hippocampal ROIs with known cortical sites supporting multisensory integration.

This approach allows us to address three questions in this study: first, how is auditory information distributed in the hippocampus? Whereas auditory processing in the hippocampus has been widely studied in other species, evidence in humans remains limited (for review, see [5]). Only a handful of human fMRI studies have directly examined auditory processing in the human hippocampus [20,21], and to our knowledge, none have mapped these representations across its anatomical subdivisions. Second, how does congruent multisensory stimulation facilitate processing in the hippocampus? Decades of research have demonstrated that congruent audiovisual stimuli enhance neural responses relative to unisensory stimuli, effects most prominent in the superior temporal sulcus (STS) and posterior superior temporal gyrus (posterior STG; e.g., [22–25]). Whether the hippocampus exhibits similar multisensory facilitation effects remains unknown. Third, what is the relationship between auditory and visual representations in the hippocampus? The anterior subdivision of the hippocampus has been shown to maintain abstract (as opposed to perceptually detailed) representations in the visual modality [26,27]. This raises the intriguing

possibility that the anterior hippocampus may support generalization across sensory modalities that enables crossmodal transfer.

Results

Behavior

Participants first viewed 10 short movie clips in their original audiovisual format in the scanner prior to data acquisition (6 s each; Fig 1A). This pre-exposure was intended to limit memory demands in the main experiment by reducing novelty and learning, consistent with prior work that isolated representational content in the hippocampus [11].

During the main experiment, participants were exposed to multiple repetitions of these movie clips across four presentation formats: auditory-only, visual-only, congruent audiovisual, and incongruent audiovisual (mismatched audio and video from different clips; Fig 1B). All 4 formats of each of the 10 movies were presented once per run (40 stimuli per run) across nine runs, enabling leave-one-run-out cross-validation analyses. For the incongruent audiovisual condition, each video was paired once with audio from each of the other movies; these re-pairings varied across runs to ensure that the visual and auditory content of each movie appeared as part of an incongruent trial once per run. A central fixation cross remained on screen throughout each 6-s trial across all conditions, superimposed over videos in the visual-only and audiovisual conditions.

To ensure attention during the movies, participants were instructed to press a button within 1 s of each movie's offset (Fig 1B). A response was scored as correct if it fell within this 1-s post-offset window, whereas presses during the movie or after the window were scored as incorrect. Performance in this offset detection task was reliable and near ceiling in every condition (auditory: mean accuracy across participants, $M = 0.94$, standard error of the mean, $SEM = 0.006$; visual: $M = 0.98$, $SEM = 0.006$; congruent: $M = 0.98$, $SEM = 0.003$; incongruent: $M = 0.99$, $SEM = 0.003$). Response times (RTs) on correct trials were fast and consistent across conditions (auditory: $M = 379$ ms, $SEM = 13$; visual: $M = 370$ ms, $SEM = 13$; congruent: $M = 337$ ms, $SEM = 11$; incongruent: $M = 334$ ms, $SEM = 10$), confirming participants' engagement throughout the experiment.

Univariate activity

To examine univariate activity, we constructed a GLM with regressors for each condition. The resulting beta estimates were extracted from each ROI and averaged across voxels per participant. We focused ROI analyses on the whole hippocampus, anterior and posterior hippocampus, the left and right hippocampus, and hippocampal subfields CA1, CA2/3, DG, and subiculum. We examined subfields given evidence for differential audiovisual selectivity across subfields in the primate hippocampus [4]. We also examined MTL cortex, comprising the parahippocampal cortex (PHC), perirhinal cortex (PRC), and entorhinal cortex (EC). We corrected for multiple comparisons separately across anterior/posterior subdivisions, hemispheres, subfields, and MTL cortex.

The whole hippocampus (Fig 2A) showed significant evoked responses to unisensory visual stimuli ($P < 0.001$) and both multisensory audiovisual conditions (congruent: $P < 0.001$; incongruent: $P < 0.001$), but not to unisensory auditory stimuli ($P = 0.55$). This pattern was consistent in the anterior hippocampus (visual: $P < 0.001$; auditory: $P = 0.89$; congruent: $P < 0.001$; incongruent: $P = 0.001$) and posterior hippocampus (visual: $P < 0.001$; auditory: $P = 0.42$; congruent: $P < 0.001$; incongruent: $P < 0.001$), as well as in the left hippocampus (visual: $P < 0.001$; auditory: $P = 0.75$; congruent: $P < 0.001$; incongruent: $P < 0.001$) and the right hippocampus (visual: $P < 0.001$; auditory: $P = 0.75$; congruent: $P < 0.001$; incongruent: $P < 0.001$; Fig 2B). All hippocampal subfields showed this same profile, with significant responses to the visual and both audiovisual conditions but no detectable response to auditory-only stimuli (S1B Fig). We found no evidence of multisensory facilitation in univariate activity: the congruent audiovisual condition did not reliably exceed the visual condition in any hippocampal ROI (all $P > 0.05$). There was a marginal trend of suppression for the incongruent condition relative to the

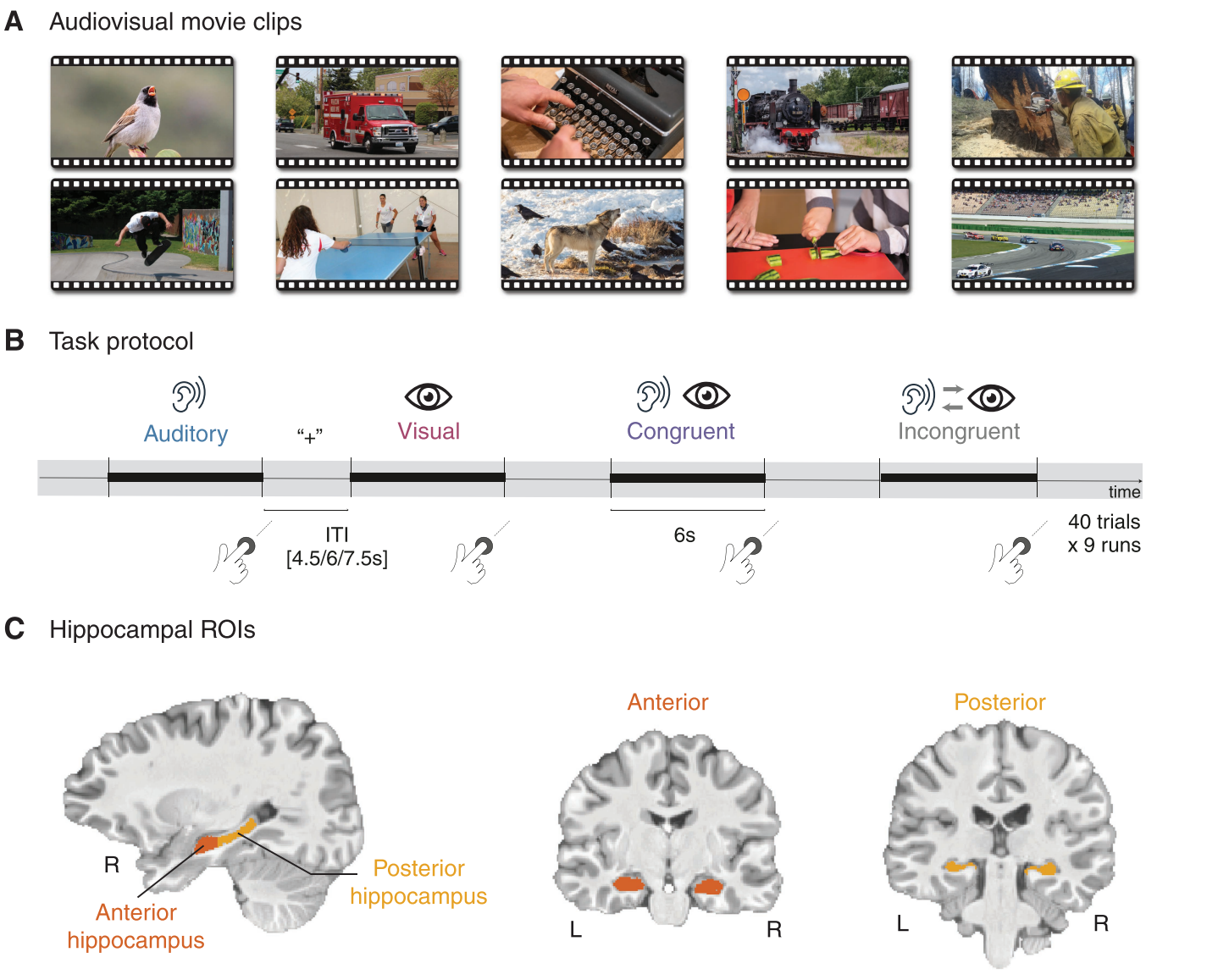


Fig 1. Experimental design and hippocampal segmentation. (A) The stimuli consisted of 10 naturalistic audiovisual movie clips (6 s each). Images are category-matched substitutes for copyright reasons. **(B)** Unisensory stimuli were created by isolating the audio or visual track of each movie, resulting in 10 auditory-only and 10 visual-only stimuli. Incongruent audiovisual stimuli were generated by pairing audio and video tracks from different movies (90 unique pairings). In each run (9 runs total), all 10 stimuli from the auditory-only, visual-only, and congruent audiovisual conditions were presented, along with 10 incongruent audiovisual stimuli (40 total trials per run). Stimulus order was randomized within each run, with jittered inter-trial intervals between trials. As a cover task to ensure attention to the stimuli, participants were instructed to press a button within 1 s of each movie's offset. **(C)** The hippocampal long axis was manually segmented into anterior and posterior subdivisions based on the last coronal slice in which the uncus was visible [28]. Example coronal slices through the anterior and posterior hippocampus.

https://doi.org/10.1371/journal.pbio.3004012.g001

congruent audiovisual condition in the whole hippocampus, its left and right hemispheres, and the anterior hippocampus (all $P < 0.1$). For full descriptive statistics and results from sensory and MTL cortical ROIs, see S1 Table and S2 Fig.

Next, we estimated group-level effects in the whole brain using voxel-wise nonparametric randomization tests (TFCE corrected at $P < 0.05$). A contrast of unisensory auditory versus visual conditions showed canonical activations in auditory and visual cortices, respectively (Fig 2C). To assess multisensory facilitation, we performed a conjunctive analysis of

