Flexible hippocampal representation of abstract boundaries supports memory-guided choice.

Mariachiara Esposito*, Lubna Abdul*, Ameer Ghouse, Marta Rodriguez Aramendía and Raphael Kaplan†

Department of Basic Psychology, Clinical Psychology, and Psychobiology, Universitat Jaume I, Castelló de la Plana, Spain

* Contributed equally

Abstract

Cognitive maps in the hippocampus encode the relative locations of spatial cues in an environment and dynamically adapt their representation when boundaries geometrically change. In parallel, hippocampal cognitive maps can represent abstract knowledge, yet it's unclear whether the hippocampus is sensitive to geometric changes to the borders, extreme coordinates, of abstract knowledge spaces. Here, we use a memory-guided decision making task to test whether the human hippocampus and medial prefrontal cortex(mPFC) flexibly learn abstract boundary representations in distinct two-dimensional(2D) knowledge spaces. Despite being unnecessary to accurately make decisions, participants conserve a 2D map-like representation of abstract boundaries after the task, where the precision of their representation relates to prior choice accuracy. Finding that the hippocampus and mPFC represent the Euclidean distance of a decision cue to the most proximal boundary during decision making, we then test whether there are brain regions sensitive to boundary-defined contextual changes in abstract spaces. We observe flexible hippocampal representation of abstract boundaries, where the fidelity of this representation relates to task performance. Taken together, our results highlight the importance of hippocampal boundary representations in facilitating flexible knowledge retrieval in dynamically changing abstract contexts.

† Correspondence: kaplan@uji.es

Introduction

The limits of our environment help structure how we think. Physical boundaries, independent of whether they are directly perceived [1] or imagined [2], are known to help guide spatial learning. Boundaries serve as a fundamental spatial orientation cue that guides learning in a host of species including humans [3–6]. In parallel, the discovery of spatially-modulated cells in the hippocampal formation [7] that are sensitive to the shape, distance, and direction to environmental borders [8] has bolstered the idea that hippocampal neurons form a cognitive map of the spatial layout of an environment [9]. Notably, a key feature of hippocampal cognitive maps is their ability to dynamically adapt to geometrical changes in environmental boundaries [10–17]. This dynamic adaptation is thought to facilitate flexible contextual processing in the service of spatial memory formation [18–23]. Taken together with theories that hippocampal cognitive maps extend to episodic memory storage [9, 24], these ideas implicate cognitive maps as an ideal format for flexibly representing abstract knowledge [25–27].

Building on these popular theories, mounting evidence in a variety of species has found that abstract knowledge is assimilated via cognitive map-like representations in the hippocampal formation [28–35, see 27, 36 for reviews] and other regions, namely medial prefrontal/orbitofrontal cortex(mPFC/OFC) [26, 37, 38]. More specifically, human hippocampus and mPFC fMRI signals reflect the Euclidean distance between learned stimuli during decision making [31], providing support for the importance of hippocampal-prefrontal map-like representation of abstract knowledge [26]. The ability of cognitive maps to flexibly code abstract knowledge raises the possibility that abstract boundaries–continuous(e.g., price range) or categorical(e.g., types of cars) limits–could help guide map-like learning of abstract knowledge. Here, we test the hypothesis that map-like representations of knowledge in the human hippocampus and mPFC are sensitive to abstract boundary changes. Addressing whether abstract boundaries are processed by the human hippocampal-prefrontal circuit in a similar way as physical borders during learning, we developed a two alternative forced-choice(2AFC) fMRI task, where participants made similarity judgements on either the price or freshness of various fruit and vegetable goods.

Unbeknownst to participants, the continuous price and freshness variables for each produce good formed two-dimensional(2D) abstract spaces, with four of the goods lying at the extreme coordinates(boundary goods), one good fixed near the center(landmark good), and sixteen cued goods that were located within the boundaries. The fMRI task consisted of encoding and decision phases (Fig. 1A). During encoding phases, the boundary goods and the landmark good were first presented without visible decision variables, and after an intertrial interval(ITI), were individually cued with their respective price and freshness variables at the same time. In the 2AFC, participants were presented with the decision variable for a cued good and decided whether it was more similar to either the landmark good or the most proximal boundary good(in 2D Euclidean distance). During 2AFC trials, the cued good was the only stimulus with a visible price or freshness variable. Crucially, the boundary produce goods had two distinct sets of price-freshness variable coordinates depending on which run they were featured, where each run was defined by one of two distinct abstract spaces for each of the four fMRI runs(square and distorted shapes). In contrast, the landmark coordinates were consistent in the two spaces(Fig. 1B, left).

Presenting participants with 2D coordinates for boundary/landmark goods and subsequently asking them

to make decisions on one variable/dimension allowed us to investigate two types of abstract distance: the distance that participants directly judged during the choices(e.g the relative choice distance based on 1D price/freshness), and the Euclidean distances based on the variables they learned during the encoding phase. We cued participants with 2D variables to avoid overtraining, but participants crucially didn't need both coordinates to accurately calculate the 1D relative choice distance. Following previous work on object and contextual novelty [39, 40], the experimental design distinguished context-related boundary changes (based on the manipulation of the boundaries in the space) from the associative structure related to the actual spaces(the contextual relevance of cued goods). Consequently for the cued goods, half of the cued goods inside the space maintained the same coordinates across abstract spaces (context-independent goods), while the other eight cued goods changed both their price and freshness variables depending on the space(context-dependent goods). Following the fMRI task, participants performed a surprise drag-and-rate task [41] on a laptop outside the scanner, where they were asked to place all the supermarket goods they previously saw on a blank map according to their 2D decision variable coordinates they most strongly associated with that particular good(Fig. 1C). We relied on this post-scan session to assess if participants retained a 2D representation of the decision spaces after the task.

Our analyses first focused on univariate fMRI signal differences related to the contextual relevance of a cued good versus the context type(type of shape) of its' run during 2AFC trials. Following previous studies [31], we tested whether the Euclidean distance between the cued goods to the landmark and closest boundary modulated fMRI representational similarity in hippocampus and mPFC during the 2AFC. Testing whether hippocampus and mPFC signals were sensitive to abstract boundaries, we then used fMRI classifiers on boundary good encoding trials when decision variables weren't visible to determine if there were distinct hippocampal-prefrontal representations for each boundary-defined abstract space. We could then relate each of these analyses back to choice and drag-and-rate performance. We expected hippocampal involvement in detecting contextual relevance and hypothesized that both mPFC and hippocampus would represent the Euclidean distance from the cued good to the proximal boundary good. Given the role of hippocampal cognitive maps in flexibly remapping their internal representation of the environment [9, 19–21, 24], we predicted that hippocampal fMRI signals would be sensitive to abstract boundary-defined contextual changes.

