Representation of Task Structure in Human Hippocampus

Eda Mizrak[1], Nichole R. Bouffard[1], Laura A. Libby[1], Erie Boorman[1,2], Charan Ranganath[1,2]

[1] Center for Neuroscience, University of California, Davis

[2] Department of Psychology, University of California, Davis

Corresponding Authors:

Eda Mızrak (edamizrak@gmail.com) and Charan Ranganath (cranganath@ucdavis.edu)

Other Author information: Nichole Bouffard (nichole.bouffard@mail.utoronto.ca), Laura Anne Libby (lauraannelibby@gmail.com), Erie Boorman (edboorman@ucdavis.edu)

Phone: 530-757-8865; Fax: 530-757-8640;

Address: 1544 Newton Court, Davis, CA 95618, U.S.A.

SUMMARY

Memories of previous experiences can be used to guide future decisions in similar situations. Recent evidence suggests that the hippocampus might support decision-making by forming representations that capture common elements across different events (e.g., "cognitive maps" or "schemas"). Here, we used functional magnetic resonance imaging (fMRI) to test how the human hippocampus represents decision relevant information extracted from previous experiences. Participants performed a task in which they learned to predict a customer preference for foods in four different store contexts. The task was structured such that we could examine the degree to which hippocampal representations reflected generalized information about the store contexts, food items, and also the kind of information that was relevant to decisions on a given trial. Results showed that hippocampal activity patterns carried information about the kind of information that was currently relevant to a decision. Across different store contexts, hippocampal representations differentiated between context-determined (deterministic) decisions and context-invariant (probabilistic) decisions. Results also showed that information about store contexts was represented by the hippocampus, but contrary to what might be expected, similar contexts were hyper-differentiated from one another. These results suggest that the hippocampus may support decision-making by systematically mapping relationships between task relevant information, decisions, and outcomes.

INTRODUCTION

Our decisions often rely on what we remember from previous experiences. For instance, in your local grocery store, you might skip the produce section because previously the produce did not live up to your expectations. Episodic memory — the ability to remember past events — enables us to build links between previous experiences and future decisions. Humans are very good at extracting regularities across experiences to make adaptive decisions (i.e. not buying the produce) and often benefit from using information derived from previous experiences (i.e. produce not being satisfactory). Recent findings are consistent with this idea (e.g., Bornstein, Khaw, Shohamy, & Daw, 2017; Bornstein & Norman, 2017; Duncan, Doll, Daw, & Shohamy, 2018; Santoro, Frankland, & Richards, 2016), and they suggest a key role for the hippocampus in memory-guided decision-making (see Doll, Shohamy, & Daw, 2015; Mizumori & Tryon, 2015 for review). Little is known, however, about what type of information is extracted from past experiences and how hippocampus organizes this information to guide future decisions.

There are at least three ways to explain how the hippocampus represents experiences that are relevant to decisions. One possibility is that the hippocampus forms a systematic cognitive map of a specific environment (O'Keefe & Nadel, 1978), in which information about an event is encoded relative to a representation of the spatiotemporal context in which the event took place. Others have argued that the hippocampus forms a cognitive map that is more general and schematic (McKenzie et al., 2016; Preston & Eichenbaum, 2013). A more general, schema like-map could be adaptive in

making decisions, as one might want to generalize across experiences that occurred in similar places. For instance, when planning a trip to the grocery store, one might rely on the hippocampus to pull up a generalized representation of overlapping experiences at Safeway™ grocery store chain. A hippocampal schema might also be useful to apply to specific items encountered across overlapping experiences, such as a specific kind of candy bar. This idea is supported by findings showing that the hippocampus organizes experiences with overlapping features to form schema-like knowledge representation and more broadly, with studies indicating that the hippocampus integrates new experiences with overlapping memories for past experiences (Mack, Love, & Preston, 2018; Morton, Sherrill, & Preston, 2017). Another possibility is that the hippocampus maps experiences in a manner that is both flexible (Eichenbaum, 2017; Ekstrom & Ranganath, 2018; Schiller et al., 2015) and goal-directed (Bornstein et al., 2017; Duncan et al., 2018; Santoro et al., 2016). For instance, some have proposed that the hippocampus encodes an internal model of tasks and situations, systematically mapping relationships between task relevant information, decisions, and outcomes (Behrens et al., 2018; Doll et al., 2015; Kaplan, Schuck, & Doeller, 2017; Wikenheiser & Schoenbaum, 2016). Representation of potential outcomes is also consistent with a growing body of evidence suggesting a role for the hippocampus in "model-based" decision making (Hampton, Bossaerts, & O'Doherty, 2006; Lee, Shimojo, & O'Doherty, 2014; Miller, Botvinick, & Brody, 2017).

A strong version of the latter view would suggest that the hippocampus represents potential decisions in a manner that cuts across contexts. Because

contexts do not always provide useful information for decision-making, it could be adaptive for the hippocampus to represent contexts according to how relevant or predictive they are for a particular decision. For instance, when deciding to buy a fruit at the grocery store, it can be useful to know whether the store generally carries fresh produce, whereas you would probably not use information about the specific store context when deciding to buy a candy bar. Research has generally been consistent with the idea that, within a particular context, the hippocampus represents task-relevant information (Aronov, Nevers, & Tank, 2017; Theves, Fernandez, & Doeller, 2019), though it is unclear whether the hippocampus represents decision-relevant information or task structure in a manner that generalizes across different contexts.

In the present study, we used representational similarity analysis of functional magnetic resonance imaging (fMRI) data to investigate how the hippocampus represents information about contexts, items, and relationships between items and contexts to guide decisions (see Fig. 1 and 2). We developed a novel paradigm in which participants were trained to learn about customer preferences for eight food items in four different grocery stores. After learning the customer preferences in each store, participants were scanned while they decided whether a food was liked or disliked by customers based on previous learning for each food in each store context. The task was designed such that, half of the foods were always "liked" or "disliked" depending on the specific store context. These foods are referred to as "context-determined" (CD) foods, because information about the context was needed in order to make accurate decisions about the customer's preference (i.e., probability of

being "liked" = 1.0 or 0 depending on the store). The other foods had customer preferences that were probabilistic, and this probability was constant across all stores (i.e., probability of being "liked" = 0.75 or 0.25 in every store). These foods are referred to as context-invariant (CI) because the context was irrelevant to decisions about the preferences. Participants were not explicitly informed about the relationship about the stores and foods, and they could therefore only acquire the task structure about the different food types (CD vs. CI) over the course of learning.

In addition to the CD/CI differentiation between foods, we manipulated the shared properties of foods across the stores (*item similarity*) and shared properties of stores across the foods (*context similarity*); see Figure 2. Regarding item similarity, apple and muffin share the same customer preference across all stores which makes these two items similar to each other. In contrast, apple and carrot have different customer preferences across all stores, which makes these two items different from each other. Regarding context similarity, store A and B had the same customer preferences for all the foods which make these two stores similar to each other. In contrast, store A and C had different customer preferences for all the foods which made them different from each other.