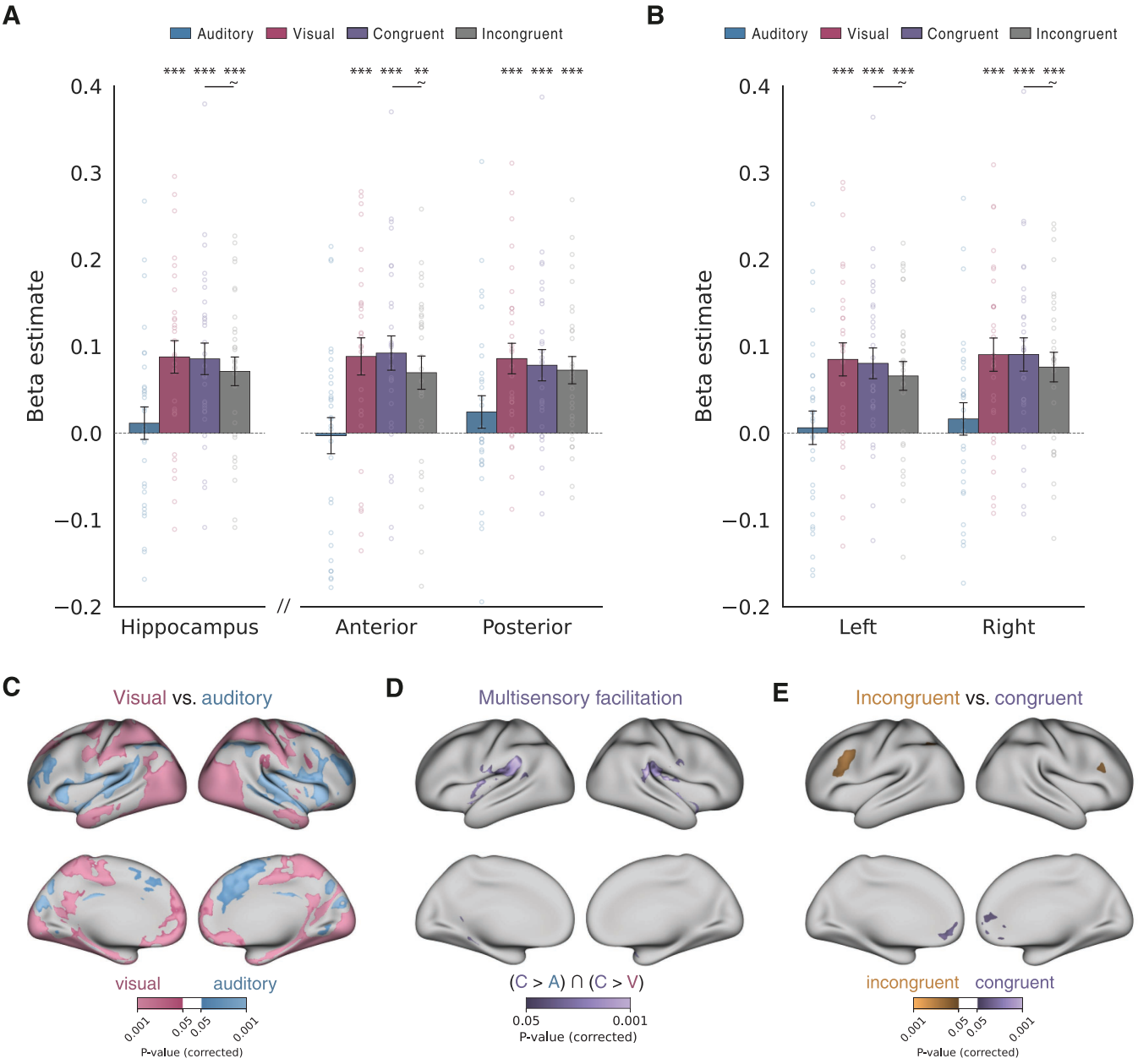


Fig 2. Univariate activation across the hippocampus and whole brain. (A) Mean beta estimates for each condition in the whole hippocampus and anterior/posterior subdivisions. **(B)** Mean beta estimates for each condition in the left and right hippocampus. **(C)** Group-level whole-brain auditory-only vs. visual-only contrast. **(D)** Conjunction analysis identifying regions in which congruent audiovisual activation exceeded both auditory-only and visual-only conditions. **(E)** Whole-brain contrast of congruent vs. incongruent audiovisual conditions. Individual participant data shown as circles. Error bars indicate SEM. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. Whole-brain maps were TFCE corrected at $P < 0.05$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.

https://doi.org/10.1371/journal.pbio.3004012.g002

voxels in which activation for the congruent audiovisual condition significantly exceeded both of the unisensory auditory and visual conditions. This contrast revealed activity in several regions, including bilateral parietal operculum, insula, angular gyrus, temporal pole, and subcortical activity in the thalamus (Fig 2D). Finally, a contrast between congruent and

incongruent audiovisual conditions revealed greater activation in the bilateral medial prefrontal cortex for congruent and in the left lateral prefrontal cortex for incongruent (Fig 2E). Together, these whole-brain effects validate the paradigm by replicating established patterns of unisensory and multisensory processing in the neocortex. The full list of peak coordinates for these contrasts is provided in S2 Table. Critically, while these univariate results confirm that the hippocampus is sensitive to the visual modality, they provide no evidence of sensitivity to the auditory modality or for multisensory facilitation.

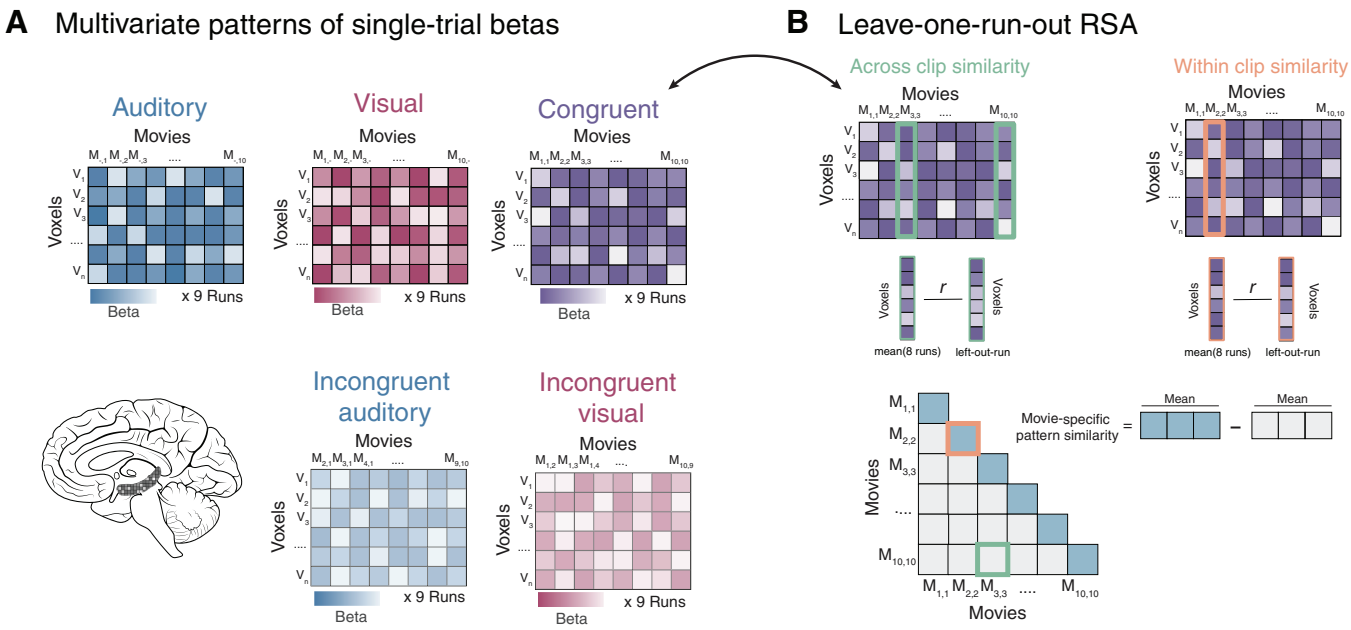
Multivariate representations

We estimated single-trial activity patterns using GLMSingle [29] and performed RSA with a leave-one-run-out cross-validation approach. For a given ROI, we extracted voxel-wise beta coefficients for each movie of a condition in every run (Fig 3A). On each fold, we correlated the patterns from one held-out run with the average patterns from the other runs (Fig 3B), yielding a 10 (movie) by 10 (movie) similarity matrix for each condition. All correlation coefficients were Fisher z-transformed and the similarity matrices were averaged across folds. Diagonal elements corresponded to within-movie similarity (same movie across runs), while off-diagonal elements reflected between-movie similarity (different movies across runs). To quantify representational distinctiveness for each condition, we computed movie-specific pattern similarity (MSPS) as the signed difference between average within-movie and between-movie similarities (mean of diagonal cells minus mean of off-diagonal cells), with positive values indicating more distinctive neural representations for individual movie clips. MSPS was computed for each participant and statistical reliability for auditory, visual, and congruent conditions was calculated against zero (null hypothesis of no diagonal minus off-diagonal; i.e., that pattern similarity was not movie-specific).

We found reliable MSPS for auditory, visual, and congruent movie clips in a subset of hippocampal ROIs (Fig 3C). Unisensory auditory representations were observed in the whole hippocampus ($P = 0.017$) and anterior hippocampus ($P = 0.007$). Unisensory visual representations were observed in the posterior hippocampus ($P = 0.031$). Congruent audiovisual representations were found in the whole hippocampus ($P = 0.001$), anterior hippocampus ($P = 0.012$), and posterior hippocampus ($P = 0.003$). The subiculum ($P = 0.031$) and DG ($P = 0.031$) showed significant congruent audiovisual MSPS but no significant unisensory MSPS (all $P > 0.05$; S1C Fig). In MTL cortex (S3B Fig), PHC showed significant MSPS for the visual ($P < 0.001$), auditory ($P = 0.002$), and congruent audiovisual ($P < 0.001$) conditions, and PRC showed significant MSPS only for the congruent audiovisual condition ($P = 0.027$); EC showed no significant effects (all $P > 0.05$).

The posterior hippocampus showed evidence of multisensory facilitation, with MSPS for the congruent condition significantly greater than for the auditory condition ($P = 0.013$) and marginally greater than for the visual condition ($P = 0.084$). The whole hippocampus showed a similar pattern, with enhanced MSPS for the congruent condition relative to both the visual ($P = 0.029$) and auditory conditions, though the latter did not reach significance ($P = 0.15$). We did not observe a facilitation pattern in the anterior hippocampus or in any hippocampal subfield (S1C Fig). Examining hemispheres separately (Fig 3D), multisensory facilitation was lateralized to the left posterior hippocampus, where congruent MSPS was significant ($P = 0.011$) and exceeded both the auditory ($P = 0.040$) and visual ($P = 0.009$) conditions. In contrast, the right posterior hippocampus showed significant visual-only ($P = 0.039$) and congruent audiovisual ($P = 0.036$) MSPS, but no facilitation. Lateralization in the whole and anterior hippocampus is shown in S4A and S4B Fig; full descriptive statistics for all ROIs are provided in S3 and S4 Tables.

Critically, congruent audiovisual MSPS could reflect a conjunctive audiovisual code, in which congruent stimuli evoke movie-specific structure not present in either modality alone, or simply a linear combination of the unisensory representations. To test this, we ran a cross-validated multiple regression, modeling the congruent audiovisual similarity matrix as a function of the auditory-only and visual-only similarity matrices (Fig 4A). Specifically, we estimated these coefficients by predicting each held-out congruent run from the remaining unisensory runs, following our main RSA approach (see Methods). Averaging across folds, we derived the auditory and visual regression coefficients, the variance explained (R^2), and the residual congruent audiovisual MSPS. The visual-only coefficient was significant in the whole hippocampus ($P = 0.003$)