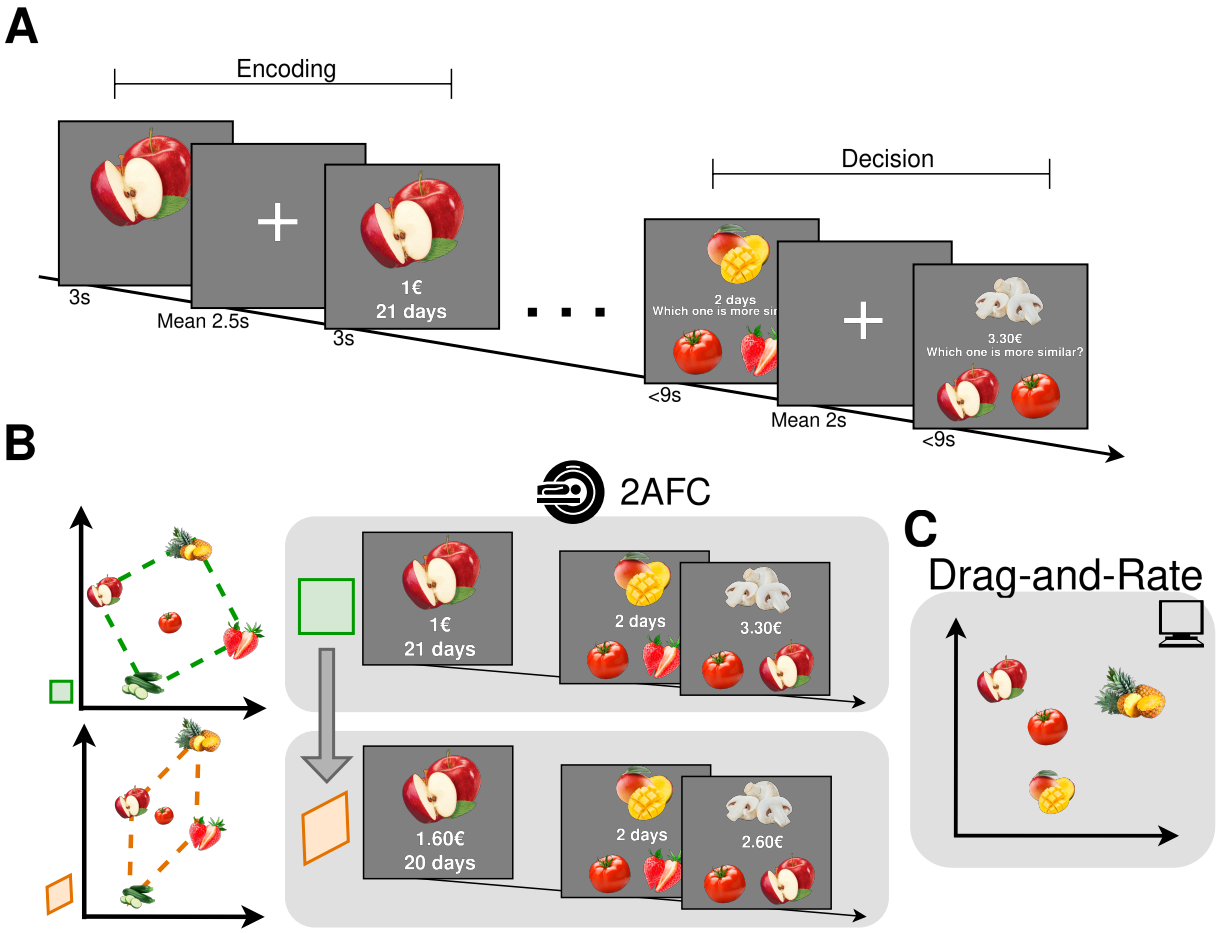


Figure 1: **Experimental paradigm.** A. fMRI Task. During encoding phases, participants were presented with continuous price and freshness variables for each of the boundary and landmark goods. Landmark and boundary goods were first presented without visible decision variables for 3s and then with visible decision variables for another 3s following a jittered intertrial fixation interval(mean=2.5s). Decision phase trials featured one of sixteen cued produce goods inside the space, and participants made a self-paced two alternative forced-choice based on whether that cued good's price or freshness was more similar to that of the landmark or its most proximal boundary good. The cued good's relevant decision variable was visible on the screen throughout the trial, while the landmark and boundary variable weren't shown during decision trials. B. Left: Unbeknownst to participants, price and freshness variable coordinates for the four boundary goods formed two abstract shapes(the "square" shape and "distorted" shape in green and orange respectively), where the boundary goods are located at the extremes of the space and have two different sets of coordinates for each space. In contrast, the landmark good kept its position close to the center. The sixteen cued goods were located within the boundaries. Right: Participants were prompted to encode boundary and landmark good decision variables and then made similarity judgements relative to a cued good in both abstract spaces twice, alternating between runs in the square and distorted shape. C. After the four fMRI runs, participants performed a surprise drag-and-rate task outside the scanner. Participants were instructed to explicitly place every cued, boundary, and landmark good they learned during the task on a blank map according to the location along the two learned price and freshness dimensions they most associated with that particular good.

Results

Behavioral results

The mean accuracy on the fMRI task was 81%(SD: 29%) and reaction time (RT) was 3.49s(SD: 1.79s), with participants responding significantly faster when they made correct decisions($t(28) = 14.06$, $p < .001$). To define the relative choice distance, we computed the absolute difference between the relative distance of each good inside the space from the landmark and from the boundary for both decision variables (Relative choice distance (RD) $= |\Delta\text{Boundary} - \Delta\text{Landmark}|$). Participants' accuracy significantly correlated with relative choice distance ($\rho = 0.23$, $p < .001$; Fig. 2A), indicating that greater distances were less demanding. Similarly, RTs were faster when relative choice distances were greater($\rho = -0.14$, $p < .001$). We then tested whether the type(contextual relevance) of cued goods or the boundary shape of the abstract space affected 2AFC task performance. There was no significant interaction in our factorial design between boundary shape and cued good type for choice accuracy($F(1,28) = 0.3$; $p = 0.87$) or RT($F(1,28) = 2.99$; $p = 0.08$). Furthermore, neither boundary shape ($F(1,28) = 0.81$, $p = 0.22$ for accuracy and $F(1,28) = 0.53$, $p = 0.46$ for RT), nor good type ($F(1,28) = 0.13$, $p = 0.72$ for accuracy and $F(1,28) = 0.73$, $p = 0.39$ for RT) influenced choice accuracy or RT (Fig. 2B). These results emphasize that participants behaved similarly during 2AFC trials independently of the contextual relevance of cued goods and the specific shape of the space where the trial occurred.

Using the drag-and-rate task after scanning, we tested whether participants retained a 2D representation of one of the abstract spaces. We computed the probability of placing the four boundaries in the specific sequences provided in the task(see Materials and Methods). Participants consistently organized boundary goods in the correct order(each boundary good shared a side with its neighboring boundary good) above chance ($p < .001$; Fig. 2C). We then examined whether the arrangement of all observed goods during the task accurately reflected the underlying 2D structure of the abstract spaces, thereby ensuring that all goods were positioned within the designated boundaries. A Monte Carlo simulation with one million iterations(see Materials and Methods for further details) was conducted to determine the likelihood of all sixteen goods falling within the predefined shapes. The simulation showed that the goods were indeed positioned within the boundaries of the abstract space significantly above chance level ($p < .001$; Fig. 2D). This outcome highlights that the boundaries delimited the placement of observed goods within the abstract space. Moreover, post-task fMRI debriefing indicated that none of the participants were aware that the goods formed any kind of 2D space when performing the task.

Next, we asked whether the placement precision was related to participant performance during the 2AFC session. We used Procrustes analysis to calculate the difference between the shape reconstructed by each participant (as defined by their boundary good placement) and the original coordinates of both the square and distorted boundaries. We then correlated the resulting Procrustes distance from each original space with fMRI task performance. The analysis showed that participants more faithfully reconstructed the original distorted shape in the drag-and-rate when they performed better on the decision task($\rho = -0.42$, $p = .026$; Fig. 2E). However, a similar relationship wasn't found for the square shape ($\rho = -0.08$, $p = .70$). Finally, we tested whether this effect could be driven by inconsistencies between the symmetry of the original square shape and participants' reconstruction. However, there was no significant correlation between the reconstruction of the

mean shape coordinates(mean coordinates of square and distorted space boundary goods) and participants' fMRI task performance(ρ = -0.36 , p = .064).

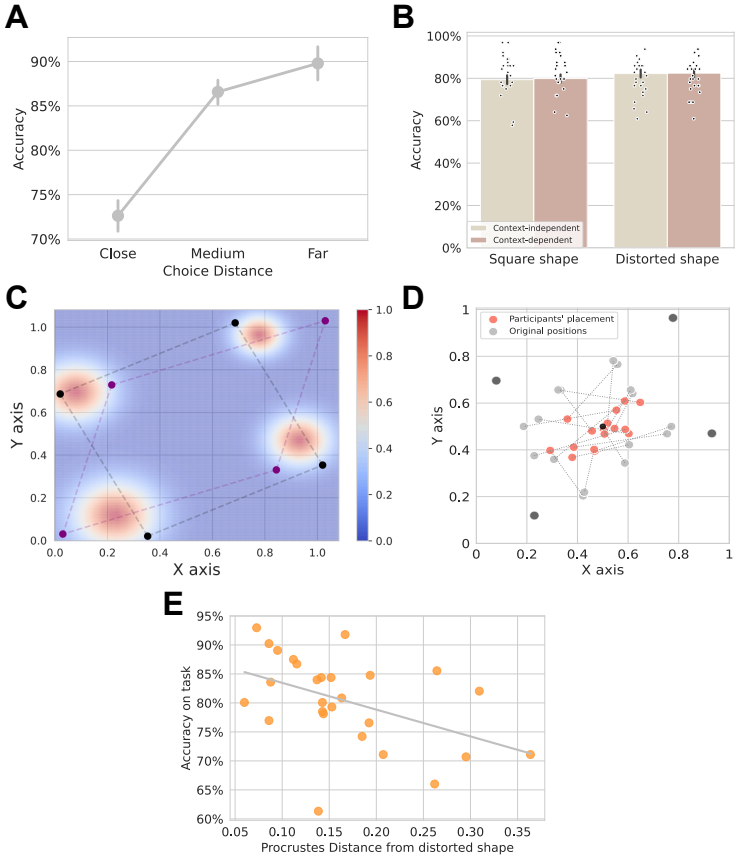


Figure 2: **Behavioral results.** A. Choice accuracy and relative choice distance. Participants' choice accuracy significantly correlated with the relative choice distance of a decision trial ($\rho = 0.23$, p $< .001$). B. Null effect of context shape and a good's contextual relevance on choice accuracy. C. Correlation between choice accuracy and drag-and-rate task precision. Plot displays correlation between Procrustes distance of participants' reconstruction from the original distorted shape with overall accuracy on fMRI task (ρ = - 0.42, p = .026). D. Post-fMRI cued good drag-and-rate placement. Pink dots represent cued goods according to participants' reconstruction, gray dots represent the original position of the goods, and black dots represent participants' mean boundary position across both shaped spaces. Dotted lines connect each cued good's true and participants' mean placement location. E. Heatmap of mean boundary good placement. Warm colors represent standard deviation of participants' placement of the borders, with the mean being the center. Square(black) and distorted(purple) shapes are overlaid for visualization purposes. Colormap reflects normalized distance units. Axes in D and E are normalized for visualization purposes.

fMRI analyses

Univariate results

In a whole-brain fMRI analysis of 2AFC decision trials, we tested whether there were fMRI signal increases that related to the contextual relevance of cued goods and the shape of the abstract space. We ran a factorial analysis comparing differences between the four conditions(contextual relevance of cued goods versus the shape of the abstract space) in a GLM with parametric modulators for relative choice distance, accuracy, and RT. We observed a significant main effect of relative choice distance in the left posterior cingulate cortex, right fusiform gyrus, right precuneus, left lingual gyrus, right superior occipital gyrus, left superior occipital gyrus and left superior parietal lobule(see Supplementary 1 for more details). Examining the effect of accuracy on decision trials, there was a significant main effect in the left angular gyrus for making accurate decisions(see Supplementary Table 2 for more details). Additionally, we observed a main effect of faster RTs right fusiform gyrus, left middle frontal gyrus, right precuneus and right middle frontal gyrus(see Supplementary Table 3 for more details). Conversely, there was a significant main effect of longer RTs in decision trials in the right anterior cingulate cortex, left superior temporal gyrus, right insula, right supramarginal gyrus, right superior frontal gyrus, right inferior temporal gyrus, right middle frontal gyrus, left superior frontal gyrus, left cerebellum, right superior temporal gyrus and left insula(see Supplementary Table 4 for more details). However, there weren't any significant effects of relative choice distance, accuracy, or RT in the hippocampus or mPFC. Contrary to our predictions, we observed no significant interaction between conditions in the hippocampus, nor anywhere else in the brain. The only significant accuracy-by-condition interaction we observed was in the right inferior occipital gyrus(see Supplementary Table 5 for more details), where the correlation with accuracy was higher in the distorted shape($t(25) = 4.90$, $p = .004$). There also wasn't any accuracy-by-condition interaction in the hippocampus or mPFC ROIs. Furthermore, we didn't observe any significant interaction in RT-by-condition or relative choice distance-by-condition in the hippocampus, mPFC, or anywhere else in the brain.