If the hippocampus forms schematic representations of shared information between items or contexts, we would expect that hippocampal activity patterns would be more similar across foods or across stores that are similar to each other (e.g., across two different foods that were both liked or disliked in different store contexts or across two different stores where the

same foods were liked or disliked). Finally, if goal-relevant information is prioritized in hippocampal representations, we would expect hippocampal activity patterns to reflect the task structure in a manner that generalizes across contexts (i.e., high pattern similarity between pairs of trials that involved both CD foods or between pairs of trials that involved both CI foods compared to pairs of trials that consisted of a CD food and a CI food). In addition to examining representation of task-relevant information in the hippocampus, we also investigated neural pattern similarity in the orbitofrontal cortex, another region that has been proposed to represent cognitive maps that support goal-directed decision making (Wikenheiser & Schoenbaum, 2016; Wilson, Takahashi, Schoenbaum, & Niv, 2014).

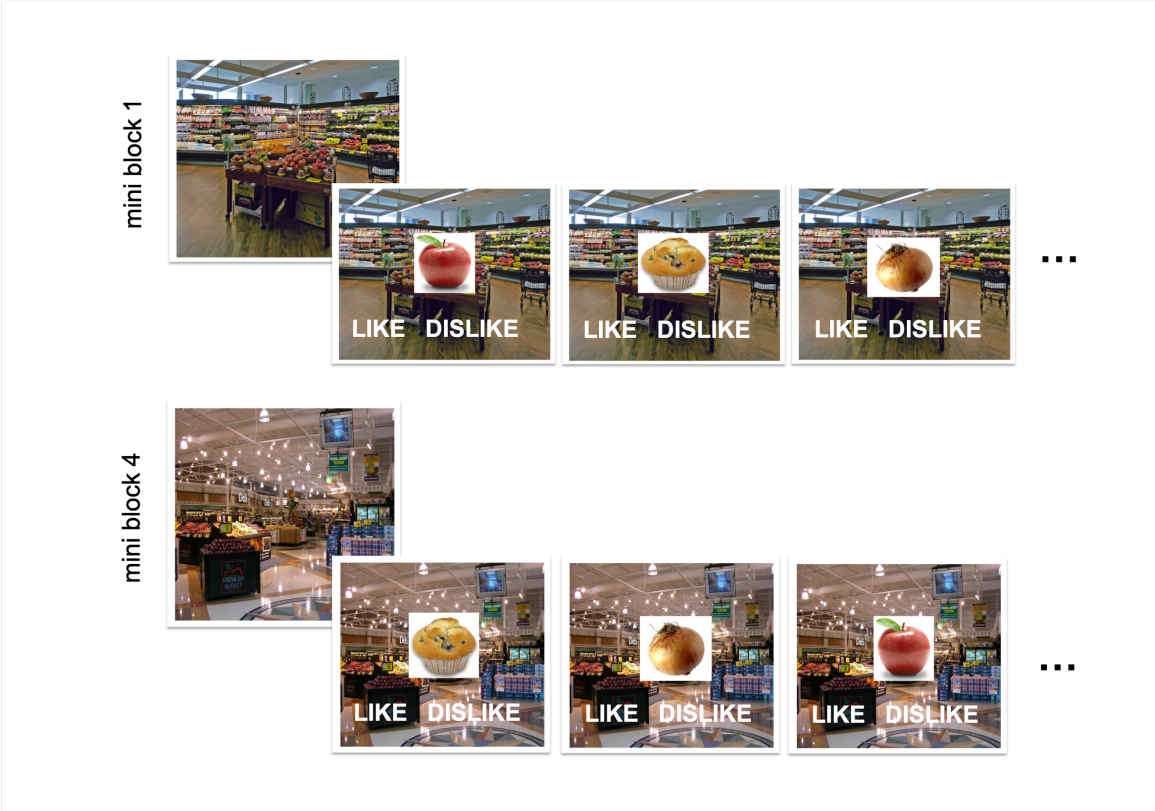

Figure 1. Example Stimuli. Illustration of two mini-blocks from a run in the decision task that took place on Day 2 (fMRI). Participants completed two runs of the decision task and each run consisted of eight mini-blocks (e.g. participants visited all four stores twice in every full run, each store was a

mini-block). Each mini-block started with presentation of one of the store contexts for 6 seconds. Following this, participants were presented with 8 different foods (2 seconds) and predicted the customer preference of the food in the store context by making like/dislike judgments. No feedback was given for decisions. Each food item was shown once in each mini-block which made 8 trials in a mini-block. The time between mini-blocks were jittered (mean = 4 secs). The next mini-block consisted of another 8 trials in another store context. Same eight foods were shown in each mini-block in a different store context but in a randomized order.

Figure 2. Task Structure. During the training phase, participants made decisions about food preferences, and then they received feedback about the actual outcome for that trial. The matrix illustrates the distribution of outcomes for eight different foods (columns) in each of the four stores (rows). Context-Determined foods are defined as those with outcomes that are determined by the store context (left side of the matrix) and Context-Invariant foods are those with probabilities that are the same across stores (right side of the matrix). The task was designed so that there were pairs of similar contexts for which the same foods were liked and disliked in both stores. Additionally, pairs of similar food items shared the same distribution of outcomes across the four store contexts.

RESULTS

Behavioral Results

Prior to scanning, participants completed 12 runs of trial-and-error learning with feedback until they had learned customer preferences for eight foods in four different store contexts up to at least 60% accuracy (see Supplemental Materials for detailed description of the learning phase). Participants were scanned during the decision phase—on each trial, they were presented with a food image overlaid on a store context and were asked to judge the customer preference (either like or dislike) for the presented food in the given store (Figure 1). During this phase, no feedback was given, so participants needed to rely on what they learned previously on Day 1 in order to choose the correct customer preferences for each food. To assess whether participants accurately learned the customer preferences in the experiment, we scored participants' performance on each trial according to whether they selected "like" for a food item that was liked more than 50 % of the time in that particular store, and selecting "dislike" for a food item that was liked less than 50 % of the time in that particular store.

Although participants were not explicitly informed about the distinction between CD and CI trials (and were not given feedback for their responses during the decision phase), we expected their decisions to reflect their past experiences in a way that was consistent with the implicit task structure. We predicted that, for CD trials, optimal decisions about the food outcome would rely on previous experience in the current store context (i.e., whether the food was previously liked in that context). For example, with the information depicted in Figure 2, it would be optimal to predict that the apple would be liked in Stores A and B, and disliked for the apple in Stores C and D. For CI

trials, optimal decisions about the food outcome instead might rely on a generalization across past experiences, irrespective of the current store context. For example, it would be optimal to choose "Dislike" for the banana and "Like" for the donut, regardless of the current store context.