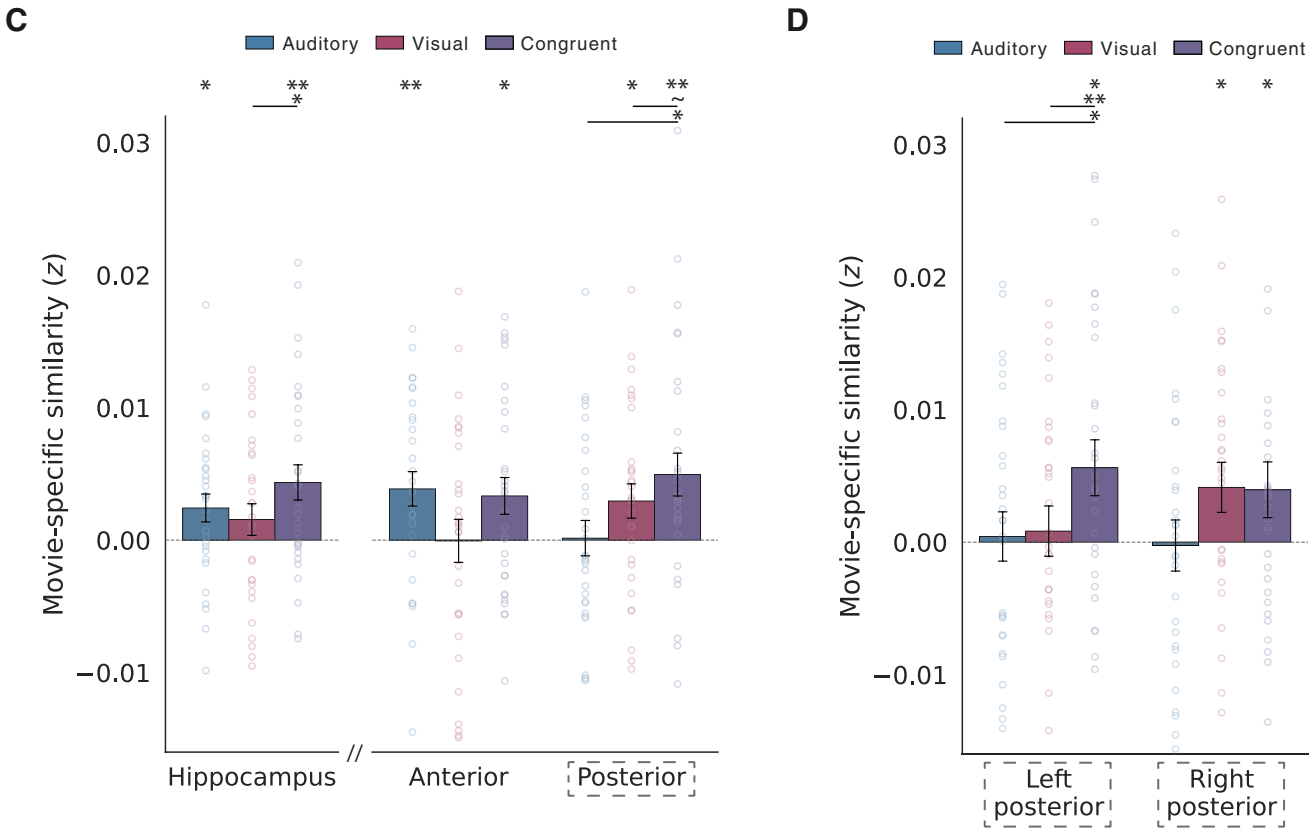


Fig 3. Movie-specific multivariate representations in the hippocampus. (A) To perform RSA, single-trial beta estimates were first extracted from a given ROI, generating multivariate activity patterns for each movie presentation. **(B)** We computed cross-validated RSA matrices independently for each condition using leave-one-run-out procedures (congruent audiovisual condition shown). Voxel-wise correlations were computed between movies (green) and within movies (orange) and averaged across cross-validation folds to construct the RSA matrix (bottom). To quantify movie-specific pattern similarity

(MSPS), we computed the signed difference between mean diagonal (within-movie) and off-diagonal (between-movie) similarity values. **(C)** MSPS across conditions for the whole hippocampus (left) and anterior and posterior subdivisions (right). The posterior hippocampus showed enhanced pattern similarity for congruent audiovisual stimuli relative to unisensory conditions (gray hatched box). **(D)** MSPS across conditions for the left and right posterior hippocampus. Six data points fall outside the plotted y-axis range; the range was chosen to match that of panel C to ease comparisons. Individual participant data shown as circles. Error bars indicate SEM. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.

https://doi.org/10.1371/journal.pbio.3004012.g003

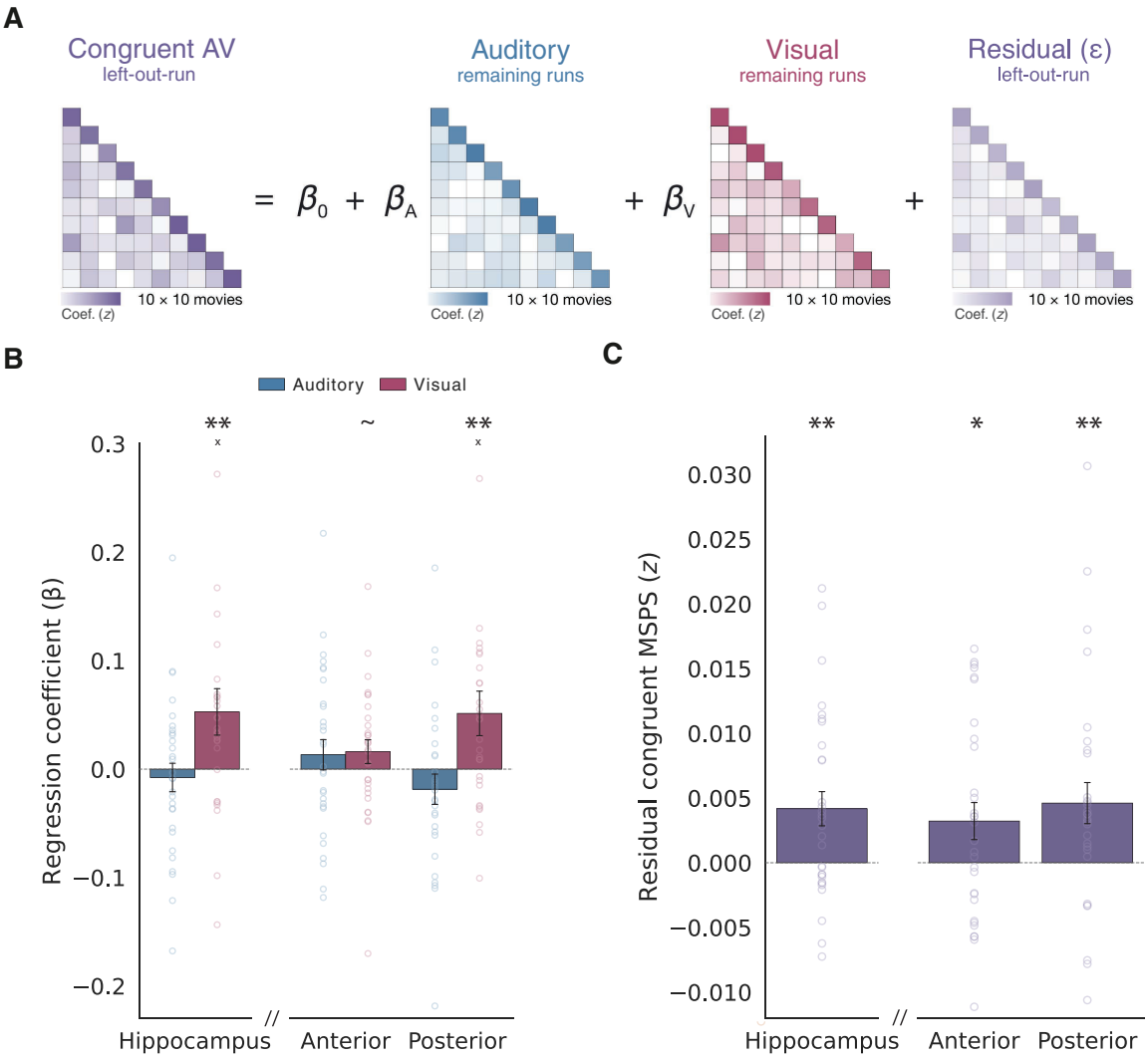


Fig 4. Conjunctive audiovisual coding in the hippocampus. (A) To test whether congruent audiovisual representations could be explained by a linear combination of unisensory representations, we performed a cross-validated multiple regression modeling the congruent audiovisual similarity matrix as a function of the auditory-only and visual-only similarity matrices. Regression coefficients (β_A, β_V) were estimated by iteratively predicting the congruent matrix on a held-out run from the remaining runs, with estimates averaged across folds. **(B)** Auditory and visual regression coefficients are shown for the whole hippocampus and anterior/posterior subdivisions. **(C)** Residual congruent audiovisual MSPS after regressing out the auditory-only and visual-only contributions. Individual participant data shown as circles; error bars indicate SEM. $\times$ markers indicate individual data points outside the plotted y-axis range. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.

https://doi.org/10.1371/journal.pbio.3004012.g004

and posterior hippocampus ($P = 0.005$; Fig 4B), whereas the auditory-only coefficient was not significant in any hippocampal ROI (all $P > 0.05$). This visual dominance was expected, as the regression models the full similarity matrix rather than isolating movie-specific structure. Critically, the unisensory matrices explained no significant variance in any hippocampal subdivision (all $P > 0.05$). This was in stark contrast to visual and auditory sensory ROIs in cortex, which accounted for substantial variance ($R^2 \approx 11\%–36\%$; for full results in sensory ROIs, see S5 Fig). Finally, the residual congruent audiovisual MSPS remained above baseline in the whole hippocampus ($P = 0.002$), anterior hippocampus ($P = 0.019$), and posterior hippocampus ($P = 0.005$; Fig 4C). Notably, residual congruent audiovisual MSPS was also significantly above zero in the sensory cortical ROIs (all $P < 0.001$; S5 Fig). Together, these findings indicate that the hippocampus carries movie-specific audiovisual structure that is not reducible to its unisensory components. For full statistical reporting, see S5 Table.

Note that this multivariate analysis could not be performed straightforwardly in the incongruent condition because each run contained a unique audiovisual combination. However, each modality still carries movie-specific information across runs; for example, an ambulance audio track is always paired with a distinct visual component within each run. This results in "incongruent visual" and "incongruent auditory" matrices, allowing us to test whether incongruence suppresses pattern similarity relative to the unisensory conditions (despite both having stable audio or video). Although we did not find any significant differences in the hippocampus (S6A Fig), we observed a significant suppression of MSPS for the incongruent relative to unisensory conditions in the PHC (S6B Fig) and at the whole-brain level (S7 Fig).

To test whether MSPS was modality-specific or amodal, we cross-correlated the voxel-wise patterns for auditory and visual presentations of the same versus different movies using the cross-validated leave-one-run-out RSA approach (Fig 5A). Reliable MSPS in this analysis would indicate the existence of movie-specific representations invariant to sensory modality. The

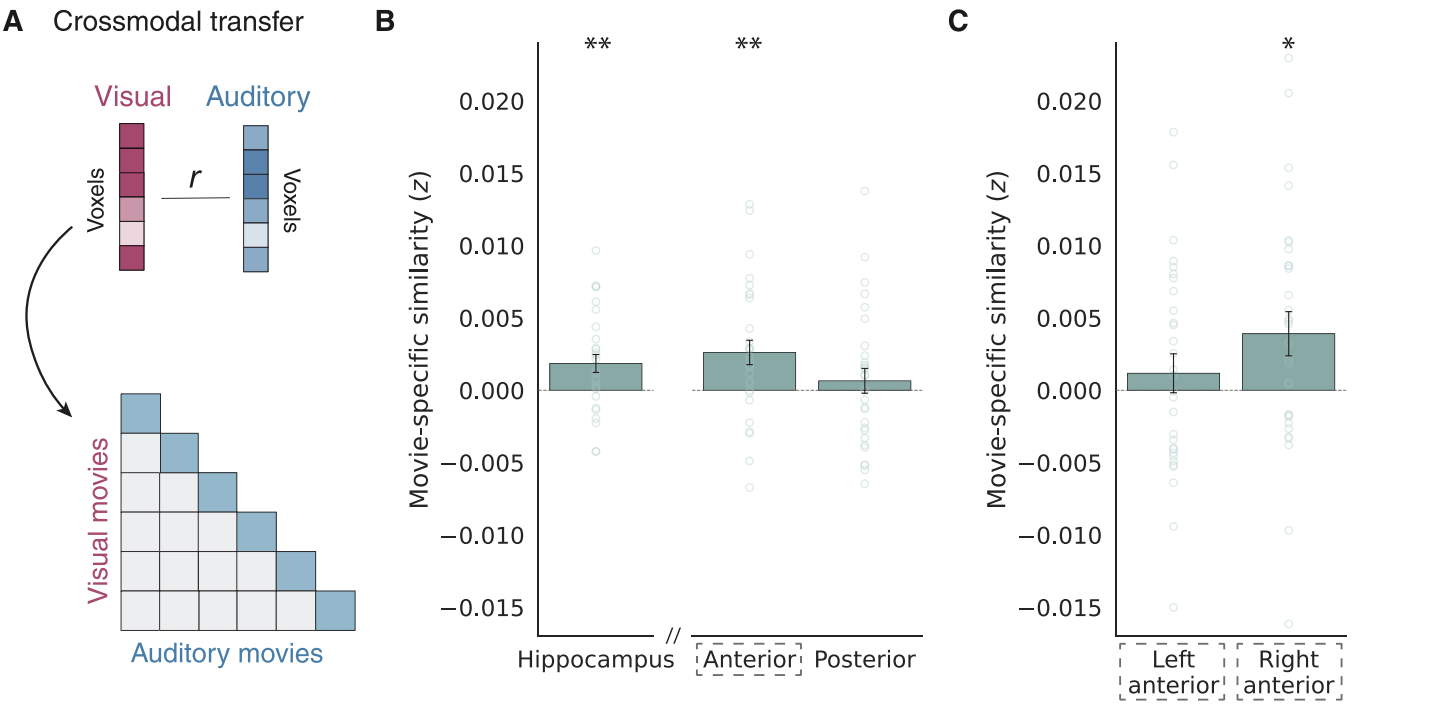


Fig 5. Crossmodal transfer in the hippocampus. (A) To estimate crossmodal transfer, we extended the cross-validated RSA approach by computing voxel-wise correlations between auditory-only and visual-only scene presentations. Movie-specific pattern similarity across modalities was quantified as the signed difference between same-movie (diagonal) and different-movie (off-diagonal) correlations. **(B)** Crossmodal MSPS in the whole hippocampus and anterior/posterior subdivisions. **(C)** Crossmodal MSPS for the left and right anterior hippocampus. Individual participant data shown as circles; error bars indicate SEM. $^{*}P < 0.05$, $^{**}P < 0.01$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.

