Hippocampal contributions to value-guided foraging

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Andrew M. Wikenheiser

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

A. David Redish, adviser

July, 2014

Acknowledgements

Foremost, I would like to thank David Redish for having me as a student. Dave's laboratory is an engaging and exciting place to do science, and I am grateful to have been a part of that environment. The Redish lab has been populated by an outstanding group of scientists that I feel privileged to call colleagues: Seiichiro Amemiya, Cassandra Barrett, Yannick Breton, Evan Carter, John Ferguson, Anoopum Gupta, Jadin Jackson, Adam Johnson, Zeb Kurth-Nelson, Andy Papale, Nate Powell, Paul Regier, Brandy Schmidt, Nathan Schultheiss, Adam Steiner, Jeff Stott, Matt van der Meer, and Adam Vogel.

I thank my thesis committee for their help and insight throughout my graduate work: Geoff Ghose, David Stephens, and Mark Thomas.

The work in this thesis was supported by NIH grants T90-DK-070106, T32-GM-008471, T32-DA-007234, R01-MH-080318, and a University of Minnesota Doctoral Dissertation Fellowship.

Finally, I'd like to thank friends and family for their support. Kat deserves special mention for radiating calm in trying times.

Abstract

The work in this thesis tested the contribution of the hippocampus to value-guided decision making in rats. A novel paradigm for testing intertemporal decision making in rats with sensitivity to the topology of choices that animals might face in natural settings was developed and validated. Rats made sequences of accept or reject decisions between feeder sites that offered an equal amount of food after differing amounts of delay. Exponential and hyperbolic delay discounting models were inconsistent with rats' behavior on the task. However, rats performing the foraging task approached maximization of long-term food intake rates, consistent with behavioral ecology models of foraging decision making. Tetrode recordings were taken from the dorsal CA1 region of hippocampus as rats engaged in the foraging task. As rats ran trajectories between feeder locations, the spiking of place-selective hippocampal neurons was organized into temporally-structured sequences bounded by the ongoing theta frequency (6-12 Hz) local field potential oscillation. Spatial representations traced out by spiking sequences were related to animals' behavior on the task, with longer sequence representations occurring as rats ran longer trajectories through the environment, suggesting that the content of hippocampal representations depended on rats' current spatial goals. At the level of single place cells, goal-dependent modulation of sequence representations manifested as an earlier activation of place fields on long trajectories. Together, these data suggest that as behavior is executed, hippocampal ensemble representations multiplex both information about the animal's current location in space and information related to current goals.

Contents

Introduction to the thesis

This thesis is divided roughly into two parts. In the first portion, we review previous approaches to studying and modeling intertemporal choice (decisions made between delayed outcomes). We develop the rationale for a new paradigm to study this behavior in rats, and analyze rats' decisions on this novel task.

- **Chapter 1** provides a historical perspective on studies of intertemporal choice, and gives a detailed specification of mathematical models that predict or describe behavior in intertemporal choice settings.
- **Chapter 2** introduces the intertemporal foraging task, and demonstrates that it is an effective means of testing intertemporal choice in rats.
- **Chapter 3** compares rats' behavior on the foraging task to previously-discussed models of intertemporal choice, and builds a quantitative framework for modeling the influence of subjective, psychological factors on behavior.

In the second portion of the thesis, we assess how a brain region called the hippocampus contributes to value-guided decision making by analyzing neural recordings taken from rats performing the intertemporal foraging task.

- **Chapter 4** reviews the anatomy and physiology of the hippocampal formation in rodents, and describes in detail the sequential spiking sequence representations characteristic of hippocampal ensemble activity.
- **Chapter 5** presents the results of electrophysiology experiments in which hippocampal activity was measured as rats performed the foraging task.

List of Figures

Chapter 1

Intertemporal choice

> Lord, give me chastity and continence,
> but not just yet.
>
> –Augustine (354–430)
> *Confessions*

Intertemporal choice treats decisions between delayed outcomes. This broad definition encompasses a wide variety of choices that humans and other species encounter frequently; because our actions generally take some time to produce their consequences, a temporal element lurks within a large fraction of the decisions we make. Examples of intertemporal choices that humans face include:

- Spend money now, or invest money for the future?
- Take a moderately-paying job, or obtain advanced schooling and aim for a high-paying job later?
- Enjoy a cigarette now, or abstain, and enjoy a healthy life later?

These examples demonstrate the breadth of situations that contain intertemporal choice components. Implicit in all of these scenarios is a tension between indulging in a more immediately available option or holding out for a more desirable, delayed outcome. An enduring finding is that many species (including humans) weight immediate outcomes more strongly than delayed ones. The tendency to ascribe greater importance to the present is known as *immediacy bias*.

As Augustine's plea at the beginning of this chapter attests, the trade-off between immediate and delayed outcomes has long been recognized; nevertheless, understanding how and why immediacy bias arises, what influences its expression, and how it is overcome in cases when delayed outcomes are selected has occupied thinkers from a variety of domains for centuries.

In this chapter, we review influential perspectives on temporal preferences (i.e. the trade-off between outcomes associated with different delays). We begin with the economic approach, which arguably represents the first systematic treatment of intertemporal choice. Next we consider the psychology of intertemporal decision making, which builds on and extends the foundations established by economists. We then examine a behavioral ecological take on intertemporal choice, which fuses ideas from economics and psychology with an ethological perspective. Finally, we review algorithms for reinforcement learning developed by integrating psychological principles from learning theory with computer science and machine learning concepts.

1.1 Economic approaches to intertemporal choice

1.1.1 Early work

Because of the ubiquity of intertemporal choice scenarios, economists have a long-standing interest in understanding how temporal preferences affect patterns of spending and consumption, both at the level of individuals and on the larger scale of organizations, communities, and nations. Given the subsequent divergence of psychological and economic approaches to intertemporal choice, it is interesting to note that the earliest economic thinking on the topic was based on distinctly psychological considerations, albeit ones derived largely from introspection or incidental observations (Loewenstein, 1992).

Rae (1834; 1905) argued that preferences were biased towards the present by both the "excitement" that comes from receiving an outcome quickly and the "discomfort" of waiting for a delay to pass. He suggested that an individual's "propensity for self-restraint" could overcome these tendencies. Rae's ideas were developed further by Jevons (1871), who suggested that delayed outcomes were favored due to their "anticipatory

utility[1]," which influences behavior in the present, but arises from looking forward to the receipt of some future outcome. This theory introduces a critical role for imagining future outcomes in determining temporal preferences.[2] Senior (1836), on the other hand, posited that the delay preceding an outcome did not explicitly affect the perception of that outcome's value, but that the misery of enduring said delay is frequently too much to endure, favoring selection of the more immediate option.

The development of psychologically-motivated theories of intertemporal choice continued with the work of von Böhm-Bawerk (1890), who attributed temporal preferences to cognitive limitations rather than emotional or motivational factors. By von Böhm-Bawerk's reckoning, decision makers were not driven to favor immediate options by the experience of positive or negative emotions; he argued instead that a consistent underestimation of future wants and needs resulted in present-oriented decisions. Erroneous predictions about the future resulted either because our ability to accurately imagine future conditions was severely limited, or because the difficulty of doing so prevented people from even trying in many cases. This sort of reasoning was the basis of a more recent explanation of intertemporal preferences, which posits that immediately available, concrete options are assigned a greater value because delayed, more abstract outcomes are difficult to find when searching through cognitive space (Kurth-Nelson et al., 2012).

All of this early theorizing was based on the idea that choices between immediate and delayed options hinge on interactions between subjective, psychological factors that could reasonably be expected to vary depending on the context of the decision and the current state of the individual.

1.1.2 Discounted utility theory

The modern era of economic thinking about intertemporal choice began with Samuelson's (1937) discounted utility model. At the heart of this model is the idea of *temporal discounting*, which asserts that a delay preceding the delivery of an outcome affects the

[1] Utility is a measure of the relative "goodness" or satisfaction associated with an outcome. Utility is roughly equivalent with the common meaning of "value" or "reward," and the psychological principle of reinforcement; unless explicitly noted, we will use these terms interchangeably.

[2] Jevons' idea received belated support from an experiment that found human subjects increased their preference for delayed monetary rewards when they were cued to episodically imagine specific events in their future (Peters and Büchel, 2010). Interestingly, this study identified the hippocampus (a brain region we shall say much of in future chapters) as a critical mediator of this preference shift.

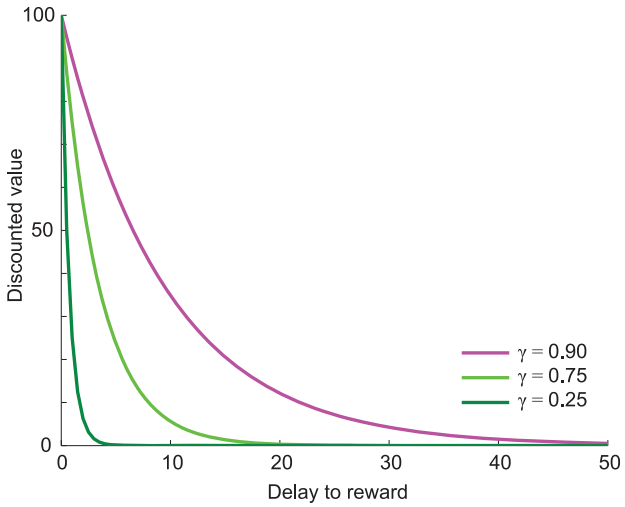

Figure 1.1: **Exponential discounting curves.** Following equation 1.1, the discounted value of an option is plotted as a function of delay. Curves were constructed for three different discounting rates (γ). With zero delay, the outcome is perceived as having its full, true value (100 units). However, as delay increases the outcome is perceived as less and less valuable. Discounted value drops more quickly with faster rates of discounting (i.e. $\gamma = 0.25$) than for slower discounting rates (i.e. $\gamma = 0.90$).

perceived or discounted value of that outcome. Temporal (or delay) discounting is the mechanism by which discounted utility theory accounts for immediacy preference. The model assigns a discounted value, DV, to a delayed outcome by decrementing its "true" value, V (i.e. the value at the moment of receipt, with no delay), in proportion to the amount of delay, d, that precedes arrival of the outcome. The discounting rate, γ (which ranges from 0 to 1), determines how quickly delay erodes an outcome's value. In modern notation, the discounted utility model takes the form:

$$DV = V\gamma^d \tag{1.1}$$

Discounted value decreases following an exponential curve, hence the term "exponential discounting," an oft-used shorthand or synonym for discounted utility theory. Using equation 1.1 to compute the discounted value of an outcome as the outcome's delay is varied produces a simple graphical representation of how delay influences perceived value, known as a *discounting curve* (fig. 1.1).