Consistent with this reasoning, we found that participants made optimal decisions on 90% (SD = 9%) of CD trials and 82% (SD = 14%) of CI trials. A direct comparison between the two conditions showed that the proportion of optimal decisions (as defined above) was higher for CD than CI trials ($F(1,21) = 4.48$, $p = .049$). The difference in decisions across CD and CI trials suggests that participants understood the task structure well enough to make different kinds of decisions for CD and CI trials. Detailed behavioral results from each phase (learning Day 1 and decision Day 2) are presented in the Supplemental Material.

Task structure is represented in Hippocampus

Although the difference between CD and CI trials was no explicitly stated, the behavioral data described above indicate that participants successfully learned and differentiated between the two types of trials during the decision phase. If the hippocampus represents task structure in a manner that generalizes across contexts, we would predict higher pattern similarity between pairs of CD trials (e.g., a correct decision about the apple in Store A & a correct decision about the carrot in Store C) and between pairs of CI trials (e.g., a correct decision about the banana in Store A & a correct decision about the donut in Store D), compared to the similarity between CD and CI

trials (e.g., a correct decision about the apple in Store A & a correct decision about the donut in Store D).

To test whether decision-relevant information was represented in a manner that generalized across contexts, we computed voxelwise hippocampal pattern similarity (PS) (collapsed across hemispheres) between trials where the food-store relationship implied with the task structure was the same (CD-CD or CI-CI) and trials where the food-store relationship was different (CD-CI). Importantly, we excluded pairs of trials with the same store context and the same food in order to make sure that any effects observed could not be due to visual similarity (see Methods section for exclusion criteria for pattern similarity analyses). We conducted a one-way ANOVA on the hippocampal PS with Food-Store relationship as a factor (same vs. different). This analysis revealed a significant effect of Food-Store relationship [$F(1,21) = 7.06$, $p = .01$: same (CD/CD or CI/CI) > different (CD/CI)], consistent with the hypothesis that the hippocampus represents information that is task-relevant, even when the store contexts differ (see Figure 3).

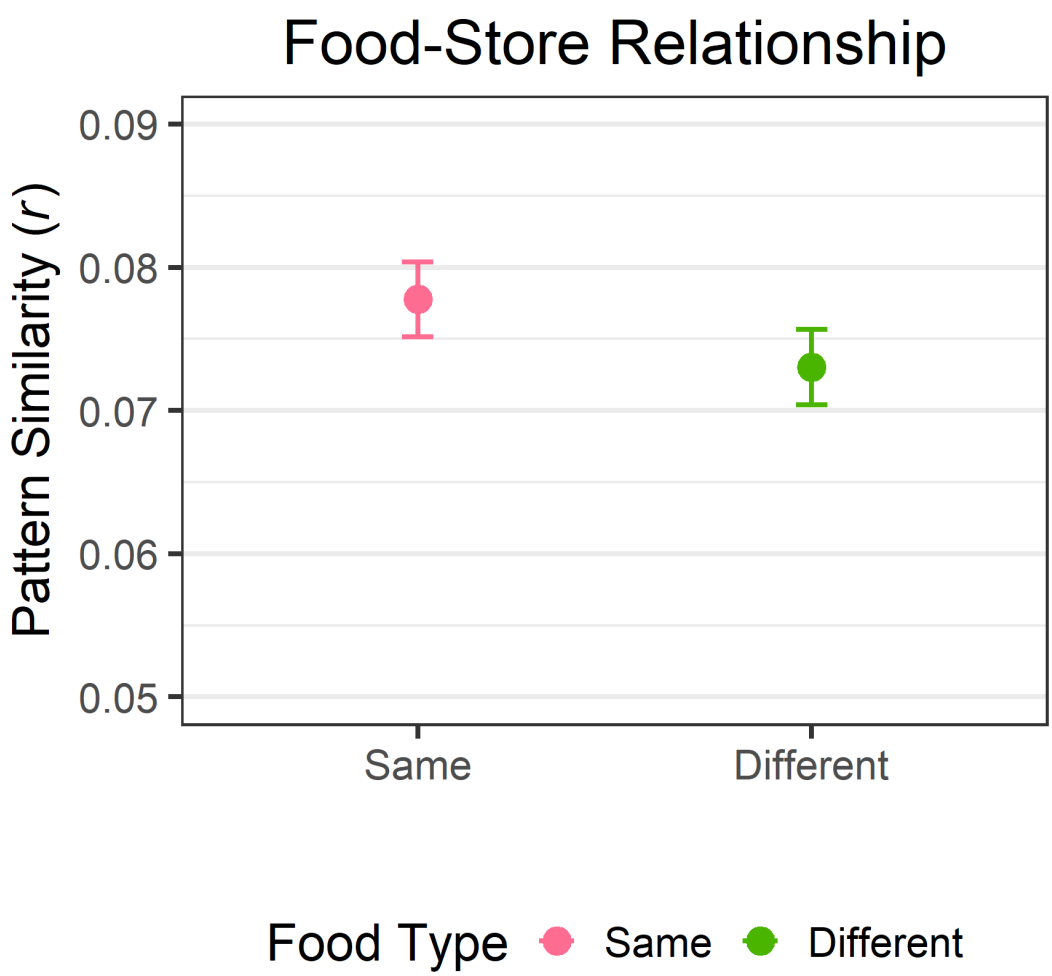

Figure 3. Hippocampal representations reflect decision relevant information. Pattern similarity values for pairs of trials that had same (pink) or different (green) food-store relationship. Trial pairs that shared the same food-store relationship (i.e., CD/CD or CI/CI) had higher similarity than pairs with different relationship (i.e., CD/CI).

Representation of Context Similarity in Hippocampus

Thus far, our results suggest that, during decision-making, the hippocampus represents information that is relevant to the decision that will be made. We next tested whether hippocampal activity patterns also carry information about the contexts in which the decisions were made. As shown in Figure 2, Stores A and B have the same customer preferences for all of the food items, and the same is also true of Stores C and D. There are at least

two ways in which the hippocampus might represent these contexts. One possibility is that the hippocampus generalizes across similar contexts. That is, overlap between the food-preference mappings between stores could result in overlapping hippocampal patterns. Another possibility is that the hippocampus might exaggerate differences between two stores that have the same customer preferences for all eight foods, creating dissimilar patterns between the two stores.

To investigate how context similarity is represented in the hippocampus, we computed neural PS between trial pairs that shared similar contexts (same context-food preference: e.g., the apple in Store A and the muffin in Store B) and trial pairs that *did not* share similar contexts (different context-food preference: e.g., the apple in Store A and the carrot in Store C). Store context has a stronger predictive relationship with customer preferences for CD foods but not CI foods, therefore it is possible that similarity between contexts might have different effects depending on the food type. Therefore, we included Food Type (context-determined vs. context-invariant) as a factor in our analysis.