Multivariate RSA

We investigated whether the similarity of neural representations of the cued goods in the abstract spaces were modulated by their Euclidean distance to the landmark good and the most proximal boundary good. Euclidean distance was based on the pre-determined objective distance between the goods. Representational dissimilarity matrices(RDMs) featuring each cued good were created, resulting in a 16x16 matrix for each participant. The RDMs were included in a GLM where the predictors were behavioral RDMs based on experimental variables, including: a) 1D relative choice distance (RD), b) Euclidean distance of a cued good to the landmark, c) Euclidean distance of a cued good to the most proximal boundary good, d) contextual relevance of a cued good (context-dependent/context-independent; Fig. 3A). We observed a significant effect of Euclidean distance to the closest boundary in both bilateral hippocampus($t(25)=2.60$, $p = .015$; mean explained variance: 35%. Fig. 3B, left) and mPFC masks($t(25)= 2.39$, $p = .024$; mean explained variance: 30%. Fig.3B, right), suggesting that sharing similar Euclidean distance from the boundary is a significant predictor of neural similarity across cued goods. We didn't observe a significant relationship

between representational similarity and Euclidean distance to the closest boundary anywhere else in the brain. Conversely, there was no significant effect of Euclidean distance of the cued goods to the landmark good in mPFC($t(25) = 0.62$, $p = .54$) or the hippocampus($t(25)= 0.50$, $p = .62$). Notably, we observed no significant difference between neural RDMs for Euclidean distances from the cued good to the landmark good and closest boundary good in the hippocampus ($t(25) = 0.88$, $p = .38$) and mPFC ($t(25) = 0.80$, $p = 0.43$). There were no significant effects for relative choice distance and representational similarity in the hippocampus($t(25) = 1.70$, $p = .10$), mPFC($t(25) = 1.79$, $p = .09$), nor any other brain region. Additionally, we didn't observe significant effects for the contextual relevance of a cued good (i.e., whether a cued good's 2D coordinates stayed the same between spaces) in the hippocampus($t(25) =-1.77$, $p = .09$), mPFC($t(25) = -0.51$, $p = .61$), or anywhere else in the brain.

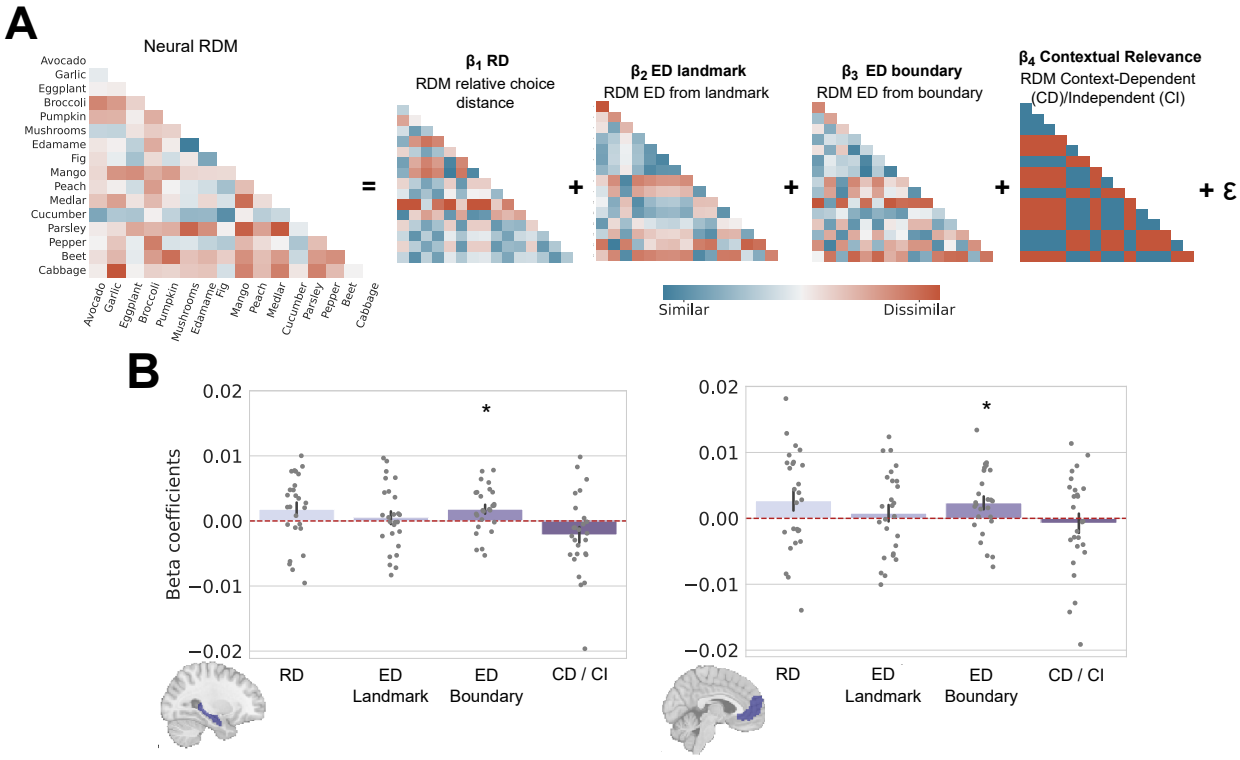


Figure 3: **fMRI Representational Similarity Analysis(RSA).** A. Representational dissimilarity matrices(RDMs) were computed by estimating neural dissimilarity from multivariate general linear models(GLM/neural RDM) and absolute difference of experimental variables(behavioral RDMs) of each pair of cued goods, resulting in 16x16 dissimilarity matrices. The behavioral RDMs were then entered as predictors in the GLM. Colder colors represent higher representational similarity. B. Beta coefficients of RSA GLM in hippocampus and mPFC. The bar plots show beta coefficients resulting from the GLM for each predictor; every participant is an individual dot. Error bars represent SEM. Left: Significant bilateral hippocampal effect of Euclidean distance from the cued good to the closest boundary good ($t(25)=2.60$, $p = .015$); Right: Significant mPFC effect of Euclidean distance from the cued good to the closest boundary good($t(25)= 2.39$, $p = .024$). Asterisks showing significant effects: *= $p< .05$.

To rule out that semantic or visual similarity were driving the hippocampal and mPFC Euclidean distance to boundary effects, we ran control RSAs in the hippocampus and mPFC for both the visual and semantic similarity of cued goods. We didn't observe significant visual or semantic similarity effects in the hippocampus (visual: T(25)=-1.44, p=0.16; semantic: T(25)=-1.38, p=0.18), or mPFC (visual: T(25)=0.39, p=.70; semantic: T(25)=-1.55, p=.13).

Classifier analysis

Asking whether the hippocampus and mPFC flexibly represented abstract spaces consisting of the same goods with different boundary coordinates(price and freshness variables), we tested whether boundary-defined contextual identity(the shape of abstract spaces) was decodable from hippocampal and mPFC signals. We modeled fMRI data from boundary good encoding trials when freshness and price variables weren't visible to the participant(i.e., same visual stimuli in both abstract spaces). We conducted multivariate pattern analysis(MVPA) on encoding phase boundary good trials to test whether the patterns elicited during encoding trials in the square shape were distinguishable from trials in the distorted shape. We used a linear support vector machine(SVM), and applied a leave-one-subject-out cross-validation procedure(Fig. 4A) to test whether the classifier could decode the shape of the abstract space above chance by running one thousand permutations with shuffled labels between runs. In other words, training was conducted on N-1 participants to classify from hippocampal and mPFC ROI encoding trial data whether the remaining participant's fMRI runs alternated between square and distorted contexts, or vice versa. The classifier yielded significant results in the hippocampus(p =.02, accuracy: 60%; Fig. 4B), but not in the mPFC(p = .29, accuracy: 53%). Furthermore, whole-brain MVPA searchlight analyses didn't uncover any significant effects elsewhere in the brain. Testing whether classification of the abstract boundary shape was specific to boundary good trials in the hippocampus, we ran the abstract shape classifier on landmark good encoding trials and observed no significant effect in the hippocampus(p = .47, accuracy: 51%), suggesting that the hippocampal classification effect was specific to the representation of boundary goods. Additionally, there was no significant classification on landmark good trials in mPFC(p = .70, accuracy: 48%).

The false discovery rate(FDR, e.g., the proportion of incorrectly classified observations for predicted class) of the hippocampal classifier was significantly higher for the square shape (Wilcoxon = 29, p = .031). Given this difference, we ran a control classifier to check whether encoding trials in both shapes were classifiable above chance. Here, Gaussian noise was included as a third category and positive predictive values (PPV, e.g., the proportion of correctly classified observations per predicted class) on both the square and distorted shape were evaluated. Square and distorted shapes were accurately classified compared to noise(p= .00098 and p = .00014 respectively; see Materials and Methods and Supplemental Fig. 1 for further details).