https://doi.org/10.1371/journal.pbio.3004012.g005

whole hippocampus ($P = 0.003$) and the anterior hippocampus ($P = 0.003$) exhibited significant crossmodal transfer (Fig 5B), whereas the posterior hippocampus did not ($P = 0.22$). Crossmodal transfer was significant in the right anterior hippocampus ($P = 0.016$) but not the left ($P = 0.20$; Fig 5C), with the same asymmetry at the level of the whole hippocampus (right: $P = 0.006$; left: $P = 0.26$; S8A Fig). The direct contrast between hemispheres was not significant in either ROI (both $P > 0.05$). In hippocampal subfields (S9A Fig), there was marginal evidence of crossmodal transfer in DG ($P = 0.091$) and subiculum ($P = 0.091$). In MTL cortex, PHC showed robust crossmodal transfer ($P < 0.001$; S9B Fig). See full descriptive statistics in S4 Table.

The multivariate effects reported above could in principle benefit from memory retrieval, with repeated stimulus exposure allowing for reinstatement of movie-specific details or mental imagery of the absent modality. If so, we would expect that the magnitude of these effects should increase across repeated stimulus presentations. We tested this possibility in two complementary analyses. First, we leveraged the leave-one-run-out cross-validation RSA procedure, in which each run is held out and correlated with the average of the remaining runs. This yields 9 estimates (one per fold), each corresponding to a run's position in the task. We therefore labeled the cross-validated folds by held-out run position (runs 1–9) and correlated run position with the magnitude of our effects of interest (multisensory facilitation and crossmodal transfer) in that run. If repeated exposure strengthened retrieval- or imagery-based contributions, we would expect a positive correlation between run position and effect magnitude. Neither multisensory facilitation in the posterior hippocampus (congruent audiovisual > visual MSPS; Spearman's $\rho = 0.055$, $P = 0.38$) nor crossmodal transfer in the anterior hippocampus ($\rho = 0.037$, $P = 0.55$) strengthened with more stimulus repetition across runs. Second, we ran a split-half analysis comparing effects estimated within the first four runs versus the last four runs (omitting the middle run for balance; S10 Fig). In the first half, the posterior hippocampus showed significant MSPS for the auditory-only ($P = 0.018$) and congruent audiovisual ($P = 0.0497$) conditions; in the second half, there were no significant effects. In contrast, crossmodal transfer in the anterior hippocampus was significant in the second half ($P = 0.005$) but not the first half ($P = 0.12$). However, neither multisensory facilitation nor crossmodal transfer differed significantly between the two halves (both $P > 0.05$). These results raise the possibility that facilitation and transfer might be affected differently by stimulus repetitions, perhaps because crossmodal transfer involves learning amodal abstractions.

To examine multivariate results beyond the hippocampus, we conducted an exploratory whole-brain searchlight analysis. In each participant, we applied the cross-validated leave-one-run-out RSA analysis within a cubic searchlight centered on each brain voxel and assigned the resulting MSPS value to that voxel (Fig 6A). We then assessed reliability across participants (TFCE corrected at $P < 0.05$). A contrast between unisensory auditory and visual conditions showed robust MSPS in auditory and visual cortices, respectively (Fig 6B). Next, we performed a conjunctive contrast to identify regions in which congruent audiovisual stimuli exceeded both unisensory conditions. This revealed multisensory facilitation across the superior temporal lobe, including posterior STG and adjacent posterior perisylvian cortex (planum temporale and parietal operculum), as well as in the insula, temporal pole, and amygdala (Fig 6C). Crossmodal transfer effects at the whole-brain level were robust across perisylvian cortex, angular gyrus, precuneus, and lateral visual cortex (Fig 6D). The full list of peak coordinates for these contrasts is provided in S6 Table.

Discussion

This study focused on the representation of unisensory and multisensory content in the human hippocampus. We found evidence of auditory and visual coding, multisensory facilitation, and crossmodal transfer along the hippocampal long axis. Namely, audiovisual clips were represented more reliably than auditory and visual versions of the same clips in the posterior hippocampus, whereas these unisensory versions shared an abstract representation in the anterior hippocampus. These findings are consistent with a posterior-to-anterior gradient in the hippocampus from fine-grained to abstract representations, respectively [26,27]. Multisensory processing in the hippocampus was also lateralized, with clearer facilitation in the left (posterior) hippocampus and clearer crossmodal transfer in the right (anterior) hippocampus; this is another case of hemispheric specialization in the human hippocampus [30].

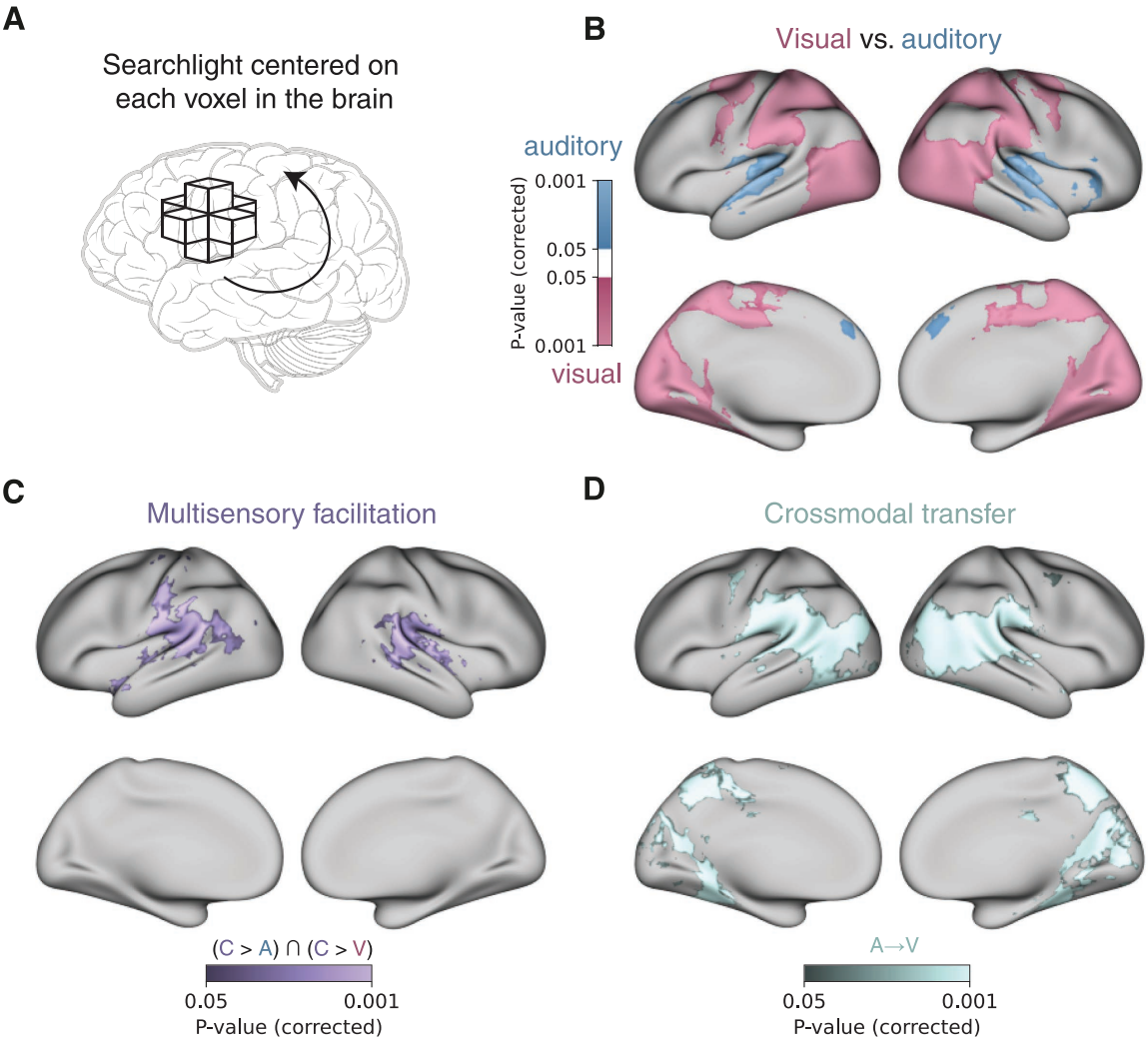


Fig 6. Whole-brain searchlight RSA results. (A) An exploratory searchlight was applied to estimate multivariate RSA effects in regions across the whole brain. **(B)** Group-level auditory-only vs. visual-only MSPS contrast. **(C)** Conjunctive map showing regions in which congruent audiovisual MSPS exceeded both auditory-only and visual-only conditions. **(D)** Above-baseline crossmodal transfer MSPS. All maps were TFCE corrected at $P < 0.05$.

https://doi.org/10.1371/journal.pbio.3004012.g006

Auditory scene representation in the MTL

We observed a clear distinction between analytical approaches: univariate analyses detected only visual responses across the hippocampus, with no activation for auditory stimuli, whereas multivariate pattern analyses revealed representations of auditory scenes. Specifically, there was significant MSPS for auditory-only stimuli in the whole hippocampus and anterior hippocampus. This suggests that auditory content can be encoded in distributed voxel-wise patterns and not readily captured in the mean activity across these voxels, underscoring the need for multivariate approaches to detect nonvisual representations in the human hippocampus.

This dissociation may be theoretically informative, as opposed to simply reflecting the greater sensitivity of multivariate methods. Whereas the univariate analysis reflects overall regional engagement, our multivariate RSA indexes stimulus-specific information carried across distributed voxel patterns. The movie-specific (event-level) structure captured in this analysis is consistent with prominent accounts implicating the hippocampus in the rapid, sparse coding of

individuated events [31–33]. Importantly, this does not preclude a univariate hippocampal response to auditory stimuli, which would be expected given extensive evidence for auditory processing in the hippocampus, primarily observed in nonhuman animals [5]. Univariate auditory responses may be identified under conditions that more strongly drive hippocampal activity, such as paradigms imposing explicit memory demands or those not relying on stimulus repetitions.

Auditory and visual stimuli may differ in perceptual salience or attentional demands, complicating the comparison between unisensory conditions. Several aspects of our data argue against a salience-driven account. First, RTs in the offset detection behavioral task did not differ between unisensory conditions, indicating relatively matched task engagement ($P > 0.05$). Second, the incongruent audiovisual condition elicited lower univariate responses than the visual-only condition across hippocampal ROIs (Fig 2A and 2B); if low-level salience in either modality were driving these responses, adding a second discordant modality should have enhanced them. Third, although salience could in principle modulate the overall magnitude of BOLD responses, our central effects are defined by MSPS, which controls for the pattern similarity across different movies; as such, our multivariate analyses are inherently robust to condition-level differences in salience or attention. Finally, Pearson correlation can be an unreliable measure under conditions of low signal-to-noise ratio [34], which could attenuate pattern similarity for nonpreferred stimuli (e.g., auditory patterns in visual areas). Importantly, our movie-specific similarity metric protects against this concern by baselining against off-diagonal structure within condition.