The discounted utility model contrasts with previous economic approaches in several

ways. Foremost, discounted utility eschews the multiple psychological factors posited in previous work, instead collapsing all influences on temporal preference into the discounting rate parameter. The model posits a literal decrease in value with delay, effectively expressing von Böhm-Bawerk's "undervaluation" of the future mathematically. Importantly, while earlier models explained immediacy bias by suggesting psychological factors were at work, the discounted utility model does not speak to *why* value decreases with delay. Thus, the model contends with immediacy preference almost tautologically, by defining a delay-dependent, but causally-agnostic, valuation system.

Samuelson himself cautioned that the discounted utility model was foremost a theoretical exercise; its form was chosen more for mathematical convenience and simplicity than to capture preferences in a realistic way (Samuelson, 1937). Samuelson repeatedly stressed the folly of considering the model too literally, as either a normative or descriptive statement, going so far as to claim that any person found to behave in accordance with such a naïve, invariant formula was likely not worth studying (Samuelson, 1937, pg. 160):

> Moreover, in the analysis of the supply of savings, it is extremely doubtful whether we can learn much from considering such an economic man, whose tastes remain unchanged, who seeks to maximise some functional of consumption alone, in a perfect world, where all things are certain and synchronised.

Nevertheless, because the discounted utility model provided a mathematical foothold for understanding intertemporal decisions, it quickly came to dominate economic analyses of temporal preferences. In addition to its mathematical facility, the model has several features economists find attractive. Some of these features are specific to discounted utility theory, particularly those that depend on the precise shape of the discounting curve defined by equation 1.1; others are generally true of any discounting curve in which value decreases with increasing delay.

According to equation 1.1, discounted value drops at a constant rate for each additional unit of delay (i.e. it follows an exponential curve). This property introduces a consistency across time intervals when choosing between different options. If one prefers

an immediately-available apple to an immediately-available orange, adding a fixed interval of delay to each option will not change their relative ordering, so the preference will hold whether the fruit are offered after an hour, a day, or a year. Variable rate discounting expressions (such as hyperbolic discounting, introduced in §1.2.3), can cause temporal inconsistencies in preference (e.g. preferring apples to oranges at one time point but oranges to apples if a fixed delay is added to each option). While such a scheme may not seem catastrophic, temporally-inconsistent preferences violate the economic principle of stationarity, a key requirement of normative, economically "rational" agents, which prevents self-contradictory valuations across different time intervals.

Eliminating conflicting psychological factors in favor of a single discounting parameter applicable to all outcomes and in all settings introduces another sort of consistency, by ensuring that all decisions are evaluated in the same manner, on a level playing field. Discounted utility theory is unaffected by secondary factors such as the way decisions are framed, the particular items being compared, or the current emotional state of the decision maker. While we might be skeptical of whether any species actually makes decisions in this way, as a normative model of how decisions *should* be made, discounted utility theory's context independence and immunity to contravening influences could be advantageous.

Finally, it is worth noting that several general economic considerations suggest that immediate outcomes should be favored. These factors do not necessarily require an exponential discounting curve, but are nevertheless addressed by the discounted utility model. For instance, decisions involving money naturally raise the question of investment. A more immediately available sum of money can be invested as soon as it is received, and might, therefore, either function as capital to support some further money-making venture, or simply accrue interest. Thus, depending on the sums of money and delays that are involved in the decision, along with other factors such as interest rates or business prospects, it is easy to imagine scenarios in which a smaller amount of money available after a shorter delay could be preferable to a larger sum of money far in the future.

The question of investment depends heavily on the particular outcomes under consideration. Money is one clear case for which investment could be useful. However, for other goods the possibility of investment is limited; an animal choosing between different amounts of food, for instance, can "invest" only a finite amount of nutrients in its

survival at one time. Even if an animal can gorge itself when an abundance of food is available, receiving vastly more food than can be consumed before it spoils may not be useful. This issue is revisited in more detail during the discussion of behavioral ecological approaches to intertemporal choice (§1.3).

Uncertainty about the future is a related issue that may favor immediate over delayed outcomes. Potential intervening events (e.g. the death of the decision maker, the collapse of whatever entity is offering the delayed outcome) may make a short-term gain preferable for its certainty. This problem is complicated by the fact that the distribution of potential future events impacting collection of the delayed outcome influences how the future should be discounted. Assuming the rate of hazards in the future is independent of time (i.e. hazards are a Poisson process) suggests that exponential discounting, which produces a constant rate of value decline with increasing delay, is the normatively correct strategy (Dasgupta and Maskin, 2005). However, if future calamities occur at an unknown or non-uniform rate, non-exponential discounting functions might be advantageous (Sozou, 1998; Dasgupta and Maskin, 2005; Henly et al., 2008).

1.1.3 Empirical challenges to discounted utility theory

Despite its attractive features, the empirical validity of discounted utility theory has been found wanting in several key respects. Temporally inconsistent preferences are a robust experimental observation across a wide range of time frames and outcomes, in both humans and other species (Rachlin and Green, 1972; Ainslie, 1974; Frederick et al., 2002). A simple thought experiment bears this out: imagine choosing between $10 available immediately and $11 delivered in one week. A sizeable fraction of people prefer the immediate sum over the larger, delayed amount. However, consider the same decision, but with a constant delay amount added to both options: now the choice is between $10 delivered in one year, and $11 delivered in a year and a week. In this scenario, the majority of people favor the slightly larger sum of money, reversing their preference and violating the principle of stationarity built in to discounted utility theory. Such temporal inconsistency implies that discounting rates are not constant per given unit of time, in at least some situations.

Other factors have been shown to have an inconvenient effect on discounting rates. In choosing between monetary sums, the amount of money under consideration influences

discounting rate. Subjects tend to discount small sums relatively steeply (for example, preferring an immediate $10 to a delayed $60), but discounting rates are generally slower for choices involving larger sums (e.g. $1000 immediately or $6000 after a delay). Equation 1.1 specifies no mechanism by which an outcome's value can influence discounting rate, making such results difficult to reconcile with standard discounted utility theory. Similarly, gains and losses are often discounted at different rates, although discounted utility theory makes no allowance for this to occur.[3]

1.1.4 Summary

Discounted utility theory has exerted a tremendous influence on the study of temporal preferences, both within and beyond economics. The model's convenient features led to wide-spread application by economists, bolstering its legitimacy, and facilitating the transfer of the temporal discounting concept to other fields. The success of discounted utility theory came at the expense of the psychological explanations of delay preferences that characterized early work, which, until recent years, were largely absent from subsequent economic models.

Although the shortcomings of discounted utility theory have been frequently recognized (e.g. Loewenstein and Prelec, 1992; Frederick et al., 2002; Rubenstein, 2003), the model's central premise, temporal discounting, has undergone a sort of reification, such that for many investigators "intertemporal choice" and "temporal discounting" are so intertwined that they have become effectively synonymous. Consequently, most attempts to reconcile discounted utility theory with experimental observations have not addressed the core assumption of temporal discounting, but have focused instead on modifying the shape of discounting curves, (in some cases by reintroducing competing psychological motives; e.g. Phelps and Pollak, 1968; Laibson, 1997), and allowing for contextual factors to modulate discounting curves.

[3] We note here that while our focus in the thesis is on choices involving appetitive (desirable) outcomes, less is known about temporal preferences for aversive outcomes. For example, rats generally favor a large shock delivered sometime in the future over a smaller shock delivered sooner, trading off a shock-free present for a more unpleasant future (Deluty, 1978). An analogous experiment in humans, however, found that most subjects preferred a smaller, more immediate shock to a larger, delayed one (Berns et al., 2006). In a slightly less visceral domain, humans generally weight monetary losses more heavily than gains (e.g. Kahneman and Tversky, 1984), and consequently discount negative financial outcomes more slowly (Loewenstein and Prelec, 1992).

Although discounted utility theory has perhaps enjoyed a somewhat unwarranted longevity, its development ushered in the modern scientific study of temporal preferences, providing a well-specified (if limited) mathematical framework for examining how decision makers trade off value and delay. Recognition of its inadequacy at accounting for behavior in intertemporal choice scenarios set the stage for development of psychological approaches to explaining temporal preferences.

1.2 Psychological approaches to intertemporal choice

If economists were the first to systematize models of intertemporal choice, psychologists deserve credit for bringing the question into the laboratory. Psychological research on delays evolved out of experiments designed to elucidate basic principles of associative learning, and eventually developed into a full-fledged branch of study. This work culminated in the development and experimental validation of hyperbolic discounting models, which remedy a number of the problematic issues that plague discounted utility theory and exponential discounting approaches.

1.2.1 Associative approaches

Some of the earliest experimental work in psychology, while not directed at intertemporal choice explicitly, recognized the importance of delays in influencing animal behavior. Thorndike's law of effect (1911) posited that actions followed by favorable consequences were more likely to be repeated, while actions followed by unpleasant consequences diminished in frequency. This general principle anticipated the modern psychological notion of reinforcement, and provided a foundation for the later development of operant conditioning procedures (Konorski and Miller, 1937; Skinner, 1935). A question embedded in Thorndike's proposal concerns how the amount of delay separating actions and outcomes impacts learning.