A repeated measures ANOVA on the PS values with factors for Food Type (CD vs. CI), and Context Similarity (similar vs. different) revealed a main effect of context similarity [$F(1,21) = 6.04$, $p = .02$], such that PS was lower for trial pairs with similar contexts than for pairs that had different contexts (Figure 4A). There were no other main effects and no interactions between the factors (smallest $p = .12$). This suggests that the hippocampus hyper-

differentiates similar contexts, as compared with different contexts, regardless of the food-store relationship that was evaluated on a given trial.

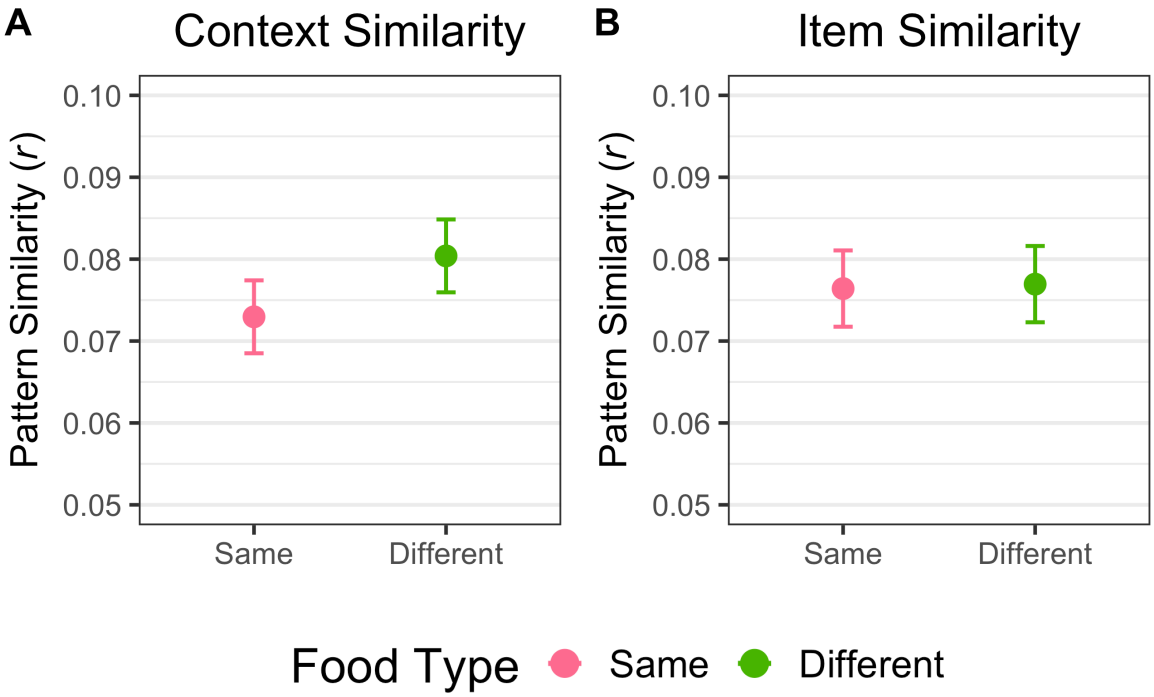

Figure 4. **The Hippocampus hyperdifferentiates between similar contexts.** A) Pattern similarity across pairs of trials that had similar vs. different contexts in the hippocampus. B) Pattern similarity across pairs of trials that had similar vs. different item types in the hippocampus.

Hippocampal activity patterns do not carry significant information about Item Similarity

We next tested whether hippocampal activity patterns carry information generalized across food items. The task was designed such that pairs of food items had identical distribution of outcomes across the four store contexts. As shown in Figure 2, the apple and the muffin had the same customer preferences in each store context, and the same is also true of the carrot and the bread. Likewise, the donut and the onion were generally liked,

and the banana and the cookie were generally disliked. If the shared outcomes across items are represented in the hippocampus, we might expect hippocampal pattern similarity values to be higher for foods with similar outcomes (i.e., similar customer preferences) than for foods with different outcomes (i.e., different customer preferences).

To test for pattern similarity effects driven by item-outcome similarity, neural PS values were calculated for trial pairs that shared the same outcome probability of being liked/disliked (e.g., apple and muffin, banana and cookie, etc.) and trial pairs that had different probability of being liked/disliked (e.g., apple and carrot, donut and banana). We computed PS for CD and CI foods separately. We conducted a repeated measures ANOVA with factors Item Similarity (similar vs. different), and Food Type (CD vs. CI). Results, summarized in Figure 4B, revealed no significant effect of item similarity on hippocampal PS [main effect of item similarity: $F(1,21) = 0.03$, $p = 0.87$]. There were no other main effects of or interactions (all p's bigger than .10).

Orbitofrontal cortex also represents task relevant information but not context similarity

Orbitofrontal cortex (OFC) has been implicated in many studies as an important region in goal-directed decision making (see reviews by Rushworth, Noonan, Boorman, Walton, & Behrens, 2011; Wallis, 2007) and in representing the current task state in a cognitive map of the task space for both learning and decision making (Schuck, Cai, Wilson, & Niv, 2016; Wilson et al., 2014). The aim of this study was to examine hippocampal patterns for

different information types related to decisions. However, given the recent literature implicating the OFC in representing the one's current position in a task space (Wilson et al., 2014), it is important to compare representations of task-relevant information between these regions. Therefore, we employed the same type of analyses presented above, this time within the OFC and examined a) food-store relationship, b) context similarity, c) item similarity. The detailed results from these analyses can be found in Supplemental Material. In brief, we found that food-store relationship was also represented in OFC such that pattern similarity values were higher for same relationship trial pairs compared to the different ones as shown in Figure 5 (main effect of food-store relationship: $F(1,21) = 5.84$, $p = .02$, same (CD/CD and CI/CI) > different (CD/CI))). Additionally, OFC activity patterns were also affected by the context similarity such that pattern similarity was also higher for OFC for similar contexts, but this effect was only apparent for CI foods and not for CD foods. Beyond this effect, OFC activity patterns were not sensitive to item similarity (smallest $p = .53$).

Figure 5. OFC represents task structure. Pattern similarity values for trials that had same (pink) or different (green) food-store relationship in different panels for Lateral and Medial parts of OFC. Trial pairs that shared the same food-store relationship (i.e., CD/CD or CI/CI) had higher similarity than pairs with different relationship (i.e., CD/CI) for both OFC parts.

DISCUSSION

The goal of the present study was to investigate how the hippocampus represents decision relevant information during memory guided decision making. In our experiment, participants needed to integrate information about items (i.e., foods), contexts (i.e., stores), and task structure (context-determined vs. context invariant probabilistic) that could help making accurate decisions. Our key findings can effectively be summarized as: Across contexts, activity patterns in the hippocampus reflected the task structure. Decisions for which the context predicted the customer preference were

grouped together and these trials were differentiated from decisions for which the context was irrelevant. The second key finding is that activity patterns depended on the store context similarity. PS during decisions made in stores that had the same food preferences was *lower* than PS for decisions made in stores that predicted different food preferences. Interestingly, we saw no evidence that the hippocampus was sensitive to item-outcome similarity.