Prior evidence for auditory coding in the human MTL is limited (cf. [35,36]). Yet, an extensive body of work has established a role for the hippocampus in visual scene perception [10,11]. Our findings provide initial evidence that the hippocampus may additionally support auditory scene perception in the absence of explicit memory demands, consistent with accounts of *domain-general* relational coding in the human hippocampus [37,38]. Testing this hypothesis more directly will require establishing a preference in the hippocampus for auditory scenes over other forms of auditory stimuli, such as auditory objects or simple acoustic stimuli, extending experiments from the visual modality [16,39]. It would further be interesting to test whether the hippocampus is specifically sensitive to the global structure or local details of auditory scenes, as has been examined in the visual modality [17].

We additionally observed robust auditory coding in PHC—a major input structure to the hippocampus via the EC. Although traditionally considered a visual region involved in place and scene processing, our findings align with recent evidence that the PHC responds to spoken descriptions of spatial information and landmarks [35]. Importantly, these effects may reflect crossmodal interactions, as evidenced by stronger visual effects and robust crossmodal generalization in our data. Nevertheless, these findings suggest a potential role for PHC in representing auditory "landscapes" that parallels its established function in visual scene processing.

Multisensory facilitation in the posterior hippocampus

We found enhanced MSPS in the posterior hippocampus for congruent audiovisual stimuli relative to both unisensory modalities, especially in the left hemisphere. Incongruent audiovisual stimuli did not show enhanced pattern similarity relative to unisensory conditions, further supporting the sensitivity of the posterior hippocampus to multisensory congruence. We did not observe multisensory facilitation in the anterior hippocampus or in any individual hippocampal subfields. Together, we interpret these findings as facilitation of perceptual processing in the posterior hippocampus when a matching and synchronized stimulus from another modality is present. This aligns with prior evidence that the posterior hippocampus exhibits stronger connectivity with sensory cortices and preferential coding of fine-grained visual features [26,40,41].

Importantly, these congruent audiovisual representations were not reducible to their unisensory components. Using a cross-validated multiple regression, we found that MSPS for congruent stimuli remained reliably above zero after regressing out the auditory-only and visual-only similarity matrices. This indicates a conjunctive audiovisual code in the hippocampus, in which congruent stimuli elicited unique movie-specific structure. This contrasts with sensory cortical ROIs, in which the unisensory matrices accounted for substantial variance in the congruent representation. Nevertheless, residual

conjunctive structure was evident not only in the hippocampus but also in sensory areas. This may indicate that unisensory matrices estimated from held-out runs could not fully account for the unisensory variance in the audiovisual matrix, leaving unexplained unisensory movie-specific structure in the residuals. Alternatively, these regions could support conjunctive coding of audiovisual inputs, with the residuals reflecting true multisensory movie-specific structure, in line with the role of early sensory regions in multisensory integration [42,43].

A whole-brain searchlight contrast between congruent audiovisual and unisensory conditions revealed robust effects in the STS and posterior STG. This aligns with an extensive literature implicating these regions in audiovisual integration (e.g., [22–25]). These findings situate the posterior hippocampus within a broader network supporting multisensory processing, raising questions about the distinct functional roles of these regions. For instance, does audiovisual facilitation have different consequences for multisensory perception versus memory across these sites? Furthermore, the functional interaction between the hippocampus and these cortical sites represents an important direction for future research.

Crossmodal generalization in the anterior hippocampus

Studies of multisensory perception interpret crossmodal transfer as reflecting "equivalence representations," indexing abstracted, modality-invariant contents [44–47]. The anterior (but not posterior) hippocampus exhibited crossmodal transfer, such that multivariate patterns belonging to the same scene's unisensory auditory and visual stimuli were more similar to each other than to patterns evoked by unisensory stimuli from different scenes. In the visual modality, the anterior aspects of the hippocampus have been implicated in representing abstract associations as opposed to perceptual details [27,40]. Notably, recent work has found that neighboring anterior temporal lobe structures show similar crossmodal object coding, with experience attenuating a visual bias in perirhinal cortex [36]. Our whole-brain searchlight analysis found crossmodal transfer in the STS, pSTG, as well as high-level visual areas, in line with research into amodal representations across vision and audition in cortical sites [48,49].

It is important to acknowledge that such effects can emerge from a range of shared contents across sensory modalities [44], including shared semantic content, low-level temporal coherence (e.g., energy fluctuations), or reinstatement of the absent modality via pattern completion or imagery. The present movie-based paradigm is not well-suited for adjudicating between these interpretations, but other paradigms that systematically manipulate stimulus features could be informative in future studies. Although our analyses across repeated exposures provided limited support for a retrieval-based account of these effects, we observed a trend toward enhanced crossmodal transfer in later runs, which suggests that these effects may be driven in part by learned audiovisual associations (S10 Fig). In fact, even without an increase across runs, it is possible that the pre-experimental exposure phase enabled retrieval processes even early in the experiment. Consistent with this possibility, we observed evidence of perceptual reinstatement in both modalities: auditory-only MSPS was significant in visual cortical ROIs, and visual-only MSPS was significant in auditory cortical ROIs (S3 Fig). These multivariate effects were notably weaker than the patterns for the presented modality. Even so, they suggest some degree of reinstatement of absent modalities. Future work that directly examines crossmodal learning will be necessary to fully adjudicate between perceptual- and retrieval-based accounts.

Implications for hippocampal long-axis specialization and lateralization

We interpret our findings in terms of a posterior-to-anterior gradient from fine-grained to abstract representations, which provides a compelling account of the observed multisensory effects [26,27]. Nevertheless, other accounts of long-axis specialization offer alternative interpretations. One prominent account distinguishes anterior encoding from posterior retrieval [50], raising the possibility that posterior audiovisual facilitation reflects enhanced recall of congruent stimuli. A more recent network-driven account instead links the posterior hippocampus to salience and target detection, and the anterior hippocampus to memory and scene construction [51]. This is broadly consistent with our analyses across repeated exposures, which showed a trend toward stronger crossmodal transfer in the anterior hippocampus (S10 Fig).

Adjudicating between these accounts will require further work, particularly designs that dissociate memory from perceptual and representational contributions to hippocampal function.

Our results additionally revealed hemispheric asymmetries along the long axis: multisensory facilitation was strongest in the left posterior hippocampus, whereas crossmodal transfer was driven by the right anterior hippocampus. The human hippocampus displays strong lateralization across a range of memory and navigation tasks [30,52–56]. Our results are broadly consistent with a left-hemisphere bias for episodic coding of stimulus content and a right-hemisphere bias for coarser, configural representations that are better suited to abstracting across modality-specific detail. The latter aligns with work showing a right hippocampal bias for statistical learning across different types of visual stimuli [57,58]. These interpretations are speculative, as the principles governing hippocampal laterality remain poorly understood. Even so, the asymmetries observed here motivate the use of multisensory stimuli to constrain hypotheses regarding lateralization in the human hippocampus.

Conclusions

Despite decades of research on the human hippocampus, we lack a basic understanding of how this brain structure processes nonvisual perceptual information and integrates across senses. By providing a comprehensive initial account of how multisensory coding is organized in the human hippocampus, this work provides a foundation for bridging across perception, memory, language, and other domains. A critical next step will be to link these unisensory and multisensory representations to the formation of episodic memories.

Methods

Ethics statement

All participants provided written informed consent and were compensated for their time ($25 per hour). The study was approved by the Yale University Institutional Review Board (protocol number: 2000022976) and conducted in accordance with the principles expressed in the Declaration of Helsinki.

Participants

Thirty-two young adults with normal or corrected-to-normal vision were recruited from Yale University and the New Haven communities. Two participants were excluded due to excessive head motion. The remaining participants comprised our final sample ($N = 30$; 16 female; mean age = 24.78 years, age range = 18.80–37.53). Although the effect sizes were unknown preventing a power analysis, this sample size is consistent with, or greater than, other studies of hippocampal function [59].

Stimuli generation

To generate the stimuli, we sourced 12 naturalistic audiovisual movies. Nine movies were from YouTube and three were from Pexels (https://www.pexels.com). In searching for stimuli, we prioritized movies that varied in content (animate/inanimate, indoor/outdoor, etc.) and that maximized audiovisual coherence (i.e., the visual and auditory components were clearly aligned to specific events). Only movies that contained high-fidelity video (1080p or 4K) and audio (48 kHz) were included. Movies containing legible speech were not considered. The final 12 movies included a ping-pong game, wolves howling, waves breaking, birds chirping, race cars passing by, and other multisensory scenes (Fig 1A).

Using Premiere Pro 2024 (Adobe), we cropped each movie into a 6-s segment. These segments were loudness-normalized to the EBU R128 loudness standard with a target loudness of –23 LUFS. The movies were exported at 1,280 × 720 resolution at 30 frames per second, with H.264 video codec and AAC audio codec. The audio was configured as stereo across all movies at 24 kHz.

From these processed movies, we selected two movies to serve as stimuli for the practice block, while the remaining 10 movies made up the experimental stimuli. To construct the task conditions, we used Premiere Pro to modify the movies. For the visual-only condition, we extracted the video component while removing the audio. For the auditory-only condition, we extracted the audio component while removing the video (Fig 1B). To create the incongruent condition, we re-paired the audio and visual components from different movies. This resulted in 90 possible incongruent stimuli (10 videos $\times$ 9 nonmatching audio tracks each). For the practice block, we created two incongruent stimuli using the same mismatching procedure. In total, the experimental stimulus set comprised 120 stimuli: 10 congruent audiovisual movies (original movies), 10 visual-only, 10 auditory-only, and 90 incongruent audiovisual movies. The practice block contained a total of 8 stimuli. Source URLs for all movies and the specific time segments extracted are available on the OSF project (https://doi.org/10.17605/OSF.IO/JCG8A).

Task design and procedures

The experiment consisted of 9 experimental runs, each containing 40 movie presentations: 10 congruent audiovisual, 10 auditory-only, 10 visual-only, and 10 incongruent audiovisual. All congruent, auditory-only, and visual-only stimuli were presented in every run. For the incongruent condition, each run contained all 10 video components and all 10 audio components, each paired with a mismatched audio or visual track. That is, each video was paired with audio from a different movie, with pairings varying across runs (e.g., ambulance video paired with ping-pong audio in one run, with waves audio in another, and so on). This ensured that the visual and auditory content from all movies appeared with equal frequency in every run while varying the specific incongruent pairings across runs, enabling leave-one-run-out cross-validation comparable to the other conditions. The order of stimuli in each run was pseudo-randomized by shuffling trials in each half of the run, ensuring an equal number of stimuli from all conditions in both halves. Finally, the order of the runs themselves was randomized across participants, to prevent order effects for the incongruent condition (the runs were all identical for the congruent, auditory-only, and visual-only conditions). In each run, the movies were shown sequentially with a jittered inter-trial interval (ITI; 4.5, 6, and 7.5 s; Fig 1B). A central fixation cross was presented for the full 6 s of every trial, including during auditory-only trials, and was superimposed over videos in the visual-only and audiovisual conditions. The onset of each movie was triggered by the start of a brain volume acquisition to ensure that any timing delays were accounted for and did not accumulate across the experimental session.

Participants performed a simple cover task while viewing the movies to ensure attention. They were instructed to press a button with their index finger within 1 s of each movie's offset (across all conditions). Visual feedback was provided after each trial, with a red fixation cross indicating an incorrect response (button press during the movie or >1 s after offset), while a green fixation indicated a correct response. At the end of each run, participants were shown their cover task accuracy (% correct responses). This paradigm was designed to maintain participants' attention throughout each movie while imposing similar task demands across conditions. Participants made all responses on an MRI-compatible button box.

In the scanner, and prior to the experimental runs, participants passively viewed the 10 movies in their original audiovisual form to gain familiarity with the auditory and visual components. This pre-exposure was intended to reduce memory demands during the main experiment, consistent with prior work [11]. Participants then completed one practice run that mirrored the experimental procedures but used the two held-out practice stimuli presented in each format.