Hull (1943) reviewed early experiments directed at answering this question. In such experiments, rats were rewarded for performing some operant response following the presentation of a cue, but the delay to reward was varied across groups of animals, such that some rats earned reward immediately after a correct response, while others had to wait for a delay (ranging from 5 s to 20 min) before receiving reward. In general, these

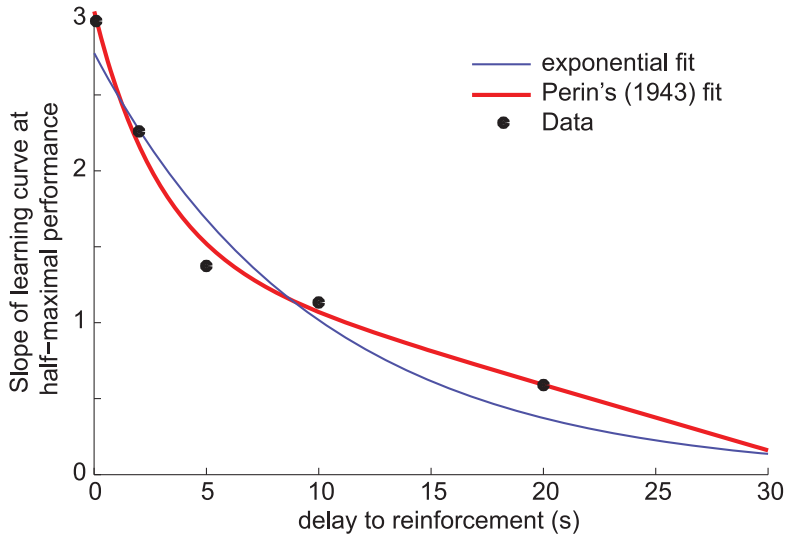

Figure 1.2: **The reinforcement gradient.** Perin (1943) tested the effect of reinforcement delay on the rate of learning an operant task. Although Hull (1943) used these data to postulate an exponential relationship between reinforcer delay and learning (blue curve), Perin's non-exponential fit (red curve) accounts for the data more accurately. Hull's exponential fit follows the form $y = a\exp(bx)$, while Perin's equation was of the form $y = a\exp(bx) - cx + d$. Data were re-plotted from Perin (1943; table 4).

experiments showed that as pre-reinforcement delay increased, learning became increasingly compromised. This suggested to Hull (1943) that the ability of reinforcement to support learning followed a gradient, with maximal efficacy occurring at zero pre-reward delay, and decreasing as delay grew longer.

Unsatisfied with a qualitative description, however, Hull set out to determine the mathematical form of the reinforcement gradient (Hull, 1943). Hull's derivation was heavily influenced by data from an experiment by Perin (1943), which Hull considered to be less affected by the influence of second-order conditioning effects than other investigations. As in previous studies, Perin's results suggested a decrease in learning rate with increasing pre-reward delay (fig. 1.2).

Interestingly, although the curve Perin (1943) fit to his data was not exponential, Hull (1943) nevertheless formulated his gradient of reinforcement efficacy as a negative growth function. In Hull's model, pre-reward delay influenced both the maximal amount of learning a reinforcer could support (the "habit strength") and the rate at which the association was acquired (the learning rate).

As Hull (1943) noted, his model offers a purely associative account of preference for more immediate outcomes. Because of the faster learning rate and the greater amount of learning that short pre-reward delays support, given the same number of experiences with a long and short outcome, the shorter of the two will have acquired a greater habit strength, increasing the probability that the behavior reinforced by the short delay will occur.

1.2.2 Matching law

A prominent thread in early studies of animal decision making was testing how subjects spread their responses between two simultaneously-available but independent streams of reinforcement. In these experiments, two response devices (e.g. levers) were available to animals in the testing chamber. Each lever was programmed to deliver reward on a variable interval (VI) schedule.[4]

Herrnstein (1961) studied pigeons performing on concurrent VI schedules, and noted that animals split their choices between the two options in proportion to the programmed rates of reinforcement of each, following the relationship:

$$\frac{P_1}{P_1 + P_2} \approx \frac{R_1}{R_1 + R_2} \tag{1.2}$$

where P is the number of pecks on response keys for options 1 and 2, and R is the programmed VI rate for the options. Thus, subjects "matched" their responding to the reward statistics of the options. Because behavior on concurrent schedules across a wide range of species and experiments was found to reliably follow this relationship, it eventually became known as the *matching law* to denote its generality. It has been suggested that matching is an innate reward-harvesting strategy that tailors behavior to the estimated rates of reward of sources of reinforcement (Gallistel et al., 2007).

Chung and Herrnstein (1967) introduced an intertemporal element to concurrent VI experiments by enforcing a time out period between responses on a primed option and

[4] Variable interval schedules program reward delivery with unpredictable inter-reward intervals. Inter-reward intervals are drawn from an exponential distribution with a characteristic mean; a VI-4s schedule would be primed, on average, every 4 s. Responses to primed levers result in reward delivery, following which the option becomes inactive until the next inter-reward interval passes and it is once again primed. However, once primed a lever remains in that state until the subject registers a response on it and earns reward.

reward delivery. In their experiments, lever priming occurred as usual for a VI schedule, both options were programmed with the same VI rate, and both options delivered the same amount of reward. However, responses to a primed lever were not rewarded until the intervening time out period had elapsed. Different pairs of time out delays were tested in blocks of trials; the delay for the "standard" option was fixed at 0, 8, or 16 s, while the delay on the "experimental" option ranged from 1–30 s. The proportion of responses to the standard option varied with the ratio of the standard and experimental time outs, as though subjects allocated their responding to the options by matching the ratio of standard and experimental delays, d:

$$\frac{P_1}{P_1 + P_2} \approx \frac{d_1}{d_1 + d_2} \tag{1.3}$$

In determining the mathematical relationship between delay and choice, Chung and Herrnstein (1967) departed from the Hull's (1943; §1.2.1) exponential formulation, noting correctly that Hull's conjecture was based on a selective reading of the literature, and moreover was predicated on the assumption that delay-dependent differences in behavior were driven purely by variation in habit strength (i.e. the strength of the stimulus-response association) rather than a choice or preference *per se*. More importantly, they observed that delay-based matching (eq. 1.3) was incompatible with an exponential relationship between response and delay ratios; instead equation 1.3 predicts a curve with more concavity than an exponential function. However, data from the experiment were variable enough to accommodate either an exponential or the matching-derived curve, leaving the issue unresolved (Chung and Herrnstein, 1967).

Chung and Herrnstein's (1967) experiment was one of several studies that helped establish the generality of the matching relationship. In addition to Herrnstein's (1961) original observations of matching to VI rate, subsequent work found that animals matched to reinforcement magnitude (Catania, 1963) when programmed rates of reinforcement were equal. The addition of matching to immediacy (the reciprocal of delay) to the list of reinforcement parameters that control allocation of behavior began to suggest a deeper underlying principle at work.

Baum and Rachlin (1969) tested pigeons on a concurrent VI task that did not require subjects to press a lever to indicate their choices. Instead, to respond to one of

the options subjects simply moved to either the left or right side of the experimental apparatus. Pigeons in the experiment matched the time they spent in each half of the chamber to the VI rate programmed for each option, confirming that the matching principle holds for the less explicitly operant response of time allocation. Integrating these results with previous work (in particular, Neuringer, 1967), Baum and Rachlin (1969) suggested a generalized matching expression:

$$\frac{T_1}{T_1 + T_2} \approx \frac{V_1}{V_1 + V_2} \tag{1.4}$$

where T is the time allocated to each option, and V is each alternative's *value*. In this formulation value multiplicatively combined parameters of reinforcement such as rate, magnitude, and immediacy to produce an integrated measure of reinforcement, similar to utility in economics. Instead of treating responding as a discrete or semi-continuous variable (e.g. number of lever presses, or rate of lever pressing), equation 1.4 suggests time allocation (be it time spent pressing the lever, or, as in Baum and Rachlin's (1969) experiment, time spent in one half of the chamber) as a more fundamental measure of behavior.

Positing value as the factor that determines behavioral allocation was an important step in the transition from the associative accounts that dominated early psychological theories of animal behavior to modern notions of decision making that incorporate subjective, unobservable factors in the valuation process, and increasingly ascribe importance to how external features of the world are represented internally.

MacCorquodale and Meehl (1948) commented on the nature of hypothetical constructs in psychology, noting a distinction between posited internal variables that were not directly measurable but were defined entirely with reference to observables, and hypothetical factors that were neither measurable nor derivable from observation. Habit strength as developed by Hull is an example of the former; although Hull (1943) was not optimistic about uncovering its brain locus, he argued that habit strength was more than a "metaphysical entity" because it was anchored directly in observation. As habit strength is mathematically defined by an animal's experiences, it can be computed accurately independent of its external reality.

Logical constructs that follow directly from observation contrast with ideas such as

those developed by Tolman. For instance, Tolman's (1932; 1938) "expectancies" were learned regularities about the structure of the world that animals could use to guide behavior. Unlike habit strength, however, knowing an animal's experience with great precision is not sufficient to infer the expectancies it has developed, because expectancy learning depends on still other internal variables (e.g. attention, motivation, etc.). Although approached with trepidation for many years, more nebulous concepts like those of Tolman are increasingly accepted, thanks in part to both behavioral, and increasingly neural evidence for their existence (Johnson and Redish, 2007; Johnson et al., 2009; van der Meer and Redish, 2009; Steiner and Redish, 2014).

The value defined by Baum and Rachlin (1969), as a simple product of experimentally-controlled reinforcement parameters, was closer in nature to Hull's habit strength than to Tolman's expectancies. However, the authors noted that their formulation could accommodate subjective, unobservable factors as well. To address an idiosyncratic side preference that subjects in their experiment exhibited, the empirically-computed value ratio (i.e. the right-hand side of equation 1.4) was multiplied by a constant factor fitted from behavior to satisfy the matching equality. Thus, while the cause of the side bias remained unspecified and unknown, its influence on behavior could be modeled quantitatively in a matching framework by a simple modification of equation 1.4.

The tautology implicit in modeling behavior this way was not lost on proponents of the approach (e.g. Rachlin, 1971); despite this, the matching framework constitutes a useful tool in the analysis of behavior. First, as Rachlin (1971) notes, casting behavior as matching forces "codification of assumptions underlying choice experiments." Frequently, interpreting behavior with respect to a theoretical model requires making assumptions; the matching formulation makes these assumptions explicit and obvious, rather than leaving them lurking unspecified and inconspicuous. The assumption that behavior matches time allocation to value permits quantification of the extent to which this end is or is not achieved, and suggests that in cases where matching does not obtain, the fault is in our understanding of the animal's mechanism of valuation, rather than some error on the subject's part. Measuring how deviations from matching vary under different sets of experimental conditions can provide insight into the decision making algorithms that drive behavior.

As Killeen (1972) commented, using the matching framework in this way to understand behavior has many parallels to the common practice in economics of positing certain properties of decision makers (e.g. rationality, utility maximization) and then deriving models based on whether or not humans behave as though these assumptions are true. The most enduring contribution of the matching law to psychological studies of animal decision making may lie as much in its method of explicitly defining a valuation framework to compare behavior against as in its actual mathematical form.

1.2.3 Hyperbolic discounting

George Ainslie (1975) marshalled decades of experiments and theory on delay preferences to propose a new model of temporal discounting. He observed that both deviations from discounted utility theory (§1.1.2) and data from psychology experiments investigating the impact of delays preceding reward delivery (§1.2.1,1.2.2) argued strongly against an exponential relationship between the perceived value of an outcome and the outcome's delay. Ainslie suggested instead that discounting functions were more correctly described by hyperbolic curves, where discounted value falls off rapidly for short delays but declines more slowly at longer delays.

