CONCLUSION

I have presented three studies that explore aspects of role of the hippocampal/episodic memory system in model-based control of behavior.

In Chapter 1, we studied the neural representations of temporal order contingencies — links between successively presented stimuli in a sequential learning task. Use of these contingencies to seek novel rewards is a key signature of model-based planning. We identified two such representations, which appear to be learned in parallel, both used to effect simple response behavior in a serial reaction time task. Neuroimaging analysis identified these representations as each uniquely reflected in activity in two separate brain systems, hippocampus and striatum, each known to support different sorts of associative learning.

Building on the results of this study, in Chapter 2 we examined how these representations were used in a goal-directed decision task. Successful performance on the choices in this task required the use of learned stimulus-stimulus contingencies. We showed that only one of these representations, the one associated with the hippocampal system, was necessary to explain the pattern of choices in this task. Further neuroimaging results are consistent with a model in which hippocampal system memories are accessed as 'samples' of past experience, used to estimate values of current choice options. Specifically, we saw that cortical areas that respond preferentially to visual presentations of certain categories of images — faces, houses, etc — were also active in response to the complexity of planning ahead via the states represented by these images, in both simple responses and goal-directed choices. Again, activity in these areas was computationally

linked to the associations learned by the hippocampal system.

This model of reinstatement of past experience in support of decisions suggested a novel behavioral manipulation, which we tested in the final study. Specifically, the idea that sampling from an episodic cache precedes choice opens the question of whether favoring particular episodes in that cache might bias choices in a way not captured by traditional reinforcement learning models. We modified a standard reinforcement learning task known as a "two-armed bandit", which asks participants to choose between two slot machines of varying reward rates. In our modification, the bandits returned trial-unique photographs ("tickets"), which individually identified the episodic event of each trial. We then showed that, when we later presented these ticket photographs as part of a choice-incidental memory probe interspersed among regular choice trials, subsequent choices are biased towards the bandit that was rewarded on the turns evoked by the photographs. This is congruent with the suggestion that the probed photographs evoke those past trial episodes, reminding participants of the reward they received from their choice on that option, and thus enhancing the likelihood that the reward — critically, a reward tied to a specific bandit — into their current value estimation process.

Together, these studies advance the notion that action selection is based on multiple factors that together comprise a context for decisions, of which the recent history of action reinforcement is just one component. Critically, in these tasks context is resolved using samples of episodic memories, of the kind strongly associated with the hippocampal memory system, therefore assigning to the hippocampus a key role in reward-guided decisions.

The process of sampling is a way to infer, stochastically, what the context is, and

thus what is the best action to take. Understanding model-based decision making as probabilistic inference accomplished by sampling from episodic memories opens up several neural, psychological, and computational questions. In this final section, I sketch the model of hippocampal control that unites these findings, and outline some of these questions and suggest further work that can build on the research program outlined in this dissertation.

Functions of the hippocampus in instrumental control

In the three studies presented here, I have described several observations of hippocampal activity linked to different instrumental behaviors. These behaviors — picture-identification responses driven by long-memory sequential expectations, deliberative choices using those same expectations to achieve unfamiliar rewards, and cue-guided recall of long-past choices and outcomes — may at first seem starkly different. I argue that these behaviors draw on a common mechanism — namely, goal-directed retrieval of episodes that are most likely to be relevant to the current action selection. That is, in each case, the hippocampal memory system supports behavior by drawing individual episodes of past experience that are selected based on their similarity to the current context. By default, this similarity is judged by temporal recency, with more recent episodes more likely to be drawn, at a ratio that decreases exponentially as experiences recede into the past. However, episodes from the far past have some small probability of being drawn, and impact behavior greatly when they do. When many trials are modeled using a simple associative learning algorithm of the form proposed by Rescorla and Wagner (1972) this "long tail" of probability is captured by a low value of the learning rate parameter.

This results in an apparently smooth effect on reaction time behavior, corresponding to that seen in Chapters 1 and 2. However, on a given trial, individual episodes may drive behavior, as in Chapter 3.

Next steps

In the introduction, I argued that the function of the hippocampus in model-based control is to serve as a store of episodic samples. I then reviewed experimental and theoretical work that situates this function within a broader architecture of multiple systems of behavioral control. I proposed that the way that episodic samples are stored adaptively links features of the task environment over time, space, or other contexts. This adaptive storage influences the way that samples are drawn at the time of decision — or in anticipation of a decision — and what features of their content are used by the value estimation system to propose optimal candidate actions.

An important feature of this control system is that it is continuously tuning behavior, even during rest; samples may *also* be drawn offline, to train representations of expected value, or to build more structured contingency representations (e.g., decision trees, or verbalizable rules) that can be used by other control systems — in e.g., prefrontal cortex — to more efficiently guide future decisions. In this view, the sampling mechanism is a critical component of action selection, as it informs decisions not just in early learning, but at every stage.

Much further work is needed to substantiate this model. Some of the more pertinent questions available to current methods are outlined here.

Which samples to draw?

Replay activity in place cells preceding choice has been observed by many labs (Ferbinteanu et al., 2003; Johnson and Redish, 2007; Singer et al., 2013), but only very recently has the replay of particular trajectories been directly linked to subsequent choices (Pfeiffer and Foster, 2013). Critically, however, while Pfeiffer and Foster demonstrate that reactivation of a particular trajectory guides choice on subsequent decisions, their results do not speak to the computational processes that may give rise to reactivating particular trajectories. There are many reasons why a particular episode might be of use to the current choice. Due to the computational infeasibility of a general solution for deciding the most relevant episodes to draw on (see *Introduction*), heuristics are likely used to select some past experiences rather than others. In Chapter 3, we explored one such heuristic, and manipulated it in a bottom-up fashion using associative cues. However there might be other factors and heuristics that make some episodes or cell assemblies more likely to be reinstated than others.