Audio delivery procedures

To deliver high-fidelity audio, we applied noise attenuation methods to overcome scanner noise. We used OptoActive inflatable headphones with active noise cancelation (OptoAcoustics, Mazor, Israel; https://www.optoacoustics.com), providing approximately 30 dB of noise suppression. Prior to the experimental session, we calibrated the active noise cancelation system based on the specific noise from the fMRI sequence while the participant was positioned in the scanner. Participants also wore semi-open earplugs, intended to provide an additional 10 dB of passive noise attenuation, while

preserving stimulus transmission from the headphones. To maintain the spectral fidelity of auditory stimuli during active noise cancelation, we applied frequency-dependent equalization through the OptoAcoustics system. Following active noise cancelation calibration, we adjusted the volume level based on participants' feedback when listening to an audiovisual clip not included in the experiment. This volume level was held constant throughout the experimental session.

MRI acquisition

Data were acquired on a Siemens Prisma 3T scanner with a 64-channel volume head coil at BrainWorks in the Wu Tsai Institute at Yale University. Functional scans were collected using an echo-planar imaging (EPI) sequence with the following parameters: repetition time (TR) = 1,500 ms; echo time (TE) = 32 ms; 90 axial slices; voxel size = 1.5 × 1.5 × 1.5 mm; flip angle = 64°; multiband factor = 6. We also collected a pair of spin-echo scans with opposite phase-encoding directions for distortion correction (TR = 11,390 ms; TE = 66 ms). We acquired two anatomical scans: a high-resolution 3D T1-weighted (T1w) MPRAGE sequence (TR = 2,400 ms; TE = 2.19 ms; voxel size = 1 × 1 × 1 mm; 208 sagittal slices; flip angle = 8°), and a T2-weighted (T2w) turbo spin-echo sequence (TR = 11,170 ms; TE = 93 ms; 54 coronal slices; voxel size = 0.44 × 0.44 × 1.5 mm; distance factor = 20%; flip angle = 150°) positioned perpendicular to the anterior-posterior axis of the hippocampus to assist with automated segmentation of MTL ROIs. These sequence parameters have been used in recent studies of the hippocampus [59,60].

fMRI preprocessing

fMRI data were preprocessed using fMRIPrep version 23.2.1 [61], based on Nipype 1.8.6 [62]. The T1-weighted anatomical image was corrected for intensity nonuniformity using N4BiasFieldCorrection [63] distributed with ANTs 2.5.0 [64]. The corrected T1w image was skull-stripped using an implementation of the antsBrainExtraction.sh workflow. Brain tissue segmentation of cerebrospinal fluid (CSF), white matter (WM), and gray matter (GM) was performed on the brain-extracted T1w image using FAST in FSL [65]. Brain surfaces were reconstructed using recon-all from FreeSurfer 7.3.2 [66]. Volume-based spatial normalization to standard space was performed through nonlinear registration with antsRegistration (ANTs 2.5.0) to the MNI152 standard template [67].

For each of the nine functional runs, slice-timing correction was applied. Head motion was corrected using mcflirt in FSL [68], and head-motion parameters (six rotation and translation parameters) were estimated for use as nuisance regressors. Susceptibility distortion correction was performed using a fieldmap-based approach, with field inhomogeneities estimated from spin-echo images with opposite phase-encoding directions using topup in FSL [69]. Fieldmaps were not acquired correctly for the first three participants and so susceptibility distortion correction was not applied in these cases. The corrected BOLD reference image was co-registered to the T1w reference using bbregister in FreeSurfer, which implements boundary-based registration [70] with six degrees of freedom.

We excluded trials in which framewise displacement exceeded 1.5 mm on any TR. Runs with more than two excluded trials were removed entirely, and participants with more than two excluded runs were removed from the dataset. This procedure resulted in the exclusion of two participants. In the remaining sample ($N = 30$), one run was excluded from each of two participants and 20 individual trials total were excluded across all participants.

Defining ROIs

MTL structures were segmented from each participant's T1-weighted and T2-weighted anatomical images using the Automatic Segmentation of Hippocampal Subfields (ASHS; [71]) machine learning toolbox. The ASHS algorithm was trained on 24 manual segmentations of hippocampal subfields [72,73]. This procedure generated participant-specific masks for hippocampal subfields CA1, CA2/3, dentate gyrus (DG), and subiculum, as well as extrahippocampal MTL regions including EC, perirhinal cortex (PRC), and parahippocampal cortex (PHC). Anterior and posterior hippocampal ROIs were manually defined for each participant on their native-space T1w image as the most posterior coronal slice in which the uncus

was visible [26,28]. We defined masks for the occipital pole (OP), lateral occipital cortex (LOC), Heschl's gyrus (HG), and posterior superior temporal gyrus (pSTG) from the Harvard-Oxford atlas [74].

Univariate GLM analysis

We constructed event-related GLMs using Nilearn version 0.9.2 [75]. For each participant and run, we included four condition regressors: auditory-only, visual-only, congruent audiovisual, and incongruent audiovisual. Each trial was modeled as a 6-s boxcar function aligned to stimulus onset and convolved with a canonical hemodynamic response function (HRF) using the Glover model with temporal and dispersion derivatives [76]. Additional regressors included six rigid-body motion parameters from fMRIPrep and nuisance regressors derived from high-variance voxels and high-pass filtering. The GLM was fit using an autoregressive AR(1) noise model. No spatial smoothing was applied during model estimation.

For ROI analyses, we extracted subject-specific beta estimates from each ROI and averaged across voxels within the region. For whole-brain analyses, subject-level contrast maps were transformed from native T1w space to MNI space using ANTs [64] and spatially smoothed with a 5 mm FWHM Gaussian kernel. Nonparametric randomization tests were applied using randomise in FSL [77].

Representational similarity analyses

We estimated single-trial activity patterns using GLMSingle [29], which modeled each 6-s movie presentation and generated regularized beta estimates for each trial and voxel. GLMSingle applied voxel-wise HRF optimization. Furthermore, using GLMdenoise, this procedure automatically estimated nuisance regressors and executed noise correction [78]. Finally, regularization was performed to improve the final beta estimates.

For each ROI, we extracted single-trial beta estimates for each voxel, generating a multivariate activity pattern for each trial. For example, each of the 10 auditory-only stimuli was presented once per run across 9 experimental runs, yielding 90 beta patterns for the auditory-only condition (assuming no trial exclusions due to motion). Similarly, we obtained patterns each for the congruent audiovisual, visual-only, and incongruent audiovisual conditions (Fig 3A).

For each condition, we computed a cross-validated representational similarity matrix using leave-one-run-out procedures (Fig 3B). We iteratively left out one run and correlated it with the average of the remaining eight runs, repeating this procedure across all cross-validation folds. All correlation coefficients were Fisher z-transformed before averaging across folds. To quantify pattern distinctiveness for each condition and participant, we computed the difference between the mean diagonal z-value (within-scene similarity) and the mean off-diagonal z-value (between-scene similarity). Higher values indicate more distinct multivariate representations for individual movies.

To assess crossmodal transfer, we applied the same RSA procedure but compared patterns between the auditory-only and visual-only conditions (Fig 5A). Specifically, we correlated multivariate patterns for the same scene presented in different modalities (diagonal elements) and for different scenes across modalities (off-diagonal elements). As before, we computed the difference between mean within-scene similarity (diagonal z-values) and mean between-scene similarity (off-diagonal z-values) to quantify crossmodal transfer within individual scenes.

Cross-validated multiple regression on similarity matrices

To assess whether the congruent audiovisual representations could be explained by a linear combination of unimodal sensory representations, we performed a cross-validated multiple regression in which the congruent audiovisual similarity matrix was modeled as a function of the auditory-only and visual-only similarity matrices. For each participant and ROI, we fit the following regression over the cells of the 10 (movie) by 10 (movie) Fisher z-transformed similarity matrices:

$$\mathrm{vec}(C) = \beta_0 + \beta_A \,\mathrm{vec}(A) + \beta_V \,\mathrm{vec}(V) + \varepsilon$$

where *A*, *V*, and *C* denote the auditory-only, visual-only, and congruent audiovisual conditions, respectively, and vec(·) denotes vectorization of a 10 × 10 similarity matrix. Model estimation followed the same leave-one-run-out cross-validation procedure as our main RSA. On each fold, regression coefficients were estimated from the cells of the eight training folds and applied to the held-out fold to compute predicted values and residuals. Per-fold residual matrices were averaged across folds to yield a single cross-validated residual matrix per participant and ROI. Cells that could not be computed for a given fold due to trial exclusions remained missing and were dropped from both estimation and evaluation.

From this model we derived three quantities: (1) the regression coefficients β_A and β_V, indexing the contribution of each unimodal condition to the congruent audiovisual similarity; (2) the cross-validated R^2 quantifying the proportion of variance in the congruent audiovisual matrix explained by the unimodal matrices on held-out folds; and (3) the residual congruent audiovisual MSPS, computed as the mean diagonal minus mean off-diagonal of the cross-validated residual matrix, indexing MSPS in the congruent audiovisual condition that was not accounted for by the unimodal conditions.

Whole-brain searchlight analyses

For exploratory analyses across the whole brain, we performed the cross-validated RSA approach in searchlights centered on every brain voxel. The searchlight analysis was implemented using BrainIAK (Brain Imaging Analysis Kit; [79]) and parallelized across voxels using the Message Passing Interface (MPI). For each participant, we swept a cubic searchlight (radius = 3 voxels) throughout the whole-brain mask. Searchlights were required to contain at least 80% of voxels within the participant's whole-brain mask to be included in the analysis. At each searchlight location, we performed our RSA procedure for each condition (auditory-only, visual-only, congruent audiovisual, and incongruent audiovisual) as well as for crossmodal transfer. We also computed participant-level conjunctive contrast maps to assess multisensory facilitation (congruent audiovisual greater than both unisensory conditions) and the disruptive effect of audiovisual incongruence (incongruent audiovisual less than mean of unisensory conditions). The resulting correlation coefficient maps were Fisher z-transformed and aligned into MNI152 standard space using ANTs [64] with linear interpolation for group-level analysis. Across participants, statistical maps were computed for each condition and contrast using FSL randomise with threshold-free cluster enhancement (TFCE; [77]).

Statistical analyses

Statistical analyses were conducted in Python (version 3.7.9) using SciPy [80] (version 1.7.3), Statsmodels [81] (version 0.13.2), and custom code. All ROI tests used nonparametric sign-flipping randomization (10,000 iterations) with $\alpha = 0.05$. Univariate GLM effects were evaluated using two-tailed tests; RSA-based pattern similarity was evaluated using one-tailed (greater than) tests given a priori hypotheses of movie-specific representational distinctiveness. Post-hoc pairwise comparisons between conditions used one-tailed randomization tests on the within-participant differences. FDR correction for multiple comparisons was applied separately across anatomical divisions of the ROIs: anterior versus posterior hippocampus; left versus right hippocampus, within each of the whole, anterior, and posterior hippocampus; hippocampal subfields (CA1, CA2/3, DG, and subiculum); MTL cortex (PHC, PRC, and EC); and Harvard-Oxford cortical ROIs (OP, LOC, HG, and pSTG). Empirical null distributions and the corresponding observed group statistics for the multivariate results are shown in S11 Fig. For whole-brain analyses (GLM and RSA searchlight), we used randomise in FSL (10,000 iterations; [77]) with TFCE correction to generate statistical maps.