Figure 1.3 shows examples of exponential and hyperbolic discounting curves. The relationship between discounted value and delay is quite similar for both curves, a fact that precluded discriminating between exponential and hyperbolic discounting curves by simply comparing how well each model fit experimental data (e.g. Chung and Herrnstein, 1967). However, the subtle differences between the curves give rise to a qualitative feature of behavior that is not easily explained by exponential discounting.

Figure 1.3 contrasts hyperbolic and exponential discounting curves for two outcomes of different magnitude. At each time point on the x-axis, the reinforcer with a greater discounted value will be preferred by a decision maker choosing between the two. When value is discounted exponentially, the discounting curves for the options never cross; decisions made at any time favor the smaller outcome. However, when reward value falls off following a more concave, hyperbolic function, the curves cross, suggesting a shift from preferring the larger option when the time to either reward is long, but the smaller option as the time frame of the decision contracts.

Hyperbolic discounting curves lead to temporally-inconsistent preferences, which,

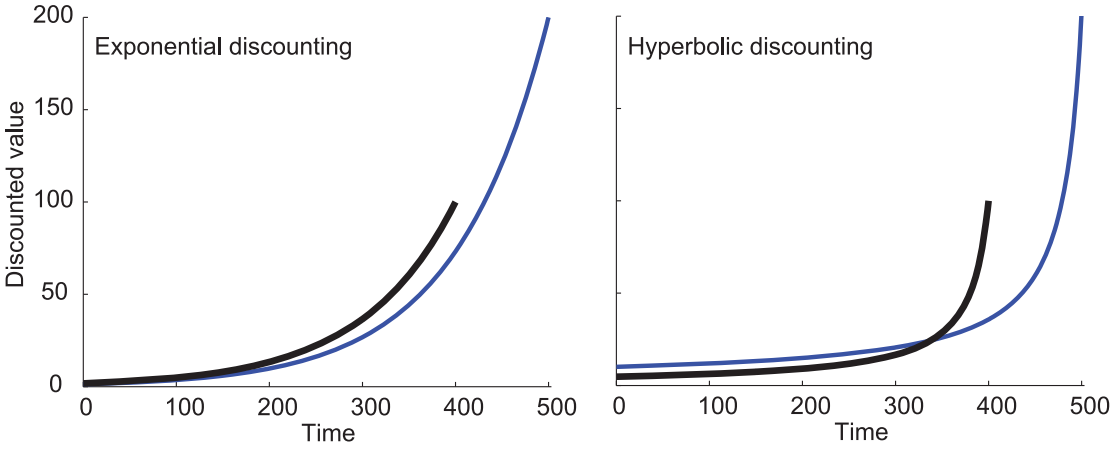

Figure 1.3: **Preference reversal.** Discounting curves are plotted for a decision between a large reward (value=200) delivered at time 500, and a smaller outcome (value=100) delivered at time 425. With exponential discounting (left panel) the value of the small outcome (black curve) exceeds the value of the larger, more delayed option (blue curve) at all time points. However, hyperbolic discounting (right panel) induces temporal inconsistency. When the decision is far in the future (e.g. time 100), the large option has the greatest value; however, as the decision draws nearer in time (e.g. time 350), the value of the outcomes is reversed, and the smaller option is preferred.

as discussed in §1.1.2, are a robust experimental finding in both animals and humans (Rachlin and Green, 1972; Ainslie, 1974), but are impossible to achieve with exponential discounting. Consequently, any model of intertemporal choice in which decisions depend on comparing the discounted values of the options under consideration requires a discounting function with greater concavity than an exponential curve to account for experimental data.

Ainslie (1975) showed that hyperbolic discounting provides a mathematical explanation for the ambivalence humans frequently report when choosing between a tempting, immediately-available outcome and a better option that will not be realized until some time in the future. For example, on the morning after a long night out on the town (discussing science with colleagues at a conference, for example), the value of going to bed early might exceed that of spending another night out late with friends. However, as the final poster session of the day concludes, and the prospect of going back to the hotel room draws near, we might reverse our previous decision, and opt for another long evening of illuminating scientific discourse with our peers.

Temporal inconsistencies also explain why, during moments of resolve, we sometimes

make decisions that limit our options in the future. For instance, having a portion of our paycheck automatically deposited in savings every month ensures that when payday comes, it will be more difficult to give into the immediate temptation of spending. Ainslie (1975) noted that such "precommitment" strategies, in which we make a choice in the present (automatically saving money) that eliminates a choice which would otherwise occur in the future (save my paycheck or spend it?), depend on non-exponential discounting. Recent work that tested precommitment in simulated decisions between delayed outcomes showed that the precise shape of discounting curves strongly influences the propensity for precommitment, with a minimum degree of hyperbolicity in discounting curves required for precommitment to occur (Kurth-Nelson and Redish, 2010).

In his original proposal, Ainslie (1975) did not specify a particular equation for hyperbolic discounting curves, but noted that matching ratio formulae proposed previously (§1.2.2) predicted hyperbolic curves. These models are not without theoretical inconveniences; strictly speaking, the value of options with zero delay is infinite by equation 1.3, suggesting that an immediate option would be rigidly preferred to any outcome with a delay greater than zero, regardless of whether that delay was very short, or the magnitude of the delayed outcome was substantial. Ainslie's analysis sidestepped this issue by considering only choices between options with non-zero delays, and showed that discounting curves constructed from equation 1.3 crossed in a manner suggestive of temporally-inconsistent preferences.

Mazur (1987) developed a novel behavioral paradigm to test intertemporal choice in animals and used data generated in these experiments to compare discounting functions. His approach centers on the idea of an *indifference point*, a set of delays for which two outcomes are equally preferred.

For instance, offered a choice between two food pellets delivered immediately and four food pellets delivered after 60 seconds, a rat may choose the immediate, "smaller-sooner" option the majority of the time, rarely selecting the delayed, "larger-later" outcome. However, if the delay on the larger option was progressively reduced, the rat would likely increasingly favor the larger option. When the delay reached a point where the rat showed no reliable preference between the larger-later and smaller-sooner options (i.e. responses were split evenly between the two options), we could infer that the discounted value of each option was equal, and that the large delay had reached the

animal's indifference point.

Indifference point delays are specific to the magnitudes of outcomes under consideration. For instance, in the previous example, if the rat was indifferent between choices when the large reward delay was 10 seconds, we would predict that increasing the number of larger-later pellets would increase the indifference point, as the increased reward magnitude could compensate for a greater amount of delay. Similarly, decreasing the number of food pellets for the larger-later reward would be expected to decrease the indifference delay, as the reduced reward magnitude could cancel out only a smaller amount delay.

Mazur's (1987) behavioral task was designed to elicit indifference points for pairs of outcomes as efficiently as possible. In the adjusting-delay task, animals make operant responses to indicate choices for the larger-later or smaller-sooner options. The task is structured in four-trial blocks; the first two trials of each block are forced choices to ensure familiarity with the delays associated with each outcome. The remaining two trials of each block are free choices. Choosing the larger-later outcome for both free-choice trials in a block results in the delay on the larger-later outcome incrementing by 1s for the next block of trials; choosing instead the smaller-sooner reward for both free choices causes the delay on the larger-later outcome to decrease by 1s in the next block. Choosing each option once in the free choice trials results in no change to the larger-later delay. This paradigm and its variations have been applied extensively in studies of intertemporal choice. Tasks of this form are often referred to as "self-control" paradigms, based on the idea that choosing a smaller-sooner option is an impulsive choice, and that selection of the larger-later option requires inhibition of this tendency, which is achieved by exerting self-control.[5]

The adjusting-delay procedure in principle allows animals to iteratively converge on indifference delays for pairs of options by adjusting the larger-later delay with their choices. By finding the indifference delays for fixed-magnitude small and large rewards while varying the smaller-sooner delay, Mazur (1987) was able to compare hyperbolic and

[5]In this thesis, we avoid using terms like "impulsive" or "self-controlled" to describe behavior, as such language immediately implies a normative judgement on the quality of the choice which may not be justified in every setting. For instance, depending on the magnitudes of reward and the delays under consideration, it is possible to construct scenarios where selecting the smaller-sooner option is economically correct under any framework that one cares to consider. Accordingly, the pejorative term "impulsive" should not be considered synonymous with preference for immediacy.

exponential discounting curves experimentally. His exponential model followed Hull's (1943) reinforcement gradient:

$$DV = Ae^{-Kd} \tag{1.5}$$

where DV is the discounted value of a reward amount A, delayed by d seconds. Mazur's K is the discounting rate (equivalent to γ in Samuelson's (1937) discounted utility theory; eq. 1.1).

As a model of hyperbolic discounting, Mazur (1987) began with Baum and Rachlin's (1969) value-matching model (eq. 1.4), noting that this formula suggested a curve with a greater concavity than an exponential. The discounting rate parameter K in the denominator scales sensitivity to delay (d). To avoid infinite values for outcomes with zero delay, Mazur simply included a 1 in the denominator:

$$DV = \frac{A}{1 + Kd} \tag{1.6}$$

To arbitrate between the models, Mazur (1987) relied on the fact that, for a given pair of large and small magnitude rewards, the indifference delay varies as a function of the delay to the smaller-sooner option. Varying the smaller-sooner delay across conditions, and allowing animals to titrate the larger-later delay to their indifference point allowed Mazur to construct *indifference curves* (plots of the larger-later indifference delay as a function of the smaller-sooner delay).

Armed with equations for exponential and hyperbolic discounting functions, Mazur computed theoretical indifference curves for each of the models and compared them to pigeon's experimentally-obtained indifference curves. While both hyperbolic and exponential models predict linear indifference curves, exponential discounting predicts a slope of one, while indifference curves under hyperbolic discounting should equal the ratio of large and small reward magnitudes. Data from Mazur's experiment strongly favored hyperbolic discounting, with slopes of indifference curves substantially greater than one.

While often effective, adjusting-delay procedures are not without their shortcomings. Cardinal and colleagues (2002) tested rats on the standard adjusting-delay task and analyzed the pattern of decisions at a trial-by-trial level. They observed that rats' behavior, did not exhibit clear sensitivity to the adjusting delay length. In many cases,

rats titrated the adjusting delay in an oscillatory pattern, first driving it up to the maximum possible length and then driving it down to the shortest possible length. Computer simulations showed that a number of decision rules (some sensitive to delay, some not) could have resulted in the pattern of choices that rats made on the task.