Relating classifier accuracy to task behavior, we then asked whether hippocampal classifier accuracy related to participant choice accuracy. Classifier accuracy didn't correlate with general task performance(ρ = 0.22, p = .30). However, since the hippocampal classifier appeared to be biased towards the distorted shape, we then investigated whether the classifier bias towards the distorted shape corresponded with participants' performance biases on the 2AFC task. Relating the difference in task performance in the distorted versus square shape to the hippocampal classifier bias, there was a significant correlation with classifier accuracy (ρ

= 0.47, p = .022; Fig. 4C) This result signifies that the classifier accuracy was highest in the hippocampus for participants that performed better in the distorted shape compared to the square shape.

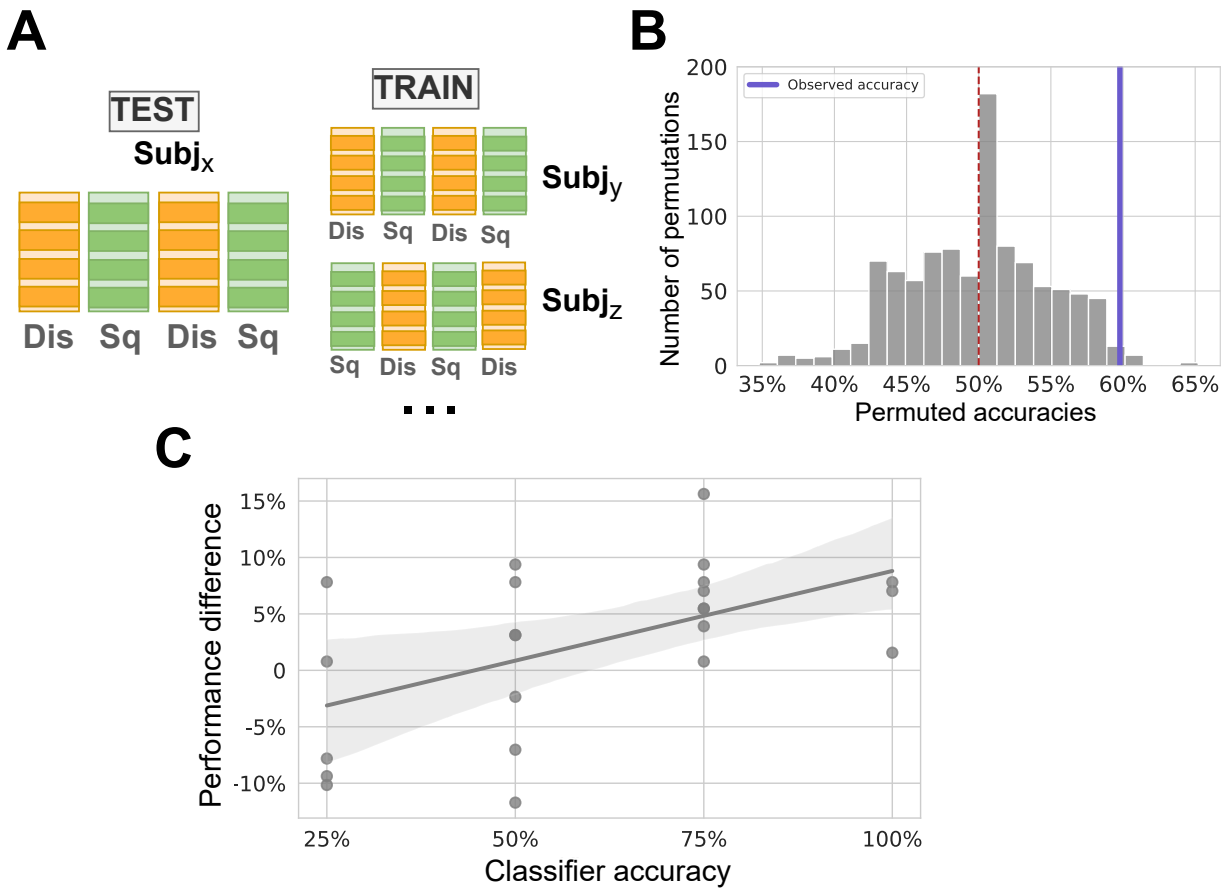


Figure 4: **Hippocampal classifier analysis.** A. Leave-one-subject-out cross-validation procedure. N validation folds were performed for N participants, where each validation comprised a classifier trained on N-1 participants and the Nth participant used as validation. In sum, training was conducted on N-1 participants to classify from their encoding trials(when price/freshness coordinates weren't visible) whether the remaining participant's fMRI runs alternated between square and distorted contexts, or vice versa. B. Hippocampal classifier results based on one thousand permutations with chance level accuracy of 50%(indicated by red dashed line). Bold purple line indicates observed hippocampal classifier accuracy(p =.02, accuracy: 60%). C. Correlation between distorted versus square shape context task performance difference and hippocampal classifier accuracy(ρ = 0.47, p = .022). Positive differences indicate higher performance in distorted versus square shape.

Discussion

Using multivariate fMRI analyses and a novel memory-guided decision making paradigm, we asked whether changing boundaries in abstract spaces are flexibly represented in the hippocampus and mPFC. Participants maintained a 2D map-like representation of the two abstract spaces after the task(Fig. 2). We observed that Euclidean distances between cued and boundary goods modulated fMRI pattern similarity in both the hippocampus and mPFC(Fig. 3). Moreover, we found that boundary-defined contextual identity can be accurately decoded from hippocampal fMRI signals, where classification accuracy is related to task performance(Fig. 4). In what follows, we relate our findings to the growing literature on hippocampal-prefrontal cognitive maps of abstract knowledge and discuss the cognitive implications of the hippocampus flexibly representing abstract boundaries.

Flexible representation of abstract boundaries in the hippocampus

We focused on hippocampus and mPFC following previous literature about the role of both these regions in forming cognitive maps and task representations [26, 27, 30–34, 36, 42]. Our results suggest that abstract boundary representations from decision spaces can be decoded from hippocampal fMRI signals. These results build on previous literature showing that neural ensembles in this region reorganize(remap) dynamically as a function of changes in the geometrical features of spatial environments, flexibly adapting cognitive maps in support of spatial memory [18–24]. Although we decode abstract spaces with different boundary representations in the hippocampus, it's still unclear whether our results reflect hippocampal remapping computations. Future studies can build on spatial paradigms investigating hippocampal remapping by gradually morphing abstract spaces or creating linear boundary changes [16, 17, 19, 20] to determine whether hippocampal spatial remapping computations are conserved in cognitive maps of abstract knowledge.

Medial prefrontal cortex and abstract boundary representations

Previous studies on rodents suggest the mPFC could be a key candidate for building task representations, which allows for rapid generalization of spatial and abstract knowledge [39, 42].Similar findings in humans have highlighted the role of mPFC in supporting relational representations of task structure, even with different sensory events [37, 38, 42, 43]. Notably, we report mPFC involvement in representing Euclidean distances of task-relevant decision variables (Fig. 3B), while the coding of the boundaries does not seem to be specific to the abstract space/context. This finding is consistent with ideas about mPFC involvement in organizing commonalities across experience [44]. A caveat to these findings is that we didn't use an fMRI sequence that robustly measures signals in the most ventral portion of OFC, so these effects should be interpreted with caution.

Implicit versus explicit 2D representations of abstract knowledge

Previous studies have shown how 2D cognitive map-like structure can be learned solely via inferred relationships [30, 31]. In contrast, we explicitly cued participants with both price and freshness decision variables during memory encoding, where they only made decisions on one dimension at a time. Consequently, it

is unclear whether participants encoding the 2D coordinates for landmark and boundary goods facilitated cognitive and hippocampal-prefrontal map-like mnemonic representations. Still, our results show that participants retained a 2D representation of the task goods after decision making that reproduced the structure of the spaces, namely the boundary goods. Each cued good was paired with the most proximal boundary good, where participants placed some of the cued goods closer to other boundary goods while still keeping them within the limits of the space. This suggests learning of an implicit knowledge of task structure in the absence of explicitly learning cued goods' coordinates relative to all four boundary goods [31, 35]. Moreover, participants reported no explicit awareness of the boundaries being positioned at the limits of the space and in a shape-like configuration when questioned after the fMRI task. Taken together, these results hint that even in the absence of extensive multi-day training, the hippocampus can learn and represent boundary-defined abstract contexts.

Absent Neural Representation of the Abstract Landmark

Our fMRI paradigm was designed to test for distinct neural processing for boundaries versus landmarks in an abstract 2D space, yet we didn't observe significant changes in representational similarity with the Euclidean distance to the landmark in the mPFC, hippocampus, or other regions. Given the putative role of the striatum in landmark-based spatial navigation [21] and alignment in abstract knowledge spaces [45], it's surprising that we didn't observe any striatal involvement in representing the Euclidean distance to the landmark. One reason for this might be that unlike the aforementioned studies, participants make 1D proximity judgments instead of aligning the cued good coordinate to the landmark good. In our task, participants only need to know whether a cued good's price or freshness is more proximal to the middle of the space or a boundary good, as opposed to whether the cued good's value is higher or lower (or to the left or right) than the boundary good. Notably, proximity judgments are typically associated with boundary-oriented navigation [21]. Furthermore, the landmark was close to the center of both spaces, so it's unclear how moving the landmark away from the center would change participant behavior. Future paradigms can move abstract landmark locations from the center and test alignment versus proximity judgments to more optimally compare landmark versus boundary learning processes in abstract knowledge spaces.