Supporting information

S1 Table. Univariate activation relative to baseline. *M* = mean β coefficient across participants. SEM = standard error of the mean. *P* = FDR-corrected statistical significance, bolded if $P < 0.05$.
(PDF)

S2 Table. List of peak activations in the whole-brain GLM analysis. Significant clusters were identified using threshold-free cluster enhancement (TFCE) at $P < 0.05$. Bidirectional contrasts are reported in both directions (auditory versus visual; congruent versus incongruent). Anatomical labels were assigned to each peak using the Harvard-Oxford atlas [74]. For each cluster, the peak and any local maxima at least 8 mm apart (labeled a, b, and c) are shown. Clusters smaller than 20 voxels are omitted.
(PDF)

S1 Fig. Univariate and multivariate results in human hippocampal subfields. (A) Hippocampal subfields (CA2/3, DG, CA1, and subiculum) were segmented using the ASHS toolbox [71] trained on an atlas of manual 3T segmentations [72,73]. **(B)** Univariate GLM results in hippocampal subfields. **(C)** Multivariate RSA results in hippocampal subfields. Error bars indicate SEM. × markers indicate individual data points outside the plotted y-axis range. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S3 Table. Multivariate pattern similarity relative to baseline. M = mean movie-specific pattern similarity across participants. SEM = standard error of the mean. P = FDR-corrected statistical significance, bolded if $P < 0.05$.
(PDF)

S4 Table. Multisensory effects. Multivariate evidence for multisensory facilitation (congruent > visual; congruent > auditory) and crossmodal transfer. $(\Delta)M$ = (difference in) mean movie-specific pattern similarity across participants. SEM = standard error of the mean. P = FDR-corrected statistical significance, bolded if < 0.05. Facilitation was tested only in regions where congruent numerically exceeded both unisensory conditions (— = not tested).
(PDF)

S5 Table. Unisensory contributions to congruent audiovisual similarity. Cross-validated multiple regression modeling the congruent audiovisual similarity matrix as a function of the auditory-only and visual-only similarity matrices. β_A, β_V = auditory-only and visual-only regression coefficients. R^2 = proportion of variance in the congruent audiovisual matrix explained by the unisensory matrices on held-out folds. Residual MSPS = movie-specific pattern similarity in the congruent audiovisual condition remaining after regressing out the unisensory contributions. M = mean across participants. SEM = standard error of the mean. P = FDR-corrected statistical significance, bolded if < 0.05. R^2 was tested one-tailed against zero; P values approaching 1 therefore indicate cross-validated variance explained reliably below zero.
(PDF)

S2 Fig. Univariate GLM results in sensory regions and MTL cortex. (A) Mean beta estimates for each condition in sensory cortical ROIs (OP, LOC, HG, and pSTG). **(B)** Mean beta estimates for each condition in MTL cortical ROIs (PHC, PRC, and EC). Individual participant data shown as circles; error bars indicate SEM. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S3 Fig. Multivariate RSA results in sensory regions and MTL cortex. (A) Movie-specific pattern similarity (MSPS) for each condition in sensory cortical ROIs (OP, LOC, HG, and pSTG). HG and pSTG showed significant multisensory facilitation. **(B)** MSPS for each condition in MTL cortical ROIs (PHC, PRC, and EC). Individual participant data shown as circles; error bars indicate SEM. × markers indicate individual data points outside the plotted y-axis range. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S4 Fig. Condition-wise lateralization in the whole and anterior hippocampus. (A) MSPS for each condition in the left and right whole hippocampus. **(B)** MSPS for each condition in the left and right anterior hippocampus. Individual

participant data shown as circles; error bars indicate SEM. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S5 Fig. Conjunctive audiovisual coding in sensory regions. (A) Auditory and visual regression coefficients from the cross-validated multiple regression in sensory cortical ROIs (OP, LOC, HG, and pSTG). **(B)** Variance explained (R^2) by the unisensory matrices. **(C)** Residual congruent audiovisual MSPS after regressing out the auditory-only and visual-only contributions. Individual participant data shown as circles; error bars indicate SEM. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S6 Fig. ROI analysis of unisensory versus incongruent MSPS. (A) No significant differences were observed between incongruent and unisensory (average of auditory and visual) MSPS in the whole hippocampus and anterior/posterior hippocampus. **(B)** PHC showed greater MSPS for unisensory relative to incongruent conditions, suggesting a destructive interaction between nonmatching auditory and visual content. There were no significant differences in PRC or EC. **(C)** Sensory ROIs showed significant suppression in HG and pSTG. Individual participant data shown as circles; error bars indicate SEM. $\times$ markers indicate individual data points outside the plotted y-axis range. $^{*}P < 0.05$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S7 Fig. Whole-brain RSA searchlight of unisensory versus incongruent pattern similarity. MSPS was significantly reduced for incongruent relative to unisensory (average of auditory and visual) throughout widespread cortical sites, including the posterior STG, lateral occipital cortex, angular gyrus, posteromedial cortex, and the lateral and medial prefrontal cortex. A full list of peak coordinates is shown in S6 Table. Corrected using TFCE at $P < 0.05$.
(TIF)

S8 Fig. Crossmodal transfer lateralization in the whole and posterior hippocampus. (A) Crossmodal transfer MSPS in the left and right whole hippocampus. **(B)** Crossmodal transfer MSPS in the left and right posterior hippocampus. Individual participant data shown as circles; error bars indicate SEM. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S9 Fig. Crossmodal transfer in hippocampal subfields and MTL cortex. (A) Crossmodal transfer MSPS across hippocampal subfields (CA2/3, DG, CA1, and subiculum). **(B)** Crossmodal transfer MSPS in MTL cortical ROIs (PHC, PRC, and EC). Individual participant data shown as circles; error bars indicate SEM. $\sim P < 0.1$, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S10 Fig. Split-half analysis of multisensory facilitation and crossmodal transfer. (A) MSPS for each condition in the posterior hippocampus, estimated separately within the first half (runs 1–4) and second half (runs 6–9) of the experiment. **(B)** Crossmodal transfer MSPS in the anterior hippocampus across the same two halves. Individual participant data shown as circles; error bars indicate SEM. $^{*}P < 0.05$, $^{**}P < 0.01$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S11 Fig. Permutation null distributions for multivariate effects. Mean MSPS across participants (colored circles) plotted against the empirical null distributions generated by the same sign-flipping randomization procedure used for statistical testing (10,000 iterations; gray violins). **(A)** MSPS for the auditory-only, visual-only, and congruent audiovisual conditions

in the whole hippocampus. **(B)** The same, shown separately for the anterior and posterior hippocampus. **(C)** Crossmodal transfer MSPS in the whole hippocampus. **(D)** Crossmodal transfer MSPS in the anterior and posterior hippocampus. Gray violins show the kernel density of the null distribution for each ROI and condition. The vertical black bar spans the 5th to 95th percentile of the null and the horizontal black tick marks the null mean. Asterisks denote FDR-corrected significance. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. The data underlying this figure can be found at https://doi.org/10.17605/OSF.IO/JCG8A.
(TIF)

S6 Table. List of peak activations in the whole-brain RSA searchlight analysis. Significant clusters were identified using threshold-free cluster enhancement (TFCE) at $P < 0.05$. Bidirectional contrasts are reported in both directions (auditory versus visual). Anatomical labels were assigned to each peak using the Harvard–Oxford atlas [74]. For each cluster, the peak and any local maxima at least 8 mm apart (labeled a, b, and c) are shown. Clusters smaller than 20 voxels are omitted.
(PDF)

Acknowledgments

We thank Roeland Hancock, Alexander Forrence, Irene Zhou, Erica Busch, Lillian Behm, and Juliana Trach for their help with data collection. We also thank all Turk-Browne lab members for their extensive feedback and support.

Author contributions

Conceptualization: Omri Raccah, Nicholas B. Turk-Browne.

Data curation: Omri Raccah, Yannan Zhu.

Formal analysis: Omri Raccah, Aryan Agarwal.

Funding acquisition: Omri Raccah, Nicholas B. Turk-Browne.

Investigation: Omri Raccah, Yannan Zhu.

Methodology: Omri Raccah, Yannan Zhu, Nicholas B. Turk-Browne.

Supervision: Nicholas B. Turk-Browne.

Visualization: Omri Raccah.

Writing – original draft: Omri Raccah, Nicholas B. Turk-Browne.

Writing – review & editing: Omri Raccah, Aryan Agarwal, Yannan Zhu, Nicholas B. Turk-Browne.

References

1. Lavenex P, Amaral DG. Hippocampal-neocortical interaction: a hierarchy of associativity. Hippocampus. 2000;10(4):420–30. https://doi.org/10.1002/1098-1063(2000)10:4<420::AID-HIPO8>3.0.CO;2-5 PMID: 10985281
2. Li T, Arleo A, Sheynikhovich D. Modeling place cells and grid cells in multi-compartment environments: entorhinal-hippocampal loop as a multisensory integration circuit. Neural Netw. 2020;121:37–51. https://doi.org/10.1016/j.neunet.2019.09.002 PMID: 31526953
3. Radvansky BA, Oh JY, Climer JR, Dombeck DA. Behavior determines the hippocampal spatial mapping of a multisensory environment. Cell Rep. 2021;36(5):109444. https://doi.org/10.1016/j.celrep.2021.109444 PMID: 34293330
4. Tyree TJ, Metke M, Miller CT. Cross-modal representation of identity in the primate hippocampus. Science. 2023;382(6669):417–23. https://doi.org/10.1126/science.adf0460 PMID: 37883535
5. Billig AJ, Lad M, Sedley W, Griffiths TD. The hearing hippocampus. Prog Neurobiol. 2022;218:102326.
6. Thiagarajan D, Sachse S. Multimodal information processing and associative learning in the insect brain. Insects. 2022;13(4):332. https://doi.org/10.3390/insects13040332 PMID: 35447774
7. Thelen A, Matusz PJ, Murray MM. Multisensory context portends object memory. Curr Biol. 2014;24(16):R734–5.