Potential problems with the standard adjusting-delay task may lie in its block-wise design and the inclusion of forced trials. This structure is enforced to ensure that subjects remain familiar with delay lengths as they change over the course of the experiment; however, it is not clear that subjects understand and use this information, and forcing choices could bring about other complications. For instance, if animals exhibit any tendency for perseveration, the order in which forced trials are presented could have an unwanted influence on the subsequent free choices. To circumvent these potential issues, Papale and colleagues (2012) adapted the adjusting-delay procedure to a spatial framework, allowing rats to select the adjusting and fixed delay options by turning left or right at the choice point of a T-maze. Their experiment did not include forced trials, but rats nevertheless showed consistent and stable patterns of titration, adjusting the delay to a steady value and then maintaining it there over the course of a single behavioral session with only 100 total trials. These data suggest that while adjusting procedures can be an efficient method of eliciting delay preferences, care must be taken to ensure that behavior is not inadvertently influenced by incidental idiosyncracies in the task design.

Since its introduction, Mazur's (1987) hyperbolic equation has enjoyed great success in fitting experimental data from a wide range of species and experimental contexts. Equation 1.6 has become the *de facto* standard for modeling hyperbolic discounting in intertemporal choice experiments. Although popular and empirically supported, hyperbolic discounting functions have not entirely replaced exponential models. Setting aside qualitative predictions such as the preference reversal phenomenon, when simply estimating discounting curves from subjects' choices between two delayed outcomes, both exponential and hyperbolic curves often adequately fit the data, and allow a parametric quantification of discounting rate. Thus, whether due to mathematical convenience or in deference to the normative considerations highlighted in §1.1.2, exponential discounting remains an often-used tool in experimental studies of decision making. Importantly, between the exponential and hyperbolic models, researchers have successfully described,

parameterized, and interpreted behavior from hundreds (perhaps thousands) of decision making experiments involving delayed outcomes.

We note here that while exponential discounting specifies a particular function that is defined solely by the discounting rate parameter, hyperbolic discounting is less tightly specified. In modeling or fitting discounting functions, any curve with a concavity greater than that of the exponential function is typically described as hyperbolic. While Mazur's (1987) instantiation of hyperbolic discounting is popular, other mathematical formulations can produce hyperbolic curves that fit experimental data (including preference reversals) equally well. For instance, as discussed above, the value matching proportion equation define a "true" hyperbolic curve that limits to infinity as delays approach zero. A model introduced by Kurth-Nelson and Redish (2009) achieves hyperbolic discounting by averaging across a distributed population of exponential curves (Kacelnik, 1997; Sozou, 1998; §1.4). Thus, when considering discounting functions, it should be recognized that hyperbolic discounting can be implemented by a family of curves rather than a single, constrained function.

1.2.4 Summary

Psychological approaches to understanding temporal preferences grew out of early work in learning theory. The development of matching law to account for choices between concurrently available reinforcers provided a framework for understanding how preferences were influenced by the value of outcomes, and opened the door to modeling subjective influences on valuation. The mathematics of matching (i.e. equating proportions of responses to proportions of value) presaged the development of hyperbolic discounting functions.

The synthesis of psychological investigations of delay preferences and economic thinking led to the recognition of exponential discounting's inadequacy in explaining experimental results. Subsequent psychological investigations of intertemporal choices have relied heavily on hyperbolic discounting functions, with many following Mazur's (1987) parameterization, but others relying on different mathematical specifications to achieve curves with more concavity than exponential functions.

1.3 Ecological perspectives on intertemporal choice

Behavioral ecology, the study of how evolutionary and ecological forces sculpt animal behavior, offers a unique and complementary perspective on decisions involving delayed outcomes. This work derives from a rather different set of core principles compared to economic and psychological approaches to studying decision making, but is concerned with answering many of the same questions. Foraging theory, a branch of behavioral ecology, attempts to understand how animals allocate their time in the pursuit of limited resources that are critical for survival.

Foraging theory models prescribe strategies that animals should use to optimize their long-term rate of energy (i.e. food) acquisition under a set of well-specified, simplifying assumptions. The energy rate maximization framework has rich connections to psychological and economic delay discounting models, and offers an account of intertemporal choice that explains delay preferences without invoking temporal discounting.

1.3.1 Holling's disc equation

Foraging theory rate maximization models draw heavily on the work of Holling (1959) and his "disc equation," so we begin our discussion by detailing its origin.

In order to model the dynamics of populations of prey and predator species in an environment, Holling (1959) was interested in understand how the density of prey animals influenced the consumption patterns of individual predators. It may seem obvious that as prey density increases, predators will necessarily encounter prey more frequently, and their consumption will increase. However, it is equally intuitive that consumption must eventually plateau, if only because the maximum rate of prey consumption and nutrient assimilation is necessarily limited by the physiology of the predator (e.g. finite stomach capacity).

To understand the function relating prey density and consumption by predators, Holling (1959) conducted a simple model experiment: a blindfolded human "predator" searched for and captured as many "prey" (sandpaper discs tacked to a table) as possible in one minute. Upon encountering a disc, the subject removed it from the table, and then continued searching. The number of discs tacked to the table in each experiment was varied to model different prey densities.

As might be expected, the number of discs participants captured increased with the number of discs on the table. Holling (1959) proposed an equation to predict how many discs would be captured given a particular disc density:

$$N_{discs} = aTx \tag{1.7}$$

where the amount of discs achieved (N_{discs}) depended on the length of the experiment (T), the density of discs (x), and a fitted constant a. However, because removing discs from the table took some amount of time, the number of discs collected within the fixed time interval of the experiment eventually reached a steady state. Equation 1.7 does not capture the saturation that occurs with increasing density.

To incorporate this observation, Holling (1959) noted that total time in the experiment, T_{total}, was actually partitioned into time spent searching for discs, T_{search}, and time spent removing or handling discs, $T_{handling}$, where $T_{total} = T_{handling} + T_{search}$. This fact is true both of Holling's model experiment and real-life prey capture scenarios (catching and consuming prey is not an instantaneous event). Accordingly, in equation 1.7, N_{discs} could be more accurately described as a function of time spent searching rather than the entire length of the experiment.

Assuming that subjects could not search for more discs while handling a captured disc suggests that the total time available for search is equal to the number of discs captured multiplied by the fixed length of time required to handle each captured disc, h. Substituting this expression for search time in place of total time in equation 1.7 and simplifying, Holling (1959) arrived at:

$$N_{discs} = \frac{T_{total}ax}{1 + ahx} \tag{1.8}$$

This formula captures the density-dependent saturation in disc capture evinced by subjects in the experiment.

Holling (1959) arrived at his equation empirically, by performing the experiment and fitting the resulting data. However, it is possible to derive his relationship by beginning with the assumption that time spent searching for and handling prey are mutually exclusive, and that the number of prey encountered increases linearly with the amount of time spent searching (Charnov and Orians, 1973; Stephens and Krebs, 1986).

Instead of solving for the absolute number of prey captured, it is useful instead to calculate the rate of prey capture per unit of time spent searching. Because the point of capturing prey is to consume them, we can equate rate of prey capture with energy intake rate, R. The energy intake rate is the total amount of energy gained from consuming prey items, E, divided the sum of search time, T_{search}, and handling time, $T_{handling}$.

$$R = \frac{E}{T_{search} + T_{handling}} \tag{1.9}$$

Because encounters with prey increase as a function of searching time, and because the total amount of time spent handling prey depends on how many prey are encountered in the first place, both $T_{handling}$ and E can be expressed as a function of T_{search}. By incorporating Holling's (1959) concept of disc density as a rate of prey encounter (λ), the number of prey items a forager could expect to find in a given search time is $T_{search} * \lambda$. If we know that individual prey items provide a fixed amount of energy, the numerator of equation 1.9 can be written as the number of prey items encountered ($T_{search} * \lambda$) multiplied by the energy each one provides (e). Similarly, the handling time for a single prey item, h, multiplied by the number of prey encountered is the total time spent handling prey. Substituting these expressions into equation 1.9 yields:

$$R = \frac{T_{search}\lambda e}{T_{search} + T_{search}\lambda h} \tag{1.10}$$

The inclusion of one further expression to the numerator of equation 1.10 allows us to model the fact searching for prey is an energetically-expensive activity. If R reflects the rate of energy gained while foraging, the cost per unit time of searching, s must be subtracted from the energy gained by capturing prey:

$$R = \frac{T_{search}\lambda e - sT_{search}}{T_{search} + T_{search}\lambda h} \tag{1.11}$$

Equation 1.11, simplified, is the rate expression of Holling's (1959) disc equation:

$$R = \frac{\lambda e - s}{1 + \lambda h} \tag{1.12}$$

Holling's (1959) disc equation, and the foraging theory rate maximization models drawn from it, have several characteristics that set them apart from the economic and

psychological models discussed thus far. Foremost, foraging theory models are developed around a sequential decision topology. While delay discounting models are often used to analyze "one-shot" decisions, many intertemporal choices are actually sequences of decisions, just as encounters with discs/prey items occurred sequentially in Holling's experiment.

In the simple model developed above, prey encounter might not seem to present any choices at all. However, imagine that Holling's (1959) experiment had been conducted with discs of different sizes, with size representing the prey item's energetic value. Upon encountering a prey item, subjects would choose to either accept that disc (and expend a fixed amount of search time in handling it), or reject the item in favor of continued searching. A foraging animal's task could then be characterized as a series of accept/reject choices upon prey encounter, interspersed with bouts of searching. This pattern is referred to as foreground/background structure, because a decision is made between the foreground option (e.g. a recently encountered prey item) and a "default" background option (e.g. continued searching for different or better prey). This arrangement contrasts with the two-alternative force-choice tasks often used to study animal decision making, in that foreground and background options need not been seen as mutually-exclusive; choosing to exploit a foreground prey item does not rule out the possibility of afterward reverting to search (the background option).

The foreground/background structure hints at another element critical to foraging theory models, which is the principle of *opportunity cost*. Performing a particular behavior generally entails not performing a host of other potential behaviors; choosing to perform one behavior comes at the expense of others. For instance, opting to exploit the foreground prey type rather than continuing with the background searching option might be either a good or bad decision depending on how it compares with options potentially present in the background. Opportunity cost, then, measures the energy intake that might have resulted from unselected, excluded behaviors. By assuming that search and handling cannot occur simultaneously, foraging models bring opportunity cost to the fore in analyses of decision making.