Relating Abstract and Event Boundaries

Segmenting the continuous flow of experience into distinct events plays a crucial role in episodic memory [46, 47]. Previous research has highlighted the hippocampus as a key region involved in event segmentation and the formation of event-specific [48–52] and context-specific representations [11, 12, 16, 20], which sometimes overlap [53]. Here, we relate hippocampal fMRI signals to more abstract boundary changes based on price and freshness decision variables of supermarket goods in abstract decision spaces. In contrast with spatial and event boundaries that share two primary dimensions, space and time[53], our results point to a more implicit use of boundaries distinct from explicit everyday experience. Given previous ideas about hippocampal cognitive maps also encoding episodic memory spaces [9, 24, 25], it's unclear precisely how abstract boundary coding in the hippocampus might relate to hippocampal event boundary coding driven by spatiotemporal contextual shifts. Future research can investigate potential computational links between flexibly adjusting

hippocampal representations of event and abstract spaces.

Conclusions

We provide evidence that the human hippocampus flexibly represents abstract boundary-defined contexts and that this process can support memory-guided choice behavior. Moreover, we find that participants learn 2D boundary-defined abstract contexts during memory-guided decisions even when it isn't necessary to effectively make a decision. Consequently, our data provide important clues on how the hippocampus can guide decision making across diverse spatial and abstract contexts.

Methods

Participants

34 participants (19 female, mean 23.4 years) recruited through an online questionnaire on Qualtrics (Provo, UT, USA, https://www.qualtrics.com) were remunerated and gave informed written consent to participate in the experiment. The study was approved by the local research ethics committee at Universitat Jaume I(ethics reference: CD/10/2022) and was conducted in accordance with the Declaration of Helsinki protocols. All participants were right-handed, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric disorders. Of the 34 initial participants, 3 were excluded due to poor performance (task accuracy below 60%) and 2 due to technical issues with data collection. For the univariate and RSA fMRI analyses, we excluded an additional three participants due to excessive movement in the scanner and technical difficulties during image acquisition. An additional two participants were excluded from the MVPA analysis as they were missing a large portion of one run due to technical MRI issues, resulting in a sample of 24 participants for the MVPA analysis, 26 participants for univariate and RSA analyses, and 29 participants in the behavioral results.

Task

We developed two abstract supermarket spaces, where the two dimensions were represented by the price(in euros) and freshness(in days) of supermarket goods. Arbitrary price and freshness values were originally set to range from 1 to 33. To prevent the 2D shape configuration from being explicitly visualized by participants and present plausible prices, we then scaled down the price range to be between €1 and €5.80 by dividing the values by five. The four boundary goods were located at the extremes of the decision space in both square and distorted shaped contexts, changing their set of coordinates based on shape. The landmark good was placed near the center of the decision space, maintaining the same coordinates in the square and distorted shaped contexts. All cued produce goods were located within the limits of the boundary goods. Each cued good was paired with a boundary good based on Euclidean proximity in the 2D space. Participants made similarity judgments on price or freshness for cued goods between the landmark and their most proximal boundary good. Additionally, eight of the sixteen cued goods inside the space

maintained the same coordinates across abstract spaces(context-independent goods), while the other eight goods changed both of their positions(context-dependent goods). This experimental manipulation resulted in a 2x2 factorial design that tested the effect of boundary good shape(square/distorted) and cued good contextual relevance(context-independent/context-dependent).

The task was run using PsychoPy toolbox [54] on Python 3.9. Stimuli were supermarket goods overlaid on a gray background, and were viewed through MRI-compatible goggles. Pictures of all goods had the same resolution and were present in a 501x451 resolution frame. Participants were instructed that they'd make similarity judgments on the price and freshness of supermarket produce goods. The price and freshness of produce goods were random and not taken from real-world values in any local supermarket chain. Controlling for participants' preconceived notions of price and freshness for different produce goods, we ensured there was no significant difference between how price/freshness values in the square and distorted spaces relate to real-life price and freshness values (price t(20): -0.06, p=.95; freshness t(20): 0.12, p= .90; real-life produce prices taken from https://www.carrefour.es/supermercado and typical shelf life/freshness values taken from https://www.webstaurantstore.com/article/570/produce-storage-guide.html). Crucially, instructions didn't mention the 2D configuration of abstract shapes or distinguish between the landmark and boundary goods.

Prior to starting the fMRI task, participants underwent a training session on a laptop to ensure adequate task performance. The training consisted of an abbreviated version of the 2AFC task, featuring a novel abstract supermarket space with three boundary goods(a triangle shape), a landmark near the center, and six cued goods located within the limits of the boundaries. After reading the instructions, participants were cued with both price and freshness variables for the boundary goods and the landmark good individually ("encoding phase"). They were instructed to passively view the goods and try to remember their price and freshness variables. Subsequently, a block of decision trials on the six cued goods started("decision phase"). Similar to the fMRI task, participants viewed a cued good on the top of the screen, while the landmark and the most proximal boundary were placed at the bottom. Participants had to decide whether the boundary and the landmark were more similar in price or freshness to the cued good. Participants performed a total of five blocks with one encoding phase at the beginning and twelve decisions each, receiving feedback on their accuracy at the end of each block. To ensure understanding of the task, participants could see the relevant decision variable of the boundary and the landmark of each trial throughout the first block. If participants reached at least 60 percent accuracy on the last block, they proceeded to the fMRI task.

The fMRI task was structured in four separate runs, with participants performing one abstract shape per session with fMRI runs alternating between the square and distorted shape. The order of the sessions was counterbalanced across participants. Participants viewed the same task instructions at the beginning of each fMRI run and were asked to press a button when they were ready to start. During the ensuing encoding phase, participants were cued with both price and freshness variables for each of the boundaries and the landmark individually. At each encoding trial, a picture of a boundary or landmark good was presented on the screen for three seconds, without any cue. Following a white fixation cross(jittered ITI, mean=2.5s), the good was presented again with its price and freshness variables for an additional three seconds(see Fig. 1A, "Encoding"). Participants were instructed to passively view the goods and try to remember their respective freshness and price variables. The order of presentation of boundary and landmark goods were randomized. Participants were never cued on the shape during the run. After the encoding phase, participants started

the decision making phase. At each decision trial, one of the sixteen cued goods inside the space was presented, with its relevant variable for the decision(price or freshness) being listed on the screen throughout the entire trial. At the bottom of the screen, the landmark and the most proximal boundary good were presented without any cued decision variables Fig. 1A, "Decision"). Participants were asked to choose which one between the boundary and landmark good was more similar to the given cued supermarket good. The decisions were made by choosing either one or the other by pressing the left or right button on the fMRI button box. Trials and positioning of correct choices were randomized across participants, where landmarks and boundaries left/right positions on the screen were counterbalanced. Trials timed out after nine seconds. Each of the cued goods inside the space had two types of decisions ("price" decisions and "freshness" decisions) for a total of thirty-two trial-unique choices per abstract space. To avoid excessive cognitive demands, participants were cued with the decision variables for the boundaries and landmark goods once they completed sixteen choices. Participants made decisions for each cued good twice, resulting in sixty-four trials per run interleaved by four encoding phases.

Following the fMRI task, participants performed a surprise drag-and-rate task from the Meadows web-based platform [41] on a laptop. They were presented with a blank map at the center of the screen with price and freshness index for x and y axes. (Fig. 1C). Axes were counterbalanced across participants, such that for some participants the x axis was indicated with "price" label and in others the x axis label was indicated with "freshness" label. All supermarket goods from the 2AFC task were presented right next to the blank map. Participants were asked to position each good according to the price and freshness coordinates(location) they most strongly associated with it, without time limitations.

For behavioral and fMRI analysis, we defined the relative choice distance for each trial. The relative choice distance was computed as the absolute difference between the distance of a given good from the landmark and from the boundary (Relative choice distance (RD) $= |\Delta\text{Boundary} - \Delta\text{Landmark}|$). We counterbalanced trials so that the mean relative choice distances for each condition–i.e. decision spaces, good types, and landmark/boundary proximity–were the same. Additionally, each of the cued goods inside the space had landmark and boundary as the correct answer depending on the trial (e.g, a good could be more similar to the landmark good in price trials and more similar to the boundary good in freshness trials), affording better coverage of the 2D abstract spaces. Additional variables were the Euclidean distance of the supermarket good from the landmark good and the boundary goods.

Procrustes analysis

Procrustes analysis [55] was used to study the correlation between choice accuracy and drag-and-rate task precision, via the Procrustes library and Python 3.10. We used this statistical method to quantify the similarity between the shapes in participants' positioning. By finding the best possible alignment or transformation between the matrices to maximize their similarity. This alignment involved scaling, rotating, and translating the data matrices to minimize the differences between corresponding points while preserving the overall structure.

Permutation testing

First, to test data from the drag-and-rate task against chance distribution, we computed the probability of having five 2D points appearing in a specific sequence. The total number of possible sequences is $5!^2$, since we have five points in total and two coordinates. Therefore, for each permutation of the points in one dimension, there is a corresponding permutation in the other dimension. The probability distribution used is a discrete uniform distribution as each sequence of the five points has an equal probability of occurring, which means that the probability of each sequence is the same. The order in which the elements appeared in a sequence doesn't affect their probabilities or outcomes. In this context, exchangeability suggests that any particular arrangement of the points is equally likely to occur as any other arrangement, provided that all points share similar characteristics and there is no intrinsic ordering among them. Specifically, the probability of any sequence is given by:

$$p_{X=x} = \frac{1}{N}$$

Where X is the random variable representing the sequence of five points, x is a specific sequence, and N is the total number of possible sequences.

Monte Carlo simulation

We used a Monte Carlo simulation algorithm to determine the probability of good placement delimited by the abstract boundaries. We drew sixteen points(the location of cued goods) in a 2D space within the defined geometric shape, a polygon made with the four vertices of the two shapes that holds the maximum area. The probability was then defined as the frequency of Monte Carlo trials, where all sixteen cued goods were placed inside the geometric shape of interest over 1,000,000 iterations.