Representation of decision-relevant information in the hippocampus

As noted earlier, one of the major questions guiding this study was how decision-relevant information is represented by the hippocampus. Many researchers have endorsed the idea that the hippocampus uses a map-like coding strategy to generalize across past experiences (O'Keefe & Nadel, 1978). A related idea is that the hippocampus might form schemas that provide a scaffold for learning and integrating new information.

Schemas (Bartlett, 1932) and cognitive maps (Tolman, 1948) are theoretical constructs that were proposed to explain how people can capture systematic regularities across experiences in a manner that organizes knowledge about individual elements. For a schema or map to be useful, however, it should enable one to generalize across some experiences. It could be very useful to generalize across contexts that are similar. For instance, you might develop a schema based on repeated visits to a particular Italian restaurant. This schema would provide a scaffold for integrating overlapping information, such as the use of parmesan cheese on different menu items. The schema could also be used to make inferences and predictions when you

dine at a new Italian restaurant in a new location. Given considerable evidence for the role of the hippocampus in coding of spatial locations (Moser & Moser, 2008; O'Keefe & Nadel, 1978) and contexts (Maren, Phan, & Liberzon, 2013), is reasonable to think that the hippocampus might support such a representation.

Recent work has also supported the idea that the hippocampus maps experiences that in a way that is optimized towards goal-directed behavior. For instance, in mice trained to do an auditory discrimination task, hippocampal neurons represented specific sound frequencies as well as other task-relevant information, but responses to auditory stimuli were only evident when the animal was performing the task (Aronov et al., 2017). Using fMRI, Theves et al. (2019) found that activity in the human hippocampus reflected distances in an abstract space that involved characteristics of a visual stimulus, and Tavares et al. (2015) found that hippocampal activity reflected information about social relationships in a task that required tracking of social relationships over time. These findings are consistent with the idea that, over the course of experience, hippocampus comes to represent the structural aspects of tasks as well as environments, in the service of goal-directed behavior (Behrens et al., 2018).

The present results add to this idea by suggesting that the hippocampus can generalize across experiences in different contexts according to the structure of the task. Over the course of training, participants learned that some food preferences could be predicted perfectly from the store context, whereas others were probabilistic and independent of

the context. Hippocampal activity patterns reflected this fundamental aspect of the task structure. The results suggest that, across different store contexts, the hippocampus assigned similar representations to context-dependent, deterministic decisions, and that these experiences were differentiated from representations of probabilistic, context-independent decisions. The design of our experiment did not permit us to definitively conclude whether the hippocampus represents context relevance per se, as this cannot be disentangled from the uncertainty of events, or the interaction of these variables. These factors could be separately tested in future studies, but the key point of the present study is that hippocampal activity patterns carried this task-relevant information in a manner that generalized across different contexts.

Our findings align well with results from a recent study that examined representation of goal-relevant information in the monkey hippocampus (Baraduc, Duhamel, & Wirth, 2019). In this study, monkeys were trained to find a hidden reward in a virtual maze, and the reward location was systematically related to visual landmarks in the maze. After learning the reward maps in one maze, monkeys were able to transfer their knowledge to a novel maze that had different landmarks but similar reward maps. Over time, a subset of hippocampal neurons ("schema cells") developed a firing map integrating goal-relevant information learned in one maze to the novel maze. The findings of Baraduc et al. (2019) converge with results from the present study to suggest that the hippocampus is able to capture goal-

relevant information in a manner that can generalize across different contexts.

The hippocampus hyperdifferentiates between similar contexts

Our task was also set up in such a way that participants could potentially form a schema (McKenzie et al., 2014) or cognitive map (Behrens et al., 2018) that generalized across similar contexts (see food-preference matrix in Figure 2). If the hippocampus indeed incorporates multiple experiences into a schema-like representation, we would have expected hippocampal pattern similarity to be higher across stores that predicted the same food outcomes (e.g., Stores A and B) than across stores that predicted opposite food outcomes (e.g., Stores B and D). Surprisingly, we found the opposite result -- contexts associated with identical preferences were more differentiated from one another than were contexts associated with opposing preferences. In other words, our results are consistent with the idea that contextual information is represented by the hippocampus, but contrary to what might be expected, similar contexts were hyper-differentiated from one another.

Although this result is counterintuitive, it fits with a growing body of evidence from fMRI studies showing that hippocampal representations of similar experiences are differentiated over the course of learning (Chanales, Oza, Favila, & Kuhl, 2017; Favila, Chanales, & Kuhl, 2016; Hulbert & Norman, 2015). For instance, Chanales et al. (2017) showed that when two spatial routes have overlapping segments, hippocampal representations of these

overlapping segments diverged with learning, and they became less similar compared to two non-overlapping routes. Another study found a learning-induced reduction in hippocampal similarity between representations with overlapping information even when the events predicted the same outcome (Favila et al., 2016). Their results suggested that pairing multiple stimuli with a common associate might have increased the probability of simultaneous activation of multiple stimuli when the common associate was presented.

Ritvo et al. (2019) proposed that overlapping experiences might become differentiated in the hippocampus through an unsupervised learning mechanism. In their framework, there can be different consequences of representational overlap depending on how learning takes place. They proposed that if a retrieval cue strongly activates two overlapping memory representations, then these representations can become integrated, as one might expect of Hebbian learning. If, however, one representation is strongly activated, and another is moderately activated, then the overlapping elements will be weakened such that differences between the two representations will become exaggerated. In our study, it is possible that, during learning, presentation of an item in one store could have activated memories for prior decisions made for the same item in other stores. For instance, discovering that an apple was liked in Store A might moderately activate a prior memory of seeing a liked apple in Store B, whereas memories of apples in stores C and D might only be weakly activated. Moderate activation of store B could start a differentiation process such that hippocampal representations of stores A and B would be pushed apart. Because participants were scanned one day

after initial learning, we cannot test whether this occurred, but this could be easily tested in a future study focused on the learning phase in this paradigm.

Representation of decision relevant information in OFC

In addition to the hippocampus, researchers have recently proposed that the OFC might also form cognitive maps that are used for decision making. Consistent with this idea, several findings suggest that the OFC represents the space of information that is used to make decisions (Farovik et al., 2015; Schuck et al., 2016). Farovik et al. (2015), for instance, examined how decision relevant task dimensions were represented in OFC of rats as they performed a context-guided object discrimination task. Results showed that OFC neurons sharply differentiated between events associated with reward outcomes and those that were not. Within each category of rewarded and nonrewarded outcomes, hippocampal neurons generalized across encounters with the same object in the same context.