Two basic foraging theory models, the prey and patch frameworks, have served as starting points for the development of more complex models adapted to specific situations. Both models share a common logic rooted in Holling's (1959) disc equation, but

diverge to treat different classes of foraging decisions.

1.3.2 The patch model

The patch model considers how foragers should best exploit a diminishing resource. For example, consider picking berries. While initially an easy task, as time passes and the low-hanging fruit are collected, with each berry that is harvested it becomes harder and harder to obtain the next one. At some point, it might be best to declare the bush exhausted, and move on to another. However, if berry bushes are few and far between, the better decision might be to work for every single berry on the bush, despite the increasing difficulty and diminishing return.

In the patch model, patches are discrete food sources (like the bush in our berry picking example) that provide some amount of energy to foragers for every unit of time the forager spends exploiting the patch. Patches are defined to be diminishing resources; accordingly, a patch's *gain function* (the relationship between time spent at the patch and the instantaneous rate of energy intake the patch provides) must be a negatively-accelerated curve.[6]

The patch model does not consider the question of which patches to enter in the first place; this problem is addressed by the prey model, described below (§1.3.3). Instead, it may be assumed that the patch in question is worth entering and occupying for some amount time. The patch forager must determine how long to remain in a particular type of patch before abandoning it and traveling to the next.

In the simplifying case of a single type of patch in the environment, and no cost for searching for new patches, the patch exploitation rate equation is:

$$R(t) = \frac{\lambda g(t)}{1 + \lambda t} \tag{1.13}$$

where $R(t)$ is the instantaneous rate of energy gained from the patch by occupying it at time t, $g(t)$ is the patch's gain function, and λ is again the rate of encountering patches in the world (Stephens and Krebs, 1986).

[6] If the gain function is not negatively accelerated, the patch would be non-depleting, providing a constant (or even increasing) rate of energy to foragers, rendering the trade-off between staying at the current location and moving on to a different patch less critical. In the case of non-depleting patches, the decision faced by the forager would be better treated with the prey model, described in the next section (§1.3.3).

Charnov's (1976a) marginal value theorem shows that the solution to this problem is for foragers to occupy a patch until the instantaneous (or marginal) rate of energy gain that the patch offers drops below the average rate of energy gain in the environment as a whole. This solution suggests that foragers must be sensitive both to the gain function of the patch (i.e. how quickly it depletes) and the statistics of patches in the environment (i.e. how long it will take to travel to the next patch). It follows then, that all else being equal, if some change in the environment suddenly increases the distance between patches, foragers should increase their occupancy time at each patch. Similarly, a step-wise decrease in travel time should result in a decreased time spent in each patch.

1.3.3 The prey model

The prey model answers the question of which potential food sources in an environment are worth pursuing. The relevant trade-off here is between how much energy an item offers and the item's handling time. Upon encountering a potential food item, the forager must decide whether the time spent capturing and consuming it might be better spent searching for something else. A prey item that is both difficult to consume (because of its speed, or physical defenses like a shell, for instance) and low in energy value (because it is small, for example) is probably best ignored in lieu of searching for something else. For an environment with multiple types of prey items, the solution to the prey problem, then, would be a set of rules that tell the forager whether, on encountering a particular prey item, it should be accepted (i.e. pursued, captured, and consumed) or rejected (i.e. ignored to resume searching).

The prey choice situation for n types of prey in an environment can be modeled as follows:

$$R = \frac{\sum_{i=1}^{n} p_i \lambda_i e_i}{1 + \sum_{i=1}^{n} p_i \lambda_i h_i} \tag{1.14}$$

where for each prey type i, p_i is the probability of accepting the type on encounter, e_i is the energy gained from accepting the item, and h_i is the type's handling time (Stephens and Krebs, 1986).

To maximize the rate of intake, equation 1.14 is computed as the probability of including each prey type is varied from 0 to 1 (i.e. the equation is differentiated with

respect to p). Doing so shows that for each prey type, the rate-maximizing probability of acceptance is either one or zero. Consequently, prey items should either always be accepted or always rejected on encounter, a result referred to as the "zero-one rule."

An intuitive solution, then, is to sort potential prey items by their short-term rates of energy gain (energy divided by handling time, a quantity also referred to as a prey type's *profitability*), and, one by one change their acceptance probability in equation 1.14 to 1, beginning with the most profitable item and working down through less profitable types. The point at which the computed rate R decreases with the addition of a prey type represents a cut-off, and prey items with profitability at or below this threshold should never be accepted on encounter (Stephens and Krebs, 1986).

In addition to the zero-one rule, another somewhat counterintuitive prediction of the prey model is that the probability of accepting an item is independent of the item's density in the environment. However, as items are only accepted if their associated opportunity cost is less than the cost of search, it matters not how frequently or infrequently a forager comes across a particular prey type; it will either always be a good deal or always be a bad deal, as determined by equation 1.14. As Stephens and Krebs (1986) note, the prey model makes no claims about which types of prey a forager should seek out; search in the prey model is a random process, and the probability of accepting a prey type is an "encounter-contingent" policy, meaning that it specifies what should be done once a particular prey type has actually been found, not how an animal ought to conduct its search.

1.3.4 Behavioral ecological perspectives on intertemporal choice

Because foraging theory models are concerned with how animals should trade off energy intake and costs over different timescales, it is clear that questions of intertemporal choice are implicit in the models detailed in previous sections. Unlike the previously-described models of intertemporal choice, the assumption of delay discounting is not built into foraging theory. Instead, delay preferences arise from considering how decisions about exploiting particular food sources influence the forager's long-term rate of energy intake. Thus, while delay discounting as a mechanism is not present in foraging theory models, behavioral ecologists have long recognized that ethological factors may give rise to a preference for immediacy.

Issues of investment and the uncertainty of the future are important considerations in food-choice decisions made by foraging animals. Achieving food and nutrients more quickly ensures that animals maintain good health and contributes both to future foraging success and general evolutionary fitness. The issue of investing immediate gains is intimately related to that of future uncertainty. In natural settings, the likelihood of collecting delayed outcomes can vary widely, depending on factors such as intra- and inter-specific competition, predation, and environmental conditions that are both difficult to predict and beyond the forager's control. To the extent that species are equipped to store energy over long intervals (i.e. by converting excess calories into fat stores, or caching food supplies in secure locations) immediate energy gains can function as a buffer against the uncertainty of future food prospects.[7]

It is worth noting here that while uncertainty undoubtedly influences temporal preferences over long timescales, it is unlikely to play a large role in the delays animals encounter in laboratory studies of decision making, which are typically less than a minute. In the majority of experimental settings, there is no collection risk at all; animals always receive the food option they select after the delay passes. Although one could argue that animals might nevertheless incorrectly perceive uncertainty in food delivery, given enough trials in a particular context it would seem likely that subjects could learn that nothing will intervene to prevent food delivery during long delays. Further, Henley and colleagues (2008) tested animals in an experiment that introduced probabilistic "interruptions" during delays, resulting in subjects not receiving some rewards. They found that the rate of interruptions did not influence discounting rate, suggesting that uncertainty about receiving delayed rewards is a minor factor in determining temporal preferences in experimental settings.

Because behavioral and physiological mechanisms for food collection and long-term energy storage vary widely across species, another factor worthy of consideration is whether the natural history of an organism might constrain its delay preferences. Stevens

[7] We acknowledge that these examples differ slightly from what economists would consider investment. Except for the case of using immediately available energy to bolster foraging efforts, and thereby obtain more food than would have otherwise been possible, strategies for saving presently-available nutrients for future use generally do not result in accumulation of additional worth in the way that a financial investment might.

and colleagues (2005), for instance, tested marmosets and cotton-top tamarins in an intertemporal choice scenario. The two species showed divergent delay preferences despite sharing a close phylogenetic relationship, with marmosets exhibiting a greater willingness to wait for larger, delayed options.

The authors argued that these preferences reflect the feeding ecology of the two species (Stevens et al., 2005). Marmosets in natural settings obtain a large fraction of their food by extracting sap from trees, a laborious process that requires long periods of time spent scratching bark. In contrast, tamarins feed primarily on insects, a task that rewards quick reflexes over a more ponderous approach to food collection. Similar studies comparing chimpanzees and bonobos have revealed similar, species-specific adaptations that influence behavior in laboratory tests of decision making (Rosati et al., 2007). Thus, when constructing general models of delay preferences, it is important to consider the extent to which evolution may have resulted in species-specific behavioral strategies. This consideration is particularly relevant for laboratory tests of intertemporal choice, in which the decision problems we test animals on may vary in both subtle and substantial ways from natural foraging contexts.

1.3.5 Rate models and intertemporal choice

Using the long-term rate maximization approach exemplified by the prey and patch models, it is possible to predict behavior in intertemporal choice scenarios. Bateson and Kacelnik (1996) performed an experiment designed to test the predictions of long-term rate maximization in the standard self-control intertemporal choice scenario, offering starlings a choice between smaller-sooner and larger-later rewards, with conditions set such that choosing the larger-later option yielded a greater long-term rate of food intake. Starlings in the experiment violated the predictions of long-term rate maximization both by their preference for the smaller-sooner option that offered a lower long-term rate, and their insensitivity to changes of the length of the inter-trial interval, a convenient proxy for modeling the travel time between food options. In contrast, animals tested in conditions that mimic foraging scenarios have consistently behaved in a manner consistent

with long-term rate maximization models.[8]

The contrast between experiments testing animals in foraging situations and choice tasks like Mazur's (1987) self-control adjusting-delay paradigm raises the possibility that the algorithms animals use to achieve long-term rate maximization in foraging scenarios are somehow deficient in other circumstances. Part of the puzzle seems to lie in precisely which time interval subjects are sensitive to in laboratory experiments. A long-term rate maximization model in its simplest form (i.e. eq. 1.9) divides all energy that is obtained in an experiment by the total experimental time, whether that time is spent waiting for food delivery (the pre-reinforcement delay) or "traveling" between food items (the inter-trial interval). However, as alluded to in the discussion of Bateson and Kacelnik's (1996) experiment, laboratory studies of intertemporal choice show that in many situations subjects seem to attend only to the pre-reinforcement delay, showing little or no change in behavior as delays following reward delivery are manipulated (e.g. Green et al., 1981; Logue et al., 1985; Mazur, 1989).