8. Duarte SE, Yonelinas AP, Ghetti S, Geng JJ. Multisensory processing impacts memory for objects and their sources. Mem Cognit. 2025;53(2):646–65. https://doi.org/10.3758/s13421-024-01592-x PMID: 38831161
9. Shams L, Seitz AR. Benefits of multisensory learning. Trends Cogn Sci. 2008;12(11):411–7.
10. Turk-Browne NB. The hippocampus as a visual area organized by space and time: a spatiotemporal similarity hypothesis. Vision Res. 2019;165:123–30. https://doi.org/10.1016/j.visres.2019.10.007 PMID: 31734633
11. Martin CB, Barense MD. Perception and memory in the ventral visual stream and medial temporal lobe. Annu Rev Vis Sci. 2023;9(1):409–34.
12. Nau M, Navarro Schröder T, Bellmund JL, Doeller CF. Hexadirectional coding of visual space in human entorhinal cortex. Nat Neurosci. 2018;21(2):188–90.
13. Hannula DE, Ranganath C. The eyes have it: hippocampal activity predicts expression of memory in eye movements. Neuron. 2009;63(5):592–9. https://doi.org/10.1016/j.neuron.2009.08.025 PMID: 19755103
14. Comrie AE, Frank LM, Kay K. Imagination as a fundamental function of the hippocampus. Philos Trans R Soc B. 2022;377(1866):20210336.
15. Hassabis D, Kumaran D, Vann SD, Maguire EA. Patients with hippocampal amnesia cannot imagine new experiences. Proc Natl Acad Sci U S A. 2007;104(5):1726–31. https://doi.org/10.1073/pnas.0610561104 PMID: 17229836
16. Hodgetts CJ, Voets NL, Thomas AG, Clare S, Lawrence AD, Graham KS. Ultra-high-field fMRI reveals a role for the subiculum in scene perceptual discrimination. J Neurosci. 2017;37(12):3150–9. https://doi.org/10.1523/JNEUROSCI.3225-16.2017 PMID: 28213445
17. Aly M, Ranganath C, Yonelinas AP. Detecting changes in scenes: the hippocampus is critical for strength-based perception. Neuron. 2013;78(6):1127–37.
18. Chun MM, Phelps EA. Memory deficits for implicit contextual information in amnesic subjects with hippocampal damage. Nat Neurosci. 1999;2(9):844–7. https://doi.org/10.1038/12222
19. Stokes MG, Atherton K, Patai EZ, Nobre AC. Long-term memory prepares neural activity for perception. Proc Natl Acad Sci U S A. 2012;109(6):E360-7. https://doi.org/10.1073/pnas.1108555108 PMID: 22109554
20. Kumar S, Bonnici HM, Teki S, Agus TR, Pressnitzer D, Maguire EA, et al. Representations of specific acoustic patterns in the auditory cortex and hippocampus. Proc Biol Sci. 2014;281(1791):20141000. https://doi.org/10.1098/rspb.2014.1000 PMID: 25100695
21. Kumar S, Joseph S, Gander PE, Barascud N, Halpern AR, Griffiths TD. A Brain system for auditory working memory. J Neurosci. 2016;36(16):4492–505. https://doi.org/10.1523/JNEUROSCI.4341-14.2016 PMID: 27098693
22. Setti F, Handjaras G, Bottari D, Leo A, Diano M, Bruno V, et al. A modality-independent proto-organization of human multisensory areas. Nat Hum Behav. 2023;7(3):397–410.
23. Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat Neurosci. 2004;7(11):1190–2.
24. Hu Y, Mohsenzadeh Y. Neural processing of naturalistic audiovisual events in space and time. Commun Biol. 2025;8(1):110. https://doi.org/10.1038/s42003-024-07434-5 PMID: 39843939
25. Pushpita SN, Wehbe L. Two cortical mechanisms of audiovisual processing in the human brain. bioRxiv. 2025.
26. Poppenk J, Evensmoen HR, Moscovitch M, Nadel L. Long-axis specialization of the human hippocampus. Trends Cogn Sci. 2013;17(5):230–40.
27. Collin SH, Milivojevic B, Doeller CF. Memory hierarchies map onto the hippocampal long axis in humans. Nat Neurosci. 2015;18(11):1562–4.
28. Warrington O, Graedel NN, Callaghan MF, Kok P. Communication of perceptual predictions from the hippocampus to the deep layers of the parahippocampal cortex. Sci Adv. 2025;11(21):eads4970. https://doi.org/10.1126/sciadv.ads4970 PMID: 40397746
29. Prince JS, Charest I, Kurzawski JW, Pyles JA, Tarr MJ, Kay KN. Improving the accuracy of single-trial fMRI response estimates using GLMsingle. Elife. 2022;11:e77599. https://doi.org/10.7554/eLife.77599 PMID: 36444984
30. Burgess N, Maguire EA, O'Keefe J. The human hippocampus and spatial and episodic memory. Neuron. 2002;35(4):625–41. https://doi.org/10.1016/s0896-6273(02)00830-9 PMID: 12194864
31. Wixted JT, Squire LR, Jang Y, Papesh MH, Goldinger SD, Kuhn JR, et al. Sparse and distributed coding of episodic memory in neurons of the human hippocampus. Proc Natl Acad Sci U S A. 2014;111(26):9621–6. https://doi.org/10.1073/pnas.1408365111 PMID: 24979802
32. Chadwick MJ, Hassabis D, Weiskopf N, Maguire EA. Decoding individual episodic memory traces in the human hippocampus. Curr Biol. 2010;20(6):544–7. https://doi.org/10.1016/j.cub.2010.01.053
33. Ison MJ, Quian Quiroga R, Fried I. Rapid encoding of new memories by individual neurons in the human brain. Neuron. 2015;87(1):220–30. https://doi.org/10.1016/j.neuron.2015.06.016 PMID: 26139375
34. Walther A, Nili H, Ejaz N, Alink A, Kriegeskorte N, Diedrichsen J. Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage. 2016;137:188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012 PMID: 26707889
35. Häusler CO, Eickhoff SB, Hanke M. Processing of visual and non-visual naturalistic spatial information in the "parahippocampal place area". Sci Data. 2022;9(1):147. https://doi.org/10.1038/s41597-022-01250-4 PMID: 35365659
36. Li AY, Ladyka-Wojcik N, Qazilbash H, Golestani A, Walther DB, Martin CB, et al. Experience transforms crossmodal object representations in the anterior temporal lobes. Elife. 2024;13:e83382. https://doi.org/10.7554/eLife.83382 PMID: 38647143
37. Davachi L. Item, context and relational episodic encoding in humans. Curr Opin Neurobiol. 2006;16(6):693–700.

38. Ranganath C, Ritchey M. Two cortical systems for memory-guided behaviour. Nat Rev Neurosci. 2012;13(10):713–26.
39. Hodgetts CJ, Shine JP, Lawrence AD, Downing PE, Graham KS. Evidencing a place for the hippocampus within the core scene processing network. Hum Brain Mapp. 2016;37(11):3779–94. https://doi.org/10.1002/hbm.23275 PMID: 27257784
40. Robin J, Moscovitch M. Details, gist and schema: hippocampal–neocortical interactions underlying recent and remote episodic and spatial memory. Curr Opin Behav Sci. 2017;17:114–23.
41. Bone MB, Buchsbaum BR. Detailed episodic memory depends on concurrent reactivation of basic visual features within the posterior hippocampus and early visual cortex. Cereb Cortex Commun. 2021;2(3):tgab045. https://doi.org/10.1093/texcom/tgab045 PMID: 34414371
42. Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? Trends Cogn Sci. 2006;10(6):278–85.
43. Ferraro S, Van Ackeren MJ, Mai R, Tassi L, Cardinale F, Nigri A, et al. Stereotactic electroencephalography in humans reveals multisensory signal in early visual and auditory cortices. Cortex. 2020;126:253–64. https://doi.org/10.1016/j.cortex.2019.12.032 PMID: 32092494
44. Lewkowicz DJ, Turkewitz G. Cross-modal equivalence in early infancy: auditory–visual intensity matching. Dev Psychol. 1980;16(6):597–607. https://doi.org/10.1037/0012-1649.16.6.597
45. Mroczko-Wąsowicz A, Grush R. Sensory individuals: unimodal and multimodal perspectives. Oxford University Press; 2023.
46. Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp M, Naumer MJ. Functional imaging of human crossmodal identification and object recognition. Exp Brain Res. 2005;166(3):559–71.
47. Calvert GA, Brammer MJ, Iversen SD. Crossmodal identification. Trends Cogn Sci. 1998;2(7):247–53.
48. Mattioni S, Rezk M, Battal C, Vadlamudi J, Collignon O. Impact of blindness onset on the representation of sound categories in occipital and temporal cortices. Elife. 2022;11:e79370. https://doi.org/10.7554/eLife.79370 PMID: 36070354
49. Xu Y, Vignali L, Sigismondi F, Crepaldi D, Bottini R, Collignon O. Similar object shape representation encoded in the inferolateral occipitotemporal cortex of sighted and early blind people. PLoS Biol. 2023;21(7):e3001930. https://doi.org/10.1371/journal.pbio.3001930 PMID: 37490508
50. Lepage M, Habib R, Tulving E. Hippocampal PET activations of memory encoding and retrieval: the HIPER model. Hippocampus. 1998;8(4):313–22. https://doi.org/10.1002/(SICI)1098-1063(1998)8:4<313::AID-HIPO1>3.0.CO;2-I PMID: 9744418
51. Angeli PA, DiNicola LM, Saadon-Grosman N, Eldaief MC, Buckner RL. Specialization of the human hippocampal long axis revisited. Proc Natl Acad Sci U S A. 2025;122(3):e2422083122. https://doi.org/10.1073/pnas.2422083122 PMID: 39808662
52. Howard LR, Javadi AH, Yu Y, Mill RD, Morrison LC, Knight R, et al. The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. Curr Biol. 2014;24(12):1331–40. https://doi.org/10.1016/j.cub.2014.05.001 PMID: 24909328
53. Maguire EA, Burgess N, Donnett JG, Frackowiak RS, Frith CD, O'Keefe J. Knowing where and getting there: a human navigation network. Science. 1998;280(5365):921–4. https://doi.org/10.1126/science.280.5365.921 PMID: 9572740
54. Maguire EA, Frith CD. Lateral asymmetry in the hippocampal response to the remoteness of autobiographical memories. J Neurosci. 2003;23(12):5302–7. https://doi.org/10.1523/JNEUROSCI.23-12-05302.2003 PMID: 12832555
55. O'Keefe J, Burgess N, Donnett JG, Jeffery KJ, Maguire EA. Place cells, navigational accuracy, and the human hippocampus. Philos Trans R Soc Lond B Biol Sci. 1998;353(1373):1333–40. https://doi.org/10.1098/rstb.1998.0287 PMID: 9770226
56. Spiers HJ, Burgess N, Maguire EA, Baxendale SA, Hartley T, Thompson PJ, et al. Unilateral temporal lobectomy patients show lateralized topographical and episodic memory deficits in a virtual town. Brain. 2001;124(Pt 12):2476–89. https://doi.org/10.1093/brain/124.12.2476 PMID: 11701601
57. Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK. Neural evidence of statistical learning: efficient detection of visual regularities without awareness. J Cogn Neurosci. 2009;21(10):1934–45. https://doi.org/10.1162/jocn.2009.21131 PMID: 18823241
58. Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. Implicit perceptual anticipation triggered by statistical learning. J Neurosci. 2010;30(33):11177–87. https://doi.org/10.1523/JNEUROSCI.0858-10.2010 PMID: 20720125
59. Wammes J, Norman KA, Turk-Browne N. Increasing stimulus similarity drives nonmonotonic representational change in hippocampus. Elife. 2022;11:e68344. https://doi.org/10.7554/eLife.68344 PMID: 34989336
60. Sherman BE, Turk-Browne NB. Statistical prediction of the future impairs episodic encoding of the present. Proc Natl Acad Sci U S A. 2020;117(37):22760–70. https://doi.org/10.1073/pnas.2013291117 PMID: 32859755
61. Esteban O, Markiewicz CJ, Blair RW, Moodie CA, Isik AI, Erramuzpe A, et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. Nat Methods. 2019;16(1):111–6. https://doi.org/10.1038/s41592-018-0235-4 PMID: 30532080
62. Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinform. 2011;5:13.
63. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29(6):1310–20.
64. Avants BB, Tustison N, Song G, et al. Advanced normalization tools (ANTS). Insight J. 2009;2(365):1–35.
65. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage. 2004;23 Suppl 1:S208-19. https://doi.org/10.1016/j.neuroimage.2004.07.051 PMID: 15501092
66. Fischl B. Freesurfer. Neuroimage. 2012;62(2):774–81.

67. Fonov V, Evans A, McKinstry R, Almli C, Collins D. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage. 2009;47:S102. https://doi.org/10.1016/s1053-8119(09)70884-5
68. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage. 2002;17(2):825–41. https://doi.org/10.1016/s1053-8119(02)91132-8 PMID: 12377157
69. Andersson JLR, Skare S, Ashburner J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage. 2003;20(2):870–88. https://doi.org/10.1016/S1053-8119(03)00336-7 PMID: 14568458
70. Greve DN, Fischl B. Accurate and robust brain image alignment using boundary-based registration. Neuroimage. 2009;48(1):63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060 PMID: 19573611
71. Yushkevich PA, Pluta JB, Wang H, Xie L, Ding S-L, Gertje EC, et al. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. Hum Brain Mapp. 2015;36(1):258–87. https://doi.org/10.1002/hbm.22627 PMID: 25181316
72. Hindy NC, Ng FY, Turk-Browne NB. Linking pattern completion in the hippocampus to predictive coding in visual cortex. Nat Neurosci. 2016;19(5):665–7. https://doi.org/10.1038/nn.4284 PMID: 27065363
73. Aly M, Turk-Browne NB. Attention stabilizes representations in the human hippocampus. Cereb Cortex. 2016;26(2):783–96.
74. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31(3):968–80. https://doi.org/10.1016/j.neuroimage.2006.01.021 PMID: 16530430
75. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. Front Neuroinform. 2014;8:14.
76. Glover GH. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage. 1999;9(4):416–29. https://doi.org/10.1006/nimg.1998.0419 PMID: 10191170
77. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE. Permutation inference for the general linear model. Neuroimage. 2014;92(100):381–97. https://doi.org/10.1016/j.neuroimage.2014.01.060 PMID: 24530839
78. Kay KN, Rokem A, Winawer J, Dougherty RF, Wandell BA. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Front Neurosci. 2013;7:247.
79. Kumar M, Anderson MJ, Antony JW, Baldassano C, Brooks PP, Cai MB, et al. BrainIAK: the brain imaging analysis kit. Aperture Neuro. 2022;1(4).
80. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in python. Nature Methods. 2020;17(3):261–72.
81. Seabold S, Perktold J, et al. Statsmodels: econometric and statistical modeling with python. SciPy. 2010;7(1):92–6.