We did not find evidence to support the idea that the OFC formed generalized representations of foods based on their probability of preference (item similarity) and instead, we found that OFC carried information about whether participants were making CD or CI decisions. There are several important differences between the study of Farovik et al. (2015) and the present study. Farovik et al. (2015) provided explicit rewards for correct performance, whereas we examined activity during trials in which participants judged customer preferences, in which participants were not presented with any feedback about the accuracy of their judgments. Another key difference

is that, in Farovik et al. (2015), the correct decision was entirely dependent on the context, whereas, in our study, context only was relevant for CD trials and not for CI trials. Putting these factors together, it is likely that our task-oriented participants to focus more on task structure, rather than item specific preferences per se.

Medial portions of OFC receive direct inputs from the hippocampus, and available evidence suggests that the hippocampus might be a source of context information for the OFC (Eichenbaum, 2017; Navawongse & Eichenbaum, 2013). Wikenheiser and Schoenbaum (2016) suggested that it is possible hippocampus and OFC encode variables in parallel but interact with each other during making of cognitive maps, a unified representation of an environment to guide future behavior. These cognitive maps consist of various information from the environment such as cues, behaviors, and their outcomes that are used to support goal-directed decision making. These findings collectively suggest that hippocampus and OFC play a crucial role in memory-guided decision-making by representing information about relevance of contextual information to future outcomes.

General Conclusions

In summary, the results suggest that hippocampus represents information related to experiences in ways that are both specific and integrative. On one hand, it appears to separate experiences in different contexts, but it also generalizes across experiences in different contexts depending on information that is relevant to a decision. This conclusion fits within a larger body of literature showing that goal- and decision-relevant

information is extensively represented in the hippocampus (Doll et al., 2015; Mizumori & Tryon, 2015; Palombo, Keane, & Verfaellie, 2015). Given that many of our experiences are encoded during temporally-extended, goal-directed behavior, these findings highlight the promise of integrative theories of hippocampal function to explain episodic memory and the use of past experiences in the service of decision-making (Gershman & Daw, 2017; Wang, Cohen, & Voss, 2015; Wikenheiser & Schoenbaum, 2016).

MATERIALS AND METHODS

Participants

Thirty-five healthy young adults (female=20; mean age = 22. 18 years, *SD* = 3.85 years) without any neurological or psychiatric disorders were recruited from the University of California, Davis Psychology Department subject pool. All thirty-five participants completed the behavioral learning session, however, eight participants performed below the set criteria for learning (60 % performance) and did not participate in the second session. One participant did not complete the second session due to a technical issue and another participant who achieved criterion level performance was unable to take part in the second session. The remaining 25 participants completed the second MRI scanning session which included a learning, decision, and a choice phase. Participants with head motion greater than 3 mm from origin (N=3) were excluded from the analysis, leaving 22 participants (female = 12) who completed both sessions and were included in the analyses reported here. Participants were compensated with either a check or an amazon gift

card. All procedures were approved by the University of California, Davis Institutional Review Board.

Stimuli and materials

Stimuli consisted of 8 food categories and 4 unique grocery store scenes. Grocery store images were the same across all participants. Participants were presented with one food item from each of the 8 categories (apple, muffin, carrot, bread, donut, onion, banana, cookie), which was randomly selected from a pool of 16 different exemplars of that food category. All participants, therefore, saw an exemplar from each of the 8 food categories, however the exact food images were different (e.g., every participant saw a banana but not every participant saw the same banana picture). Additionally, food items were randomly assigned to either "like" or "dislike" preferences in each of the four stores. This food-preference mapping was consistent within each participant but randomized across participants.

Procedure

The experiment consisted of two sessions that took place across two consecutive days (see Figure 1 and Figure S1). During session 1, participants completed the learning task outside of the scanner, where they learned which foods were preferred in which stores through repeated trials of cycles of decisions and feedback. Participants returned the next day to complete session 2 and were scanned while they completed two additional runs of the learning phase (to ensure that the associations were well learned), two runs of the decision task, and four runs of a choice task. The fMRI analyses

presented in this paper focused on the decision phase which included participants deciding whether customers preferred presented foods in a given store or not based on previous learning (See Supplemental Material for more details on the other phases).

During the decision phase, participants completed trials where a grocery store appeared on the screen followed by a food item, which was overlaid on top of the store. Using what they had learned the previous day, participants had 2 seconds to predict what the customer reactions were to that food item in that store by selecting either "like" or "dislike" using the j and k keys. Participants were not given any feedback about their answer (they did receive feedback in the learning trials). Participants completed two runs of the decision phase, where each run consisted of 8 mini-blocks. Each mini-block had 8 trials (so every run had 64 trials). During the mini-blocks, the background image of the store remained on the screen for the entire mini-block and changed to a new store after each mini-block (e.g. participants visited all four stores twice in every full run, each store was a mini-block). Participants made like/dislike judgments on all 8 foods in each mini-block. The time between each mini-block was jittered (mean = 4 seconds). Each mini-block started with the presentation of the store for 6 seconds, followed by the eight foods. The order of the foods was randomized within each mini-block. The mini-blocks following each other always presented a different store.

"Like" probabilities. The food item and store context pairs varied in their "like" probability so that there were context-determined (CD) and context-invariant (CI) foods. Foods with "like" probability of 1 or 0 depending on the store they

were presented in are considered CD, while the foods whose "like" probability are .75 or .25 across all stores are CI. There are also two stores that share the same "like" preference across all 8 foods, so these are considered as "similar contexts." The food preference probability matrix is shown in Figure 2. The mean like preference of all foods presented within a store was .5 such that no store had more favorable preferences (probability of a row in the matrix) than another store. CI foods had the same "like" preference probability across all stores and the average probability of them was either .75 or .25 (mean probability of a CI food column). CD foods on the other hand were 100% liked in half of the stores and 0% liked in the other half of the stores. In order to predict the like preference of the CD foods, participants had to understand the relationship between the store and food preference. The overall average like probability of CD foods was .5 (mean probability of a CD food column). This probability did not provide any information about the likeability of that food across all the stores.

Image Acquisition and Preprocessing

Scanning was performed on a Siemens Skyra 3T scanner system with a 32-channel head coil at the UC Davis Facility for Integrative Neuroscience. High-resolution T1-weighted structural images were acquired using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) pulse sequence (1 mm^3 voxels; matrix size=256 x 256; 208 slices). An additional T1-weighted image with only sagittal oriented slices was used for aligning the field of view box in the subsequent functional scans (i.e., the box was adjusted for each participant to make sure that the temporal lobes were not

cut off and as much of the brain as possible was in the box). Functional images were acquired over eight runs using a whole-brain multiband gradient echo planar imaging (EPI) sequence (TR = 1220 ms; TE = 24 ms; flip angle = 67°; FOV = 192 mm; multi-band factor = 2; 38 interleaved slices; voxel size = 3 mm^3; image matrix = 64 × 64).