These results raise an important question: is the idea that animal behavior maximizes energy intake rate hopelessly misguided, or is it the case that self-control tests of intertemporal choice are asking animals to answer an ill-posed question? Stated another way, we can consider whether there exists some choice structure in which the apparent preference for immediacy that self-control experiments uncover can achieve a high long-term rate of energy intake.

An experiment by Stephens and Anderson (2001) set out to answer this question directly, by examining decisions in economically-equivalent patch and self-control scenarios. Blue jays (*Cyanocitta cristata*) were trained to perform two intertemporal choice tasks, as diagrammed in figure 1.4. The self-control condition offered a binary, mutually-exclusive choice between a smaller-sooner option (delivered after delay t_1) and a larger-later option (delivered after delay t_2). In the patch condition, subjects made a stay/go decision. After waiting for t_1 to elapse, the smaller-sooner food reward was delivered. Subjects were then free to leave the patch, and re-enter a new patch of the same type

[8] A full review of experimental results supporting foraging theory predictions is beyond the scope of the thesis. Suffice it to say, however, that foraging theory models successfully predict behavior in both laboratory experiments and field observations across an impressive range of species. See Pyke et al. (1977); Krebs et al. (1978); Stephens and Krebs (1986); Stephens et al. (2007); Stephens (2008) for further information.

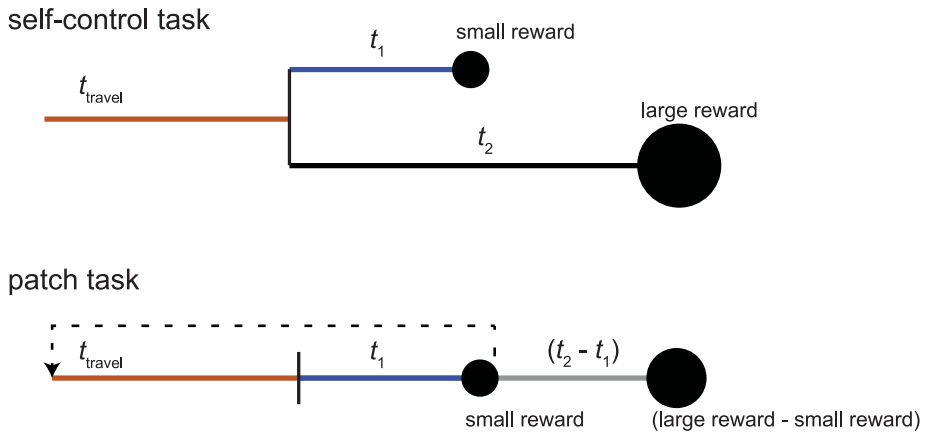

Figure 1.4: **Comparing patch and self-control decision topologies.** Stephens and Anderson (2001) tested blue jays on two intertemporal choice tasks whose options were equilibrated with respect to long-term rates of food intake, but differed in intake rates over the short term.

after the travel time/inter-trial interval passed. Otherwise, they could remain in the patch and collect more food after the remainder of the delay period ($t_2 - t_1$) expired. In this arrangement, leaving the patch after the first reward produced the same rate of intake as the smaller-sooner option in the self-control treatment. Similarly, choosing to wait in the patch for both rewards yielded the same rate of food intake as choosing the larger-later self-control option. As is clear from the diagram, however, the options varied in the short-term rate they afforded.

From the previous discussion of how animals in laboratory settings often do not attend to inter-trial intervals, one can appreciate how behavior guided by short-term rates (i.e. rates computed excluding the inter-trial interval) could lead to preference for the smaller-sooner option in the self-control treatment, but could favor staying in the patch until the second reward was delivered.

Indeed, this is what Stephens and Anderson (2001) found. Their experiment tested jays on multiple permutations of the patch/self-control tasks, in which the delays to reward, the travel time, and the reward magnitudes were varied parametrically (but in tandem for the two conditions to preserve the equality in their long-term rate outcomes). In the patch condition, birds generally followed the predictions of long-term rate maximization, remaining in the patch for the second reward when this option produced the greatest long-term rate of intake, but abandoning the patch after the first reward when

delays and amounts were such that waiting for the second reward reduced their long-term intake rate. Further, subjects' tendency to choose the larger-later option became more pronounced as the difference in long- and short-term rates increased. In contrast, compared to long-term rate equilibrated patch conditions, subjects more frequently opted for the smaller-sooner option in the self-control treatment.

These results demonstrate a subtle and important point: behavior driven by the same "short-sighted" decision rule (e.g. choose options to maximize the rate of intake over the next – and only the next – time interval) can have very different consequences for the long-term rate of food intake depending on the topology of the choice situation. Examining only the self-control treatment, it would be tempting to conclude that subjects in the experiment were "impulsive" given their maladaptive preference for the smaller-sooner option. Data from the patch treatment, on the other hand, suggests that animals maximized their long-term intake rate as prescribed by foraging theory models.

Stephens and Anderson's (2001) results suggest a peculiar sort of rigidity in subjects' behavior. Having extensive experience with the patch and self-control conditions, it seems possible that subjects might have noticed that their decision making strategy was more effective in one situation than the other. Given that behavior in the patch scenario did vary as a function of the difference between long- and short-term rates, it is clear that animals could discriminate the differences in rate that the options provided. However, the data suggest that jays did not use this information to adapt their behavior and earn more food in the self-control task, raising an interesting question: is it possible to design intertemporal choice tasks that shift behavior to favor options with a greater long-term rate when preferences would otherwise skew towards short-term options?

Pearson and colleagues (2010) showed that changing how animals attend to task components influences their decision making strategies. They tested primates (*Macaca mulatta*) in the self-control setting (choice between larger-later and smaller-sooner rewards) and inferred discounting rates. In one version of the task, information about the length of pre-reward delays was cued by displaying a progress bar that shrunk as time passed. In a task variant, both pre- and post-reward delays were cued with the progress bar. Providing an explicit cue indicating the length of post-reward delays caused subjects to increase their selection of the larger-later option, which resulted in discounting rates decreasing by nearly an order of magnitude compared to the version of the task

that lacked a post-reward cue. While one might be tempted to invoke a species difference in comparing these results to Stephens and Anderson's (2001) experiment, a related primate study found that subjects choosing between un-cued delayed outcomes failed to associate their decisions with a contingent post-reward delay, and that although subjects were sensitive to the length of post-reward delays, their behavior suggested they vastly underestimated their length (Blanchard et al., 2013). A similar study found that cuing pre-reward delays common to both the larger-later and smaller-sooner options caused pigeons to discount long delays more slowly (Calvert et al., 2011), consistent with the idea that subjects' understanding of task structure is an important determinant of their decisions. A study involving rats found similar effects (Zeeb et al., 2010).

These experiments highlight the importance of understanding how subjects use the information that is available to them in decision making tasks. Foraging theory models require "complete information" about the structure of the world in order to correctly determine the rate-maximizing set of behaviors. Experimental data suggest that subjects do not always learn the statistics of a task and apply this information in the way we might hope they would. Although it is convenient (and economically-equivalent) to model activities such as traveling or handling prey items by imposing a delay, it is not clear that animals react to such delays in the same way they do to their real-world counterparts. Further, factors such as attention, cue saliency, and the level of subjects' engagement with the task are all likely to influence behavior and should temper the conclusions we draw.

1.3.6 Summary

Foraging theory offers a unique perspective on understanding temporal preferences that complements thinking from economics and psychology. The assumption of evolution as an optimizing agent that drives behavior towards maximizing energy intake rates has spurred the development of mathematically-rigorous, well-specified models that can be solved to find the best foraging strategies for a particular environment.

Experiments testing animals in foraging situations (choosing how long to exploit a patch, for instance) are generally well-described by long-term rate maximization models. However, behavior from intertemporal choice experiments, like Mazur's (1987) adjusting-delay task, have been more difficult to account for with long-term rate models. Recent

work suggests that this discrepancy arises from a mismatch between strategies that maximize long-term intake rate in natural settings and the structure of self-control intertemporal choice tasks. A rule that maximizes short-term intake rates may also maximize long-term rates in natural settings, but fail to do so in some laboratory tests.

1.4 Reinforcement learning approaches to intertemporal choice

Reinforcement learning (RL) is determining which actions to take based on experience alone (Sutton and Barto, 1998). RL differs from supervised learning, traditionally the domain of machine learning algorithm development, which treats the problem of learning from exemplars or templates. Supervised learning algorithms operate with a "ground truth" example that can be compared to incoming data streams in order to effect pattern recognition or classification. Although useful in some scenarios, supervised learning is insufficient when organisms lack a template to work from.

The problem of reinforcement learning is typically framed in the context of an abstract decision making *agent* that interacts with its surroundings by taking actions, and receives feedback from the world in the form of reward, or reinforcement. The RL framework closely approximates the predicament faced by a laboratory animal upon encountering decision making task for the first time; subjects are given little information about the structure of the world but are tasked with uncovering the set of behavior that results in the best possible consequences. RL algorithms describe strategies for learning optimal behavior efficiently, and with minimal information about the world beyond that gained from experience. Because RL algorithms consider not just instantaneous reward, but also reward that is available after completing a sequence of actions with temporal extent, they are inherently well-suited for modeling and understanding behavior in intertemporal choice scenarios.

1.4.1 States, transitions, and reinforcement

The RL agent's world is discretized into *states*, which encompass all the features of the environment that characterize a given situation. An important simplifying assumption is that situations in the world which are sufficiently similar are recognized as the same state despite any incidental differences that might be present. For instance, if we were

to model a Pavlovian conditioning task in which a rat undergoes tone-food pairings in an operant box, we might define three possible states: the interval between tone-food pairing trials, the interval during cue presentation, and reward delivery. At any point between pairing trials, the agent/rat would be reside in the inter-trial interval state, regardless of where precisely in the chamber the rat happens to be. Frequently it is assumed that agents identify their current world state correctly without issue; more complex models allow for state uncertainty or ambiguity in state recognition, which can have a strong influence on behavior (e.g. Daw et al., 2006; Redish et al., 2007). The collection of possible world states defines the environment, or *state space*.

Agents *transition* between states in multiple ways. In the Pavlovian conditioning example above, movement from one state to the next is controlled entirely by the world; at the end of the cue state, for example, the agent automatically moves to the reward state. In some cases, however, state transitions are caused by the agent's actions (i.e. pressing a lever in an instrumental conditioning task). Depending on the task being modeled, transitions between states may also be probabilistic. A description of how world states are connected (including, in the case of stochastic transitions, the likelihood of moving from one state to another) is called a *transition function*.