MRI Data acquisition

Whole-brain structural and functional MRI data were acquired using a 3T General Electric Signa Architect magnetic resonance imaging (MRI) scanner using a 32-channel head coil. Structural T1-weighted images were acquired using an MPRAGE sequence (TR = 2.5 s; flip angle = 8, and slice thickness = 1 mm). Four whole-brain functional runs were acquired using a T2*-weighted multi-band echo-planar imaging (EPI) sequences (voxel size= 2.5 mm isotropic, TR = 2 s; TEs = 20ms; flip angle $\theta = 90$; resolution matrix = 80 $\times$ 80, multi-band acceleration factor = 2).

Preprocessing

Data were processed within the fMRIPrep 23.0.2 framework [56]; which is based on Nipype 1.8.6 [57]. The T1-weighted image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection, distributed with ANTs 2.3.3, and used as T1w-reference. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using FAST (FSL 6.0.5.1). Brain surfaces were reconstructed using

recon-all (FreeSurfer 7.3.2), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle. Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. For each of the functional scanning runs the following preprocessing within the fMRIPrep 23.0.2 framework was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. The estimated fieldmap was then aligned with rigid-registration to the target EPI (echo-planar imaging) reference run. The field coefficients were mapped on to the reference EPI using the transform. fMRI runs were slice-time corrected to 0.96s (0.5 of slice acquisition range 0s-1.92s) using 3dTshift from AFNI. The blood-oxygen-level-dependent (BOLD) fMRI signal reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration. Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 6.0.5.1). Several confounding time-series were calculated based on the preprocessed BOLD signal: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by [58]). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (CompCor [59]). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. The exclusion criteria for excessive movement was set with a FD greater than 0.20. The pre-processed BOLD time series data of each run were modeled with a general linear model (GLM) separately for univariate (using SPM 12) and multivariate analyses (using Nilearn 0.10.1 on Python 3.9, (https://nilearn.github.io/). In addition to task-relevant regressors, both GLMs contained twelve nuisance regressors, six of which pertaining to participants' head motion (three rotation parameters and three translation parameters) and six pertaining to anatomical physiological noise (aCompCor). Moreover, BOLD fMRI data were smoothed with a 8 mm full-width half-maximum Gaussian kernel for univariate analyses, while no smoothing was performed for multivariate analyses.

Univariate analysis

The preprocessed smoothed BOLD fMRI time series data of each run were modeled with a general linear model (GLM) on SPM 12 (https://www.fil.ion.ucl.ac.uk/spm/). We used a box-car function and the default canonical hemodynamic response function (HRF). The GLM contained a regressor for each of the four factorial conditions (shape by cued good contextual relevance: square context-independent, square context-dependent, distorted context-independent, and distorted context-independent), resulting in four conditions of interest. For each condition, we added parametric regressors based on relative choice distance, response times and accuracy. This resulted in a β image of each condition of interest and its parametric modulators

for each participant. For second-level analyses, we computed group level T-maps for each regressor and ran a 2x2 ANOVA(shape x cued good contextual relevance) based on our factorial design.

Multivariate analysis

The preprocessed (unsmoothed) BOLD time series data of each run were modeled with a general linear model (GLM) on Nilearn using a boxcar function and the default canonical hemodynamic response function(HRF). For multivariate analyses, we ran two separate GLMs. In GLM1, we added regressors accounting for each of the sixteen supermarket goods inside the spaces during decision making. This resulted in a β estimate of activity for all trials of each good while participants were making decisions, on which we computed good-by-good RDMs. In GLM2, we accounted for timepoints in the encoding trials when price and freshness were not shown. A β estimate of activity was thus obtained for each boundary good, excluding visual differences between stimuli in the square and distorted shape.

Representational similarity analysis

We ran representational similarity analysis (RSA) to investigate whether the behavioral variables in our memory-guided decision task relate to fMRI voxel patterns [60]. Representational dissimilarity matrices(RDMs) were compared across experimental conditions. We extracted voxels from hippocampus and mPFC for the β estimate of each good (GLM1), and computed RDMs of each cued good pair, resulting in a 16x16 neural RDM for each participant(Fig. 3A). RDMs of experimental variables were computed by calculating the absolute difference of that particular variable for each good pair, and entered those behavioral RDMs as predictors in a GLM where the dependent variable was the neural RDM. All RDMs were standardized before performing regression. The analysis resulted in a β coefficient of each predictor, for which we computed the mean of each participant and tested statistical significance against zero with a one sample t-test.

To rule out potential confounding effects of similarity of the products presented on screen, we ran a control RSA for both visual and semantic similarity of the cued goods. For visual features, we generated vector embeddings from the pictures of the products used during the task (https://github.com/minimaxir/imgbeddings) and computed RDMs of their cosine similarity. For semantic similarity, we used a pre-trained word2vec model (https://github.com/piskvorky/gensim-data) to compute cosine similarity of word vector embeddings. We performed Spearman correlations between the resulting RDMs with neural RDMs of cued goods for every participant, and assessed statistical significance of the correlation coefficients on the group level with a one sample t-test.

In addition to ROI-based RSA, a searchlight-based RSA was conducted to investigate changes in neural pattern similarity related to 2D distances outside of our ROIs. We used β estimates of each good obtained from GLM1 and ran a four voxel radius searchlight analysis. We computed a neural RDM of the searchlight spheres based on each voxel of the brain [60]. Similarly to the ROI-based RSA, we then ran a GLM including the following predictor variables: a) relative choice distance; b) Euclidean distance of a good to the landmark; c) Euclidean distance of a good to the most proximal boundary good; d) contextual relevance of cued good (context-dependent/context-independent). The regression coefficients of the GLM for each searchlight sphere

resulted in a whole-brain β-map of each predictor variable for each participant. In the second-level analysis, the regression coefficients were tested against zero using a one sample t-test. Analysis was performed using a Python open-source RSAtoolbox v.3.0 (https://github.com/rsagroup/rsatoolbox), and statsmodels toolbox v. 0.14.0 (https://github.com/statsmodels/statsmodels).

Multivariate pattern analysis (MVPA)

We tested whether decision spaces with different shaped abstract boundaries were decodable from hippocampus and mPFC. A decoding analysis was conducted by using data from encoding trials of each shape and modeled data from the portion of the encoding phases when boundary good coordinates(price and freshness variables) weren't visible to participants (GLM2). This resulted in a whole-brain β-map for each boundary in a run, averaged across encoding phases. Voxel patterns were extracted from the boundary β-maps in the hippocampus and mPFC ROIs. Since we were interested in decoding the activity from the different abstract spaces and not from the single boundary goods, mean signals were computed across the four boundary goods, obtaining mean patterns in ROIs for each run/shape. These patterns were included as features for the classifier. Asking whether we could decode which runs were square or distorted from encoding phase trial data in hippocampus and mPFC, two class labels of "square" and "distorted" were used for the four runs. We implemented a linear support vector machine(SVM) [61] with a principal component analysis(PCA) accounting for 90% of variance and applied a leave-one-subject-out cross-validation procedure [62]. Consequently, N validation folds were performed for N participants, where each validation comprised a classifier trained on N-1 participants and the Nth participant used as validation. The average validation classification accuracy across validation folds was used to characterize the expected decoding accuracy for the ROI. To check for differences in classification accuracy between square and distorted shapes, we computed the false discovery rate (FDR) of both shapes from the classifier confusion matrix. The FDR represents the proportion of incorrectly classified observations for predicted class, and is computed by calculating the ratio of false positives on all the positive calls (e.g, incorrectly predicted squares on all predicted squares). Testing whether the classifier could accurately decode above chance, we ran one thousand permutations with shuffled labels. Investigating whether the classification effect was exclusive to the boundary goods in the abstract spaces, we then ran the same classifier based on the landmark good, similarly extracting signals from our ROIs for when participants were looking at the landmark good (without price and freshness features) during the encoding phases of each shape.

To determine hippocampal classifier performance in distinguishing noise from the two distinct abstract space configurations, we implemented a linear SVM with a standard scalar preprocessing and ANOVA feature selection, selecting the 200 most responsive voxels in the hippocampus [61]. Data was augmented with a counterbalanced set of noise patterns by sampling and assigning two gaussian noise pattern vectors per participant for each cross-validation. We then computed the positive predictive value (PPV) for the square and distorted shape. The PPV represents the proportion of correctly classified observations per predicted class, and is defined by the proportion of true positives (TP) on all the positive calls; therefore, it gives information about the precision of the classification analysis in decoding each shape against the other and against noise. To check whether the decoding precision was above chance, we conducted five hundred label

permutations by shuffling the square, distorted, and noise labels. These permutations were nested with five hundred validations corresponding to sampling a unique realization of gaussian noise pattern vectors.

To detect potential signals outside of our ROIs, we conducted searchlight-based MVPA, testing whether signals from square and distorted boundaries in other areas across the brain could be accurately decoded. The same procedure as the ROIs classifier was followed, using β estimates of each boundary obtained from GLM2. We applied a linear SVM kernel, PCA accounting for 90% of variance, and leave-one-subject-out cross-validation to searchlight spheres (radius=4 voxels) across each voxel of the brain. This resulted in whole-brain F-scores maps for which p-values were calculated. For the whole-brain searchlight MVPA classifier, Nilearn 0.10.1(https://nilearn.github.io/) was used.