Processing of the fMRI data was carried out using FEAT (FMRI Expert Analysis Tool) Version 6.00, part of FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). The two functional scans from the decision phase underwent the following preprocessing steps: (1) Skull stripping of the anatomical images was carried out using FSL's BET function (Smith, 2002). (2) Motion correction was carried out using FSL's MCFLIRT function (Jenkinson, Bannister, Brady, & Smith, 2002), whereby volumes in the functional scans were aligned to the center volume of each scan with rigid body registration. (3) Spatial smoothing was carried out using a Gaussian kernel of 6.0 mm FWHM. (4) Grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor. (5) A high pass filter cut off was set at 100 and a high pass temporal filter was applied to remove low frequency noise from the signal (Gaussian-weighted least-squares straight line fitting, with sigma=50.0s). (6) Functional scans were coregistered to each subject's native space anatomical scan using the boundary-based registration (BBR) (Jenkinson et al., 2002; Jenkinson & Smith, 2001) cost function with FLIRT.

ROI definition and masks

Bilateral ROI masks for the hippocampus were manually segmented on the high-resolution anatomical using the guidelines supplied by Frankó et al. (2014). Masks for the OFC were labeled with FreeSurfer (Desikan et al., 2006; Fischl, 2004).

All ROIs were co-registered to the example functional image of the participant in FSL by using the FLIRT function and applxfm option applying the same parameters that were used to co-register their anatomical image with the functional images and pattern similarity analyses were done with the co-registered ROI masks.

fMRI data analysis

Pattern similarity analysis

Multivoxel pattern similarity analyses was performed on the fMRI data from the decision task runs. The analyses were conducted on unsmoothed functional images in the native space and the EPI timeseries underwent motion correction and high-pass filtering (0.01 Hz) in FMRIB's Software Library (FSL).

For each trial, a single beta image was estimated by single trial models for event-related blood oxygenation level-dependent (BOLD) signal change, controlling for signal change due to all other trials and motion artifact, using ordinary least squares regression, resulting in 128 single-trial beta images (Mumford, Turner, Ashby, & Poldrack, 2012). Parameter estimates for each

trial were computed using a general linear model, with the first regressor as a stick function placed at the onset of each trial and a second regressor containing all the other trials.

Single-trial beta images from run 2 were coregistered with single-trial beta images from run 1 using FSL's FLIRT linear registration software (6 degrees of freedom). Coregistered single-trial beta images with atypically high mean absolute z-score (based on the distribution of beta estimates for each grey matter voxel across all trials) were excluded from further analysis. Based on a mean absolute z threshold of 1.5, between 0 and 10 trials were excluded per subject with a median of 4. Beta images went through a second visual inspection to make sure all the deviant trials were excluded. This noise trial exclusion procedure is adopted from the previous pattern similarity studies in the lab (Libby, Reagh, Bouffard, Ragland, & Ranganath, 2019).

The representational similarity analyses were then conducted using the RSA toolbox by Nili et al.(2014). For each region of interest (HP and OFC), all trial patterns were correlated with each other using Pearson's r resulting in a 128 * 128 pattern similarity matrix. This matrix was then masked with binary masks of pairwise combinations of trials of interest that were created specifically for each analysis.

Our PS analysis included trial pairs that consisted of trials from the same run (within run analysis). We excluded trial pairs if the two trials: a) shared the same store as background image, b) had the same food image, c) if the trials were proximally close in time (in order to control for adjacent trial similarity – 8 trial distance was required), or if the trials had global signal values that were outliers based on the global average of absolute standardized values calculated for each within-brain voxel. These exclusions assured that the trial pairs from the same mini-block were not correlated as well. After the exclusion process, the remaining trial pairs were correlated with each other that resulted in a number of correlation values for each ROI, participant, and condition of interest. These values were then averaged across trial pairs and resulted in one value for each participant per ROI per condition. These values were then passed to repeated measures of ANOVAs for each ROI.

We performed an additional PS analyses to examine whether the similarity between trials that required same motor response was a source of pattern similarity in hippocampus. We did not find any evidence for the motor response similarity resulting in higher pattern similarity in hippocampus. The

detailed results from these analyses are presented in the Supplemental Material Pattern Similarity Results section.

References

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, *543*(7647), 719–722. https://doi.org/10.1038/nature21692

Baraduc, P., Duhamel, J.-R., & Wirth, S. (2019). Schema cells in the macaque hippocampus. *Science*, *363*(6427), 635–639. https://doi.org/10.1126/science.aav5404

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Bornstein, A. M., Khaw, M. W., Shohamy, D., & Daw, N. D. (2017). Reminders of past choices bias decisions for reward in humans. *Nature Communications*, *8*, 15958.

Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature Neuroscience*, *20*(7), 997–1003. https://doi.org/10.1038/nn.4573

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology*, *27*(15), 2307-2317.e5. https://doi.org/10.1016/j.cub.2017.06.057

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., … Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, *31*(3), 968–980. https://doi.org/10.1016/j.neuroimage.2006.01.021

Doll, B. B., Shohamy, D., & Daw, N. D. (2015). Multiple memory systems as substrates for multiple decision systems. *Neurobiology of Learning and Memory*, *117*, 4–13. https://doi.org/10.1016/j.nlm.2014.04.014

Duncan, K., Doll, B. B., Daw, N. D., & Shohamy, D. (2018). More Than the Sum of Its Parts: A Role for the Hippocampus in Configural Reinforcement Learning. *Neuron*, *98*(3), 645-657.e6. https://doi.org/10.1016/j.neuron.2018.03.042

Eichenbaum, H. (2017). Time (and space) in the hippocampus. *Current Opinion in Behavioral Sciences*, *17*, 65–70. https://doi.org/10.1016/j.cobeha.2017.06.010

Ekstrom, A. D., & Ranganath, C. (2018). Space, time, and episodic memory: The hippocampus is all over the cognitive map: EKSTROM and RANGANATH. *Hippocampus*, *28*(9), 680–687. https://doi.org/10.1002/hipo.22750

Farovik, A., Place, R. J., McKenzie, S., Porter, B., Munro, C. E., & Eichenbaum, H. (2015). Orbitofrontal Cortex Encodes Memories within Value-Based Schemas and Represents Contexts That Guide Memory Retrieval. *Journal of Neuroscience*, *35*(21), 8333–8344. https://doi.org/10.1523/JNEUROSCI.0134-15.2015

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*(1). https://doi.org/10.1038/ncomms11066

Fischl, B. (2004). Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex*, *14*(1), 11–22. https://doi.org/10.1093/cercor/bhg087