One particularly relevant application of sampling methods to action selection is embodied in the DYNA architecture (Sutton, 1991), in particular, the DYNA-2 version (Silver and Sutton, 2007). In DYNA, samples of past experience are used to train the value estimates that directly guide behavior — in DYNA terms, the "reactive policy". Critically, these samples are drawn during offline states, not necessarily to guide any particular behavior; Sutton describes this as "planning is 'trying things in your head'." DYNA-2 extends this approach by using a prediction error-like quantity ('regret' Auer et al., 2002) to guide which samples are drawn during planning. For the purposes of this discussion, the procedure can be thought of as ranking episodes according to the

prediction error that was computed when they were originally experienced. This feature embodies the suggestion that surprising outcomes need to be explored more often.

Do humans and animals draw samples based on their prediction error at the time of experience? The question remains unapproached, experimentally. There may be behavioral consequences of drawing samples in this fashion, or according to other metrics. Several authors have proposed that sampling might underlie idiosyncratic choice behavior (Stewart et al., 2006; Biele et al., 2009). An understanding of the factors that guide sampling could improve the predictive power of these models in situations of idiosyncratic choice behavior (Peters and Büchel, 2010). Therefore, a comparison of multiple sampling approaches could be of major import to the fields of neuro- and behavioral economics.

Critically, in adjudicating among competing hypotheses for experimental data, the order of samples drawn may be as instructive as the fact that they are drawn. Therefore, recording techniques with high temporal resolution (Hunt et al., 2012), in combination with content decoding approaches like those we use in Chapter 2, could be necessary for a full understanding of the mechanisms employed.

Effect of sampling on subsequent memory

If episodes are sampled to guide behavior, and represented in a manner to enhance their effective use when sampled, then their use in behavior should be reflected in the manner in which they are represented. We have already discussed some evidence for the effect of memory use on memory representation. Namely, the reactivation of cell assemblies by reverse replay (Foster and Wilson, 2006) engages Hebbian mechanisms that strengthen the cell assembly in a direction opposite to that in which it was experienced. This new

directionality thus improves the suitability of the assembly for later goal-directed behavior — when the representation of the goal state is reactivated, so too will be the members of the cell assembly that lead up to that goal state. This feature, critical to the adaptiveness of reverse replay (Foster and Wilson, 2006), may also underlie the usefulness of offline replay in consolidation and behavioral memory improvements (Carr et al., 2011).

Untested is the impact of this rewiring on other types of memory behaviors, such as episodic probes and source memory tests of the kind that are predominantly used in humans. A key question is the degree to which this adaptive rewiring extends beyond simple place-place associations. Importantly, if it is found to exist, then a number of questions open up about the extent of the consequences of that rewiring — on, e.g., other associations that may be disfavored, or new associations that may be bound.

For instance, it might be that repeated sampling of a cell assembly during planning for decisions should then improve the later strength of that memory. If participants are encouraged by task demands to sample a particular assembly more than others, or more often than otherwise, then that enhanced deliberation activity should lead to better subsequent memory for the assembly as a whole — and in particular in the direction in which the assembly was reactivated. Follow-on effects of the strengthening of one set of associates, favored over others, might have more interesting consequences for both memory and decisions, for instance by encouraging association with other contextual features present at the time of reactivation.

From episodes to structured models

A key question opened by the above studies is that of when model-based decisions rely on episodic memories, and when they rely on more structured representations traditionally associated with model-based control (such as decision trees or rules Daw et al., 2005). I proposed above that the transition between reliance on either type of memory corresponds to the consolidation of relevant episodes of experience into semanticized versions suitable for evaluation. This suggestion concords with experimental observations that resting periods — thought to promote consolidation — improve performance in a variety of instrumental tasks (Ego-Stengel and Wilson, 2010; Fischer et al., 2002; Albouy et al., 2006).

Though it has a strong theoretical history (McClelland et al., 1995), a direct relationship between semanticization and behavioral performance in model-based control has yet to be explored experimentally. One testable aspect of the hypothesis is the suggestion of a distinct effect of offline activity that results in successful structure learning, driven by replay from the hippocampal memory system, as opposed to offline activity that does not result in successful structure learning. This structure learning may correspond with new patterns of neural activity appearing during offline states, or with signatures of structural learning, such as *State Prediction Errors* (Gläscher et al., 2010) during offline rest. Specifically, it may be possible to observe reinstatement episodes over the course of rest after training in a structured environment, and assess whether these reinstatements result in new correlations arising among previously distinct cortical regions. These new "functional connectivity" patterns should then be predictive of specific behavioral improvements in post-rest testing.

Summary

The study of the contributions of episodic memory to decisions for reward is in its early stages. Whereas the study of decision neuroscience is relatively new, an extensive literature describes the architecture of episodic memory. However, fundamental questions remain. Because decisions operate on external quantities that are readily measurable, and produce behavior that is straightforward to quantify, computational modeling approaches have yielded enormous insight into the normative justifications for much choice behavior. But when applied to choices that rely on less directly observable inputs, these models have had less success.

The results presented here suggest new directions for applying the wealth of knowledge about the cognitive neuroscience of episodic memory to the study of idiosyncracies in decisions. I look forward to building on this research for some time to come.