Statistical Analyses

Given our hypothesis, we focused our analyses on hippocampal and mPFC regions of interest(ROI) masks. We defined the bilateral hippocampal ROI using the Wake Forest University (WFU) Pickatlas [63, 64] integrated on SPM. The mPFC ROI was defined as the rostral medial wall of the prefrontal cortex, omitting ventral and frontopolar portions of mPFC, where robust signals weren't observed in all participants. For univariate and RSA/MVPA searchlight analyses outside the two ROIs, we report effects surviving an uncorrected statistical threshold of p¡.001 and correction for multiple comparisons(family-wise error correction, FWE) of p< .05 at the whole-brain level.

Acknowledgements

We thank Doerte Kuhrt, Matthias Nau, and Ignacio Polti for helpful advice on multivariate fMRI analyses. We thank Michal Zareba for fMRI preprocessing advice. This research is supported by grants awarded to RK from the Valencian Community's Program for the Support of Talented Researchers (CIDEGENT/2021/027), Universitat Jaume I Research Advancement Plan(UJI-B2022-45), and Spanish Science, Innovation, and University Ministry(PID2021-122338NA-I00).

References

[1] S. A. Lee, J. F. Miller, A. J. Watrous, M. R. Sperling, A. Sharan, G. A. Worrell, and J. Jacobs. Electrophysiological signatures of spatial boundaries in the human subiculum. *Journal of Neuroscience*, 38(13):3265–3272, 2018.

[2] C. M. Bird, C. Capponi, J. A. King, C. F. Doeller, and N. Burgess. Establishing the boundaries: the hippocampal contribution to imagining scenes. *Journal of Neuroscience*, 30(35):11688–11695, 2010.

[3] Steven Poulter, Tom Hartley, and Colin Lever. The neurobiology of mammalian navigation. *Current Biology*, 28(17):R1023–R1042, 2018.

[4] K. Cheng and N.S. Newcombe. Is there a geometric module for spatial orientation? squaring theory and evidence. *Psychonomic Bulletin & Review*, 12(1):1–23, 2005.

[5] J. B. Julian, A. T. Keinath, S. A. Marchette, and R. A. Epstein. The neurocognitive basis of spatial reorientation. *Current Biology*, 28(14):R1059–R1073, 2018.

[6] J. Margules and C. R. Gallistel. Heading in the rat: determination by environmental shape. *Animal Learning & Behavior*, 16(4):404–410, 1988.

[7] J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175, 1971.

[8] C. Lever, S. Burton, A. Jeewajee, J. O'Keefe, and N. Burgess. Boundary vector cells in the subiculum of the hippocampal formation. *Journal of Neuroscience*, 29:9771–9777, 2009.

[9] John O'keefe and Lynn Nadel. *The hippocampus as a cognitive map*. Oxford university press, 1978.

[10] R. U. Muller and J. L. Kubie. The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. *Journal of Neuroscience*, 7(7):1951–1968, 1987.

[11] C. Barry, R. Hayman, N. Burgess, and K. J. Jeffery. Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*, 10(6):682–684, 2007.

[12] S. Leutgeb, J. K. Leutgeb, C. A. Barnes, E. I. Moser, B. L. McNaughton, and M. B. Moser. Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, 309(5734):619–623, 2005.

[13] K. Hardcastle, S. Ganguli, and L. M. Giocomo. Environmental boundaries as an error correction mechanism for grid cells. *Neuron*, 86(3):827–839, 2015.

[14] F. Savelli, D. Yoganarasimha, and J. J. Knierim. Influence of boundary removal on the spatial representations of the medial entorhinal cortex. *Hippocampus*, 18(12):1270–1282, 2008.

[15] J. Krupic, M. Bauza, S. Burton, C. Barry, and J. O'Keefe. Grid cell symmetry is shaped by environmental geometry. *Nature*, 518(7538):232–235, 2015.

[16] T. J. Wills, C. Lever, F. Cacucci, N. Burgess, and J. O'Keefe. Attractor dynamics in the hippocampal representation of the local environment. *Science*, 308(5723):873–876, 2005.

[17] T. Stensola, H. Stensola, M.-B. Moser, and E. I. Moser. Shearing-induced asymmetry in entorhinal grid cells. *Nature*, 518:207–212, 2015.

[18] J. L. Bellmund, W. De Cothi, T. A. Ruiter, M. Nau, C. Barry, and C. F. Doeller. Deforming the metric of cognitive maps distorts memory. *Nature Human Behaviour*, 4(2):177–188, 2020.

[19] J. B. Julian and C. F. Doeller. Remapping and realignment in the human hippocampal formation predict context-dependent spatial behavior. *Nature Neuroscience*, 24(6):863–872, 2021.

[20] A. T. Keinath, O. Rechnitz, V. Balasubramanian, and R. A. Epstein. Environmental deformations dynamically shift human spatial memory. *Hippocampus*, 31(1):89–101, 2021.

[21] C. F. Doeller, J. A. King, and N. Burgess. Parallel striatal and hippocampal systems for landmarks and boundaries in spatial memory. *Proceedings of the National Academy of Sciences*, 105(15):5915–5920, 2008.

[22] T. Hartley, I. Trinkler, and N. Burgess. Geometric determinants of human spatial memory. *Cognition*, 94(1):39–75, 2004.

[23] B. Steemers, A. Vicente-Grabovetsky, C. Barry, P. Smulders, T. N. Schröder, N. Burgess, and C. F. Doeller. Hippocampal attractor dynamics predict memory-based decision making. *Current Biology*, 26 (13):1750–1757, 2016.

[24] Howard Eichenbaum and Neal J Cohen. Can we reconcile the declarative memory and spatial navigation views on hippocampal function? *Neuron*, 83(4):764–770, 2014.

[25] D. Schiller, H. Eichenbaum, E. A. Buffalo, L. Davachi, D. J Foster, S. Leutgeb, and Charan R. Memory and space: towards an understanding of the cognitive map. *Journal of Neuroscience*, 35(41):13904–13911, 2015.

[26] R. Kaplan, N. W. Schuck, and C. F. Doeller. The role of mental maps in decision-making. *Trends in Neurosciences*, 40(5):256–259, 2017.

[27] T. E. Behrens, T. H. Muller, J. CR. Whittington, S. Mark, A. B. Baram, K. L. Stachenfeld, and Z. Kurth-Nelson. What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, 100 (2):490–509, 2018.

[28] D. Aronov, R. Nevers, and D. W. Tank. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, 543(7647):719–722, 2017.

[29] E. H. Nieh, M. Schottdorf, N. W. Freeman, R. J. Low, S. Lewallen, S. A. Koay, and D. W. Tank. Geometry of abstract learned knowledge in the hippocampus. *Nature*, 595(7865):80–84, 2021.

[30] M. M. Garvert, R. J. Dolan, and T. E. Behrens. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife*, 6:e17086, 2017.

[31] S. A. Park, D. S. Miller, and E. D. Boorman. Inferences on a multidimensional social hierarchy use a grid-like code. *Nature Neuroscience*, 24(9):1292–1301, 2021.

[32] A. O. Constantinescu, J. X. O'Reilly, and T. E. Behrens. Organizing conceptual knowledge in humans with a gridlike code. *Science*, 352(6292):1464–1468, 2016.

[33] N. W. Schuck and Y. Niv. Sequential replay of nonspatial task states in the human hippocampus. *Science*, 364(6447):eaaw5181, 2019.

[34] E. B. Knudsen and J. D. Wallis. Hippocampal neurons construct a map of an abstract value space. *Cell*, 184(18):4640–4650, 2021.

[35] S. Theves, G. Fernández, and C. F. Doeller. The hippocampus maps concept space, not feature space. *Journal of Neuroscience*, 40(38):7318–7325, 2020.

[36] R. A. Epstein, E. Z. Patai, J. B. Julian, and H. J. Spiers. The cognitive map in humans: spatial navigation and beyond. *Nature Neuroscience*, 20(11):1504–1513, 2017.

[37] Y Niv. Learning task-state representations. *Nature neuroscience*, 22(10):1544–1553, 2019.

[38] N. W. Schuck, M. B. Cai, R. C. Wilson, and Y. Niv. Human orbitofrontal cortex represents a cognitive map of state space. *Neuron*, 91(6):1402–1412, 2016.

[39] R. Kaplan, A. J. Horner, P. A. Bandettini, C. F. Doeller, and N. Burgess. Human hippocampal processing of environmental novelty during spatial navigation. *Hippocampus*, 24(7):740–750, 2014.

[40] E. Mızrak, N. R. Bouffard, L. A. Libby, E. D. Boorman, and C. Ranganath. The hippocampus and orbitofrontal cortex jointly represent task structure during memory-guided decision making. *Cell reports*, 37(9), 2021.

[41] N. Kriegeskorte and M. Mur. Inverse mds: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in psychology*, 3:28167, 2012.

[42] R. C. Wilson, Y. K. Takahashi, G. Schoenbaum, and Y. Niv. Orbitofrontal cortex as a cognitive map of task space. *Neuron*, 81(2):267–279, 2014.

[43] A. B. Baram, T. H. Muller, H. Nili, M. M. Garvert, and T. E. Behrens. Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron*, 109(4):713–723, 2021.

[44] A. Gilboa and H. Marlatte. Neurobiology of schemas and schema-mediated memory. *Trends in cognitive sciences*, 21(8):618–631, 2017.

[45] R. Kaplan and K. J. Friston. Entorhinal transformations in abstract frames of reference. *PLoS biology*, 17(5):e3000230, 2019.

[46] S. DuBrow and L. Davachi. The influence of context boundaries on memory for the sequential order of events. *J. Exp. Psychol. Gen.*, 142:1277–1286, 2013.

[47] I. K. Brunec, M. Moscovitch, and M. D. Barense. Boundaries shape cognitive representations of spaces and events. *Trends in cognitive sciences*, 22(7):637–650, 2018.