Frankó, E., Insausti, A. M., Artacho-Pérula, E., Insausti, R., & Chavoix, C. (2014). Identification of the human medial temporal lobe regions on magnetic resonance images: Human Medial Temporal Lobe Landmarks. *Human Brain Mapping*, *35*(1), 248–256. https://doi.org/10.1002/hbm.22170

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*(1), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The Role of the Ventromedial Prefrontal Cortex in Abstract State-Based Inference during Decision Making in Humans. *Journal of Neuroscience*, *26*(32), 8360–8367. https://doi.org/10.1523/JNEUROSCI.1010-06.2006

Hulbert, J. C., & Norman, K. A. (2015). Neural Differentiation Tracks Improved Recall of Competing Memories Following Interleaved Study and Retrieval Practice. *Cerebral Cortex*, *25*(10), 3994–4008. https://doi.org/10.1093/cercor/bhu284

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Kaplan, R., Schuck, N. W., & Doeller, C. F. (2017). The Role of Mental Maps in Decision-Making. *Trends in Neurosciences*, *40*(5), 256–259. https://doi.org/10.1016/j.tins.2017.03.002

Lee, S. W., Shimojo, S., & O'Doherty, J. P. (2014). Neural Computations Underlying Arbitration between Model-Based and Model-free Learning. *Neuron*, *81*(3), 687–699. https://doi.org/10.1016/j.neuron.2013.11.028

Libby, L. A., Reagh, Z. M., Bouffard, N. R., Ragland, J. D., & Ranganath, C. (2019). The Hippocampus Generalizes across Memories that Share Item and Context Information. *Journal of Cognitive Neuroscience*, *31*(1), 24–35. https://doi.org/10.1162/jocn_a_01345

Mack, M. L., Love, B. C., & Preston, A. R. (2018). Building concepts one episode at a time: The hippocampus and concept formation. *Neuroscience Letters*, *680*, 31–38. https://doi.org/10.1016/j.neulet.2017.07.061

Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear conditioning, extinction and psychopathology. *Nature Reviews Neuroscience*, *14*(6), 417–428. https://doi.org/10.1038/nrn3492

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal Representation of Related and

Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas. *Neuron*, *83*(1), 202–215. https://doi.org/10.1016/j.neuron.2014.05.019

McKenzie, S., Keene, C. S., Farovik, A., Bladon, J., Place, R., Komorowski, R., & Eichenbaum, H. (2016). Representation of memories in the cortical–hippocampal system: Results from the application of population similarity analyses. *Neurobiology of Learning and Memory*, *134*, 178–191. https://doi.org/10.1016/j.nlm.2015.12.008

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*, *20*(9), 1269–1276. https://doi.org/10.1038/nn.4613

Mizumori, S. J. Y., & Tryon, V. L. (2015). Integrative hippocampal and decision-making neurocircuitry during goal-relevant predictions and encoding. In *Progress in Brain Research* (Vol. 219, pp. 217–242). Retrieved from https://linkinghub.elsevier.com/retrieve/pii/S0079612315000448

Morton, N. W., Sherrill, K. R., & Preston, A. R. (2017). Memory integration constructs maps of space, time, and concepts. *Current Opinion in Behavioral Sciences*, *17*, 161–168. https://doi.org/10.1016/j.cobeha.2017.08.007

Moser, E. I., & Moser, M.-B. (2008). A metric for space. *Hippocampus*, *18*(12), 1142–1156. https://doi.org/10.1002/hipo.20483

Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel

pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643. https://doi.org/10.1016/j.neuroimage.2011.08.076

Navawongse, R., & Eichenbaum, H. (2013). Distinct Pathways for Rule-Based Retrieval and Spatial Mapping of Memory Representations in Hippocampal Neurons. *Journal of Neuroscience*, *33*(3), 1002–1013. https://doi.org/10.1523/JNEUROSCI.3891-12.2013

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLoS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford : New York: Clarendon Press ; Oxford University Press.

Palombo, D. J., Keane, M. M., & Verfaellie, M. (2015). The medial temporal lobes are critical for reward-based decision making under conditions that promote episodic future thinking: DECISION MAKING IN AMNESIA. *Hippocampus*, *25*(3), 345–353. https://doi.org/10.1002/hipo.22376

Preston, A. R., & Eichenbaum, H. (2013). Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, *23*(17), R764–R773. https://doi.org/10.1016/j.cub.2013.05.041

Ritvo, V. J. H., Turk-Browne, N. B., & Norman, K. A. (2019). Nonmonotonic Plasticity: How Memory Retrieval Drives Learning. *Trends in Cognitive Sciences*, *23*(9), 726–742. https://doi.org/10.1016/j.tics.2019.06.007

Rushworth, M. F. S., Noonan, M. P., Boorman, E. D., Walton, M. E., & Behrens, T. E. (2011). Frontal Cortex and Reward-Guided Learning and Decision-Making. *Neuron*, *70*(6), 1054–1069. https://doi.org/10.1016/j.neuron.2011.05.014

Santoro, A., Frankland, P. W., & Richards, B. A. (2016). Memory Transformation Enhances Reinforcement Learning in Dynamic Environments. *Journal of Neuroscience*, *36*(48), 12228–12242. https://doi.org/10.1523/JNEUROSCI.0763-16.2016

Schiller, D., Eichenbaum, H., Buffalo, E. A., Davachi, L., Foster, D. J., Leutgeb, S., & Ranganath, C. (2015). Memory and Space: Towards an Understanding of the Cognitive Map. *Journal of Neuroscience*, *35*(41), 13904–13911. https://doi.org/10.1523/JNEUROSCI.2618-15.2015

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Tavares, R. M., Mendelsohn, A., Grossman, Y., Williams, C. H., Shapiro, M., Trope, Y., & Schiller, D. (2015). A Map for Social Navigation in the Human Brain. *Neuron*, *87*(1), 231–243. https://doi.org/10.1016/j.neuron.2015.06.011

Theves, S., Fernandez, G., & Doeller, C. F. (2019). The Hippocampus Encodes Distances in Multidimensional Feature Space. *Current Biology*, *29*(7), 1226-1231.e3. https://doi.org/10.1016/j.cub.2019.02.035

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*(4), 189–208. https://doi.org/10.1037/h0061626

Wallis, J. D. (2007). Orbitofrontal Cortex and Its Contribution to Decision-Making. *Annual Review of Neuroscience*, *30*(1), 31–56. https://doi.org/10.1146/annurev.neuro.30.051606.094334

Wang, J. X., Cohen, N. J., & Voss, J. L. (2015). Covert rapid action-memory simulation (CRAMS): A hypothesis of hippocampal–prefrontal interactions for adaptive behavior. *Neurobiology of Learning and Memory*, *117*, 22–33. https://doi.org/10.1016/j.nlm.2014.04.003

Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: Cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience*, *17*(8), 513–523. https://doi.org/10.1038/nrn.2016.56

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, *81*(2), 267–279. https://doi.org/10.1016/j.neuron.2013.11.005