[48] A. Ben-Yakov, N. Eshel, and Y. Dudai. Hippocampal immediate post-stimulus activity in the encoding of consecutive naturalistic episodes. *J. Exp. Psychol. Gen.*, 142:1255–1263, 2013.

[49] J. M. Zacks, T. S. Braver, M. A. Sheridan, D. I. Donaldson, A. Z. Snyder, J. M. Ollinger, R. L. Buckner, and M. E. Raichle. Human brain activity time-locked to perceptual event boundaries. *Nat. Neurosci.*, 4:651–655, 2001.

[50] C. Baldassano, J. Chen, A. Zadbood, J. W. Pillow, U. Hasson, and K.A. Norman. Discovering event structure in continuous narrative perception and memory. *Neuron*, 95:709–721, 2017.

[51] M. Silva, C. Baldassano, and L. Fuentemilla. Rapid memory reactivation at movie event boundaries promotes episodic encoding. *Journal of Neuroscience*, 39(43):8538–8548, 2019.

[52] J. Chen, Y. C. Leong, C. J. Honey, C. H. Yong, K. A. Norman, and U. Hasson. Shared memories reveal shared structure in neural activity across individuals. *Nature neuroscience*, 20(1):115–125, 2017.

[53] A. J. Horner, J. A. Bisby, A. Wang, K. Bogus, and N. Burgess. The role of spatial boundaries in shaping long-term event representations. *Cognition*, 30:1–30, 2016.

[54] J Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, and J. K. Lindeløv. Psychopy2: Experiments in behavior made easy. *Behavior research methods*, 51:195–203, 2019.

[55] F. Meng, M. Richer, A. Tehrani, J. La, T. D. Kim, P. W. Ayers, and F. Heidar-Zadeh. Procrustes: A python library to find transformations that maximize the similarity between matrices. *Computer Physics Communications*, 276:108334, 2022.

[56] O. Esteban, C. J Markiewicz, R. W. Blair, C. A Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, and M. Snyder. fmriprep: a robust preprocessing pipeline for functional mri. *Nature methods*, 16(1):111–116, 2019.

[57] K. Gorgolewski, C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. S. Ghosh. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics*, 5:13, 2011.

[58] J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L Schlaggar, and S. E. Petersen. Methods to detect, characterize, and remove motion artifact in resting state fmri. *Neuroimage*, 84:320–341, 2014.

[59] Y. Behzadi, K. Restom, J. Liau, and T. T. Liu. A component based noise correction method (compcor) for bold and perfusion based fmri. *Neuroimage*, 37(1):90–101, 2007.

[60] N. Kriegeskorte, M. Mur, and P. A. Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:249, 2008.

[61] F. Pereira, T. Mitchell, and M. Botvinick. Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1):S199–S209, 2009.

[62] G. Valente, A. L. Castellanos, L. Hausfeld, F. De Martino, and E. Formisano. Cross-validation and permutations in mvpa: Validity of permutation strategies and power of cross-validation schemes. *Neuroimage*, 238:118145, 2021.

[63] J. A. Maldjian, P. J. Laurienti, R. A. Kraft, and J. H. Burdette. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. *Neuroimage*, 19(3):1233–1239, 2003.

[64] J. A. Maldjian, P. J. Laurienti, and J. H. Burdette. Precentral gyrus discrepancy in electronic versions of the talairach atlas. *Neuroimage*, 21(1):450–455, 2004.

Supplementary material

Region	x	y	z	Cluster Size	Cluster Level p(FWE-corr)	T	Peak Level p(FWE-corr)
L posterior cingulate	-2	-38	39	120	< .001	6.28	.061
R fusiform gyrus	29	-55	-16	285	< .001	6.2	.074
R precuneus	6	-58	24	533	< .001	5.96	.133
L lingual gyrus	-12	-78	-9	404	< .001	5.76	.221
R superior occipital gyrus	31	-78	22	159	< .001	5.46	.294
L superior occipital gyrus	-24	-90	19	109	< .001	5.01	.434
L superior parietal lobule	-32	-63	54	45	.003	4.58	.994

Supplementary Table 1: Relative Choice Distance (RD)

Region	x	y	z	Cluster Size	Cluster Level p(FWE-corr)	T	Peak Level p(FWE-corr)
L angular gyrus	-57	-33	17	30	.030	4.46	.998

Supplementary Table 2: Accuracy

Region	x	y	z	Cluster Size	Cluster Level p(FWE-corr)	T	Peak Level p(FWE-corr)
R Anterior cingulate	6	25	32	673	< .001	8.46	< .001
L superior temporal gyrus	-49	20	-1	1509	< .001	8.33	.001
R insula	44	15	-6	1080	< .001	8.33	.001
R supramarginal gyrus	64	-33	39	701	< .001	8.28	.001
R superior frontal gyrus	29	50	29	301	< .001	7.21	0.006
R inferior temporal gyrus	54	-35	-16	9	0.869	6.98	0.012
L middle temporal gyrus	-59	-30	-4	226	.001	6.85	0.016
R middle frontal gyrus	39	18	47	69	.001	6.58	0.03
L superior frontal gyrus	-27	58	27	127	< .001	6.33	0.054
L Cerebellum	-32	-70	-31	58	.003	6.08	0.1
R superior temporal gyrus	59	-25	-4	279	< .001	5.75	0.224
L insula	-37	-15	17	59	.003	5.37	0.545

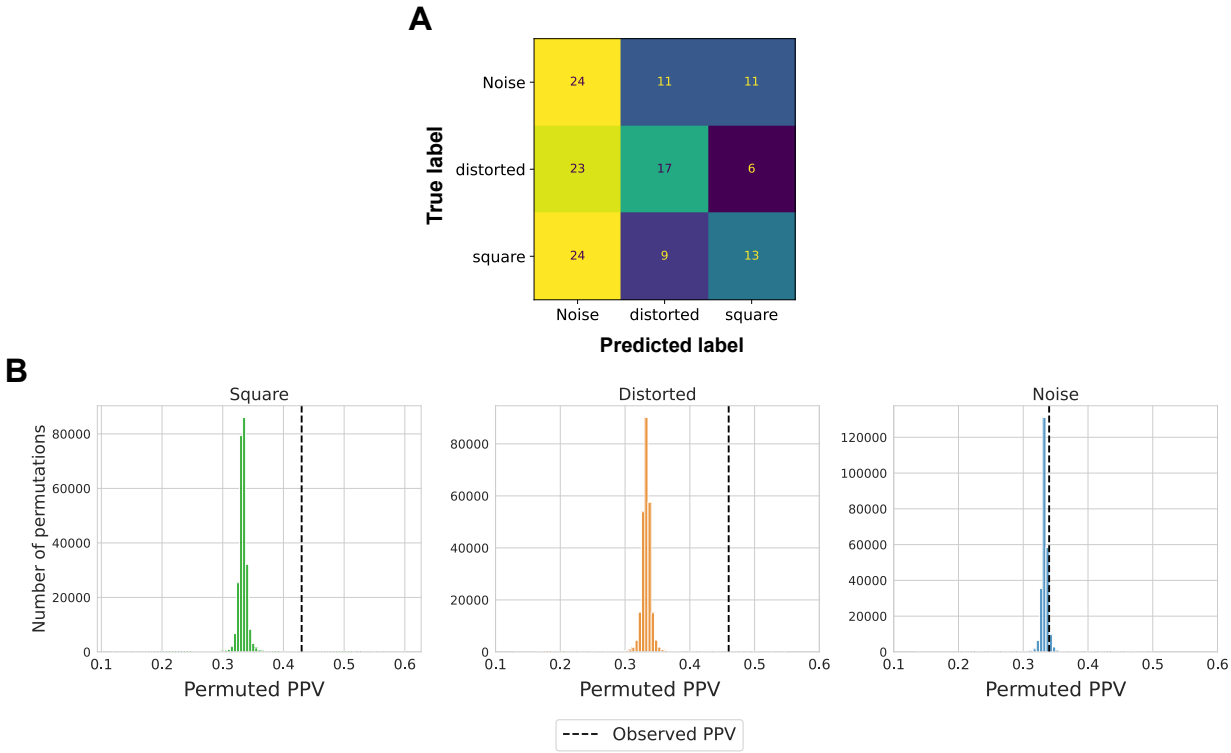
Supplementary Table 3: Reaction Times +

Region	x	y	z	Cluster Size	Cluster Level p(FWE-corr)	T	Peak Level p(FWE-corr)
R fusiform gyrus	41	-70	-11	5724	$< .001$	9.68	$< .001$
L middle frontal gyrus	-27	0	57	42	.018	4.68	.952
R precuneus	1	-55	42	67	.001	4.65	.961
R middle frontal gyrus	26	0	39	35	.041	4.64	.964

Supplementary Table 4: Reaction Times -

Region	x	y	z	Cluster Size	Cluster Level p(FWE-corr)	F	Peak Level p(FWE-corr)
R inferior occipital gyrus	49	-63	-9	45	.049	21.01	.620

Supplementary Table 5: Accuracy By Shape



Supplementary Figure 1: **Control classifier analysis.** A. Confusion matrix of the control classifier, with labels "Square", "Distorted" and "Noise". Noise pattern vectors were sampled from a gaussian distribution. B. Permuted PPVs of control classifier. We ran two hundred and fifty thousand permutations by nesting five hundred unique realizations of gaussian noise pattern vectors and five hundred permutations with shuffled labels. We then assessed whether classifier precision was above chance by comparing observed PPV (positive predictive value) with the null distribution. Plot shows PPVs obtained from permutations. Black dashed line indicates observed PPV (square = 0.43, distorted = 0.46, noise = 0.34).