Reinforcement is a property associated with certain world states; upon entering a state that contains reward, subjects achieve a scalar, numerical reinforcement value. This conception is consistent with the psychological definition of reinforcement and the economic definition of utility. Qualitatively different reinforcers (e.g. two different types of food) are discriminable only to the extent that the magnitude of reinforcement each provides differs. Agents that have full knowledge of an environment's state space and transition function can use this information to make decisions based on outcomes (i.e. which states and reinforcers will result from actions) rather than scalar reward alone. The simplest RL models, however, assume agents lack such information; more complex, outcome-specific RL models are discussed later.

The system of states and transitions described above describes a Markov decision process (MDP). In a discrete MDP, events (e.g. observations of states, receipt of reward, action selection) occur at discrete time steps $t = 1$, $t = 2$, $t = 3, \cdots$. Agents are coupled to their surroundings by their observations of state and reinforcement, and the actions (when available) that they select. The environment determines how agents transition

between states, in a manner that may or may not depend on responses from the agent, and that may include varying levels of stochasticity. A defining feature of a Markov process is lack of memory; that is, the likelihood of entering future world states depends only on the current world state, and not the precise path of previous states by which the agent arrived there, a property that will prove advantageous for determining the agent's best course of action in any arbitrary world state. Later, in discussing RL models explicitly designed to tackle questions of discounting and intertemporal choice, we will consider semi-Markov worlds with continuous time. In this case, states have characteristic dwell times that must elapse before the agent can transition to a new state. However, the lack of history dependence (i.e. the Markov property) is maintained.

1.4.2 Value learning

In order to maximize reinforcement, agents must have some means of estimating the value of different world states and the actions afforded to them. The ability to predict the value associated with potential future states would allow the agent to direct its behavior towards high-value states in the world. Several approaches for estimating the value of states exist. Critical to all of these is the idea that a current state's value should depend on both how much reward it offers, and which future states it permits the agent to access, as these future states might in turn vary in how much reward they contain.

Bellman's method for computing state value

Bellman (1958) first described an approach now commonly used to estimate state values. Consider a simple world, in which transitions between states are deterministic and contingent on the action an agent selects in each world state. This example temporarily sets aside the question of how an agent knows which actions to take; instead, we assume the agent has a pre-defined rule for action selection, and in each state simply consults that rule to determine what to do. Such an action selection rule is called a *policy*, and describes a mapping from states to actions. An intuitive way estimating the value of the current world state at t_0 is to sum the amount of reward available in the current state ($r(t_0)$) and all the future reward that will result in in subsequent time steps. If we know the agent's policy and the reward structure of the world, we have all the information

necessary to compute the value at the current time step ($V(t_0)$):

$$V(t_0) = r(t_0) + r(t_1) + r(t_2) + \cdots \tag{1.15}$$

Because the agent's trajectory through the world's state space depends on which actions it selects, it is important to note that the quantity $V(t_0)$ is an estimate of state value particular to the agent's policy; the value at t_0 under a different policy might well differ. However, a problem is immediately apparent with this way of measuring state value. Summing over an infinite number of future states results in a state value estimate that is also infinite. This problem highlights the difficulty of computation with an unlimited temporal horizon; the infinite state value estimates that result from weighting all future rewards equally are not particularly useful to the agent. For finite chains of states, the sum terminates at the last possible state of the world (called the *absorbing state*) and is therefore bounded. Another solution to the problem of temporal horizon is discounting delayed, future rewards. This is accomplished by multiplying future rewards with an exponential discounting factor γ raised to a power that equals their delay:

$$V(t_0) = r(t_0) + r(t_1) * \gamma + r(t_2) * \gamma^2 + \cdots \tag{1.16}$$

In modeling animal behavior, the discounting approach is often preferable to the inclusion of an absorbing state, as world states will recur throughout the course of the experiment instead of progressing inexorably towards one, final state. Further, as it is usually unclear whether subjects are able to measure the time remaining in the experiment while they are performing a task, it is not unreasonable to imagine that they perceive the task as an effectively infinite (or at least very long) series of choices. The discounting approach allows agents to compute sensible value estimates over a sequential state space with a temporal horizon defined by the discounting parameter rather than the task's structure.

Equation 1.16 measures state value as current reward plus the sum of all future reward, discounted by delay. This formulation lends itself to a recursive definition of state value:

$$V(t_0) = r(t_0) + V(t_1) * \gamma \tag{1.17}$$

Here, expected future values are incorporated into the value estimate of the next state the agent will enter, discounted by one time step.

Equation 1.17 is the form of Bellman's (1958) equation most frequently used for computing state values in MDPs, and it can be solved in a number of ways to obtain a value function (an estimate of each state's expected future reward) for the world. Bellman's method is a useful computational tool, but several difficulties suggest it is unlikely that animals use such an approach for estimating state values. Foremost, Bellman's technique requires a complete and accurate description of the world (the state space and the transition function). Further, the complexity of the Bellman equation grows dramatically as more realistic conditions (e.g. large, richly-interconnected state spaces, probabilistic transitions and rewards) are simulated. While these expressions are not difficult to evaluate mathematically, they impose prohibitively-large memory demands as state spaces are scaled up to approximate realistic conditions.

Temporal difference value learning

A value function like that computed by Bellman's method would undoubtedly be useful to RL agents or animals performing a behavioral task. Ideally, agents could learn such a value function from scratch, by observing and interacting with the world. The method of temporal difference (TD) learning (Sutton, 1988; Sutton and Barto, 1998) provides an algorithm that makes this possible.

At the heart of the TD method is a comparison between how good an agent expects a state to be before it is entered, and how good the state actually is when the agent experiences it. In a stationary environment, if an agent's value function is correct, estimated state value and experienced state value will be identical. However, if the agent's value function does not match the true reward structure of the world, estimated and experienced state values will not match. The discrepancy between these temporally-successive value estimates is termed the *prediction error* (δ). After transitioning to a new state and observing its value directly, the prediction error term can be computed and used to update the agent's value estimate for the preceding state, improving the agent's value function in light of new experience. The value updating rule takes the

form:

$$V(s_t)_{new} = V(s_t)_{old} + \alpha * \delta_t \tag{1.18}$$

In equation 1.18, the prediction error term is computed as $\delta_{t_0} = r(t_0) + \gamma * V(t_1) - V(t_0)$. The learning rate α ranges from 0–1, and scales how quickly state value estimates are updated. Note that when the δ is zero, value estimates are not updated (i.e. $V(t_0)_{new} = V(t_0)_{old}$).

The TD method performs a sort of bootstrapping, iteratively adjusting earlier predictions of state values as the agents gains new information about the world. The value function that agents learn therefore reflects both the immediate reward each state provides and the (discounted) future reward each state predicts. Value functions learned with the TD algorithm converge with those computed by the more computationally-intensive Bellman approach, and importantly, the TD method learns on-line, as experience occurs (Sutton and Barto, 1998).

The method of learning from differences in predictions across time has its basis in earlier theories of Pavlovian value learning (Bush and Mosteller, 1951, 2006). The most successful of such models, proposed by Rescorla and Wagner (1972), accounts for many features of simple animal learning. The Rescorla-Wagner rule computes the associative strength of cues (conditioned stimuli; CS) as they are paired with reward (unconditioned stimuli; UCS) in Pavlovian conditioning experiments. Associative strength (V) is updated according to:

$$V(CS_i)_{new} = V(CS_i)_{old} + \alpha * (\lambda_{US} - \sum_i V(CS_i)_{old}) \tag{1.19}$$

The λ_{US} term represents the maximal associative strength that the unconditioned stimulus can support, reflecting the fact that greater rewards support more learning. The summation reflects the fact that associative strength is distributed across all the cues present in the experiment.

The Rescorla-Wagner (1972) equation assigns a finite associative strength to cues that predict reward. Once the cues present in the environment have absorbed the entirety of associative strength a particular US can provide, no further learning occurs. This updating rule is similar in many ways to the TD learning rule (eq. 1.18), but subtle

differences between the two lead to qualitatively different predictions.

The fundamental difference between the two models concerns the times at which updates occur. The Rescorla-Wagner (1972) equation updates associative strength following the delivery of reinforcers, and thereby implicitly defines the time unit of reference as whole conditioning trials, which culminate in reward. TD learning, on the other hand, updates value predictions at every time step (i.e. each state transition), whether or not reward is immediately available in the newly-entered state. Consequently, TD methods are able to update predictions on the basis of subsequent predictions (Sutton and Barto, 1998; Glimcher et al., 2008).

The TD rule accounts for many well-documented learning phenomena (e.g. Pavlov, 1927; Kamin, 1969; Rescorla, 1969), including some that Rescorla and Wagner's (1972) model cannot accommodate (e.g. second-order conditioning; Rescorla, 1973). In addition to behavioral support for TD models, there is neural evidence suggesting that brains implement a TD-like mechanism, with the phasic activity of midbrain dopamine neurons encoding the prediction error signal (δ) during learning (Montague et al., 1996; Schultz et al., 1997; Waelti et al., 2001; Bayer and Glimcher, 2005; Roesch et al., 2007). Although current understanding of whether and how RL methods are implemented in the brain remains incomplete (e.g. Dayan and Niv, 2008; Maia, 2009), the convergence of neural and behavioral evidence has led to the widespread popularity of TD algorithms for interpreting data from learning and decision making experiments (Daw, 2003; Daw et al., 2005; Glimcher et al., 2008; Kalenscher and Pennartz, 2008).

1.4.3 Instrumental action

The previous section detailed methods for estimating state values within the Markov decision framework. Realistic agents, however, do more than simply learn state values– they also act on those value estimates by making choices. RL models for action selection are built around the foundation of TD value learning, with the addition of mechanisms for transforming estimated values into actions.

The credit assignment problem

The addition of action selection complicates the problem of value learning. The agent's actions influence which state it will transition to and how much reinforcement will be

obtained on the next time step, but actions can have consequences which are not realized until many time steps in the future. When an agent earns reward, it is unclear which previous action or actions were responsible for arriving at the rewarding state. The challenge of discerning which previous features of behavior were determinants of reward is called the credit assignment problem (Minsky, 1961).

The credit assignment problem can be solved with methods rooted in the recursive form of Bellman's equation (eq. 1.17). Dynamic programming techniques describe algorithms for modifying the agent's policy to ensure that choices maximize reinforcement (Bellman, 1958; Watkins, 1989; Sutton and Barto, 1998). However, as with using the Bellman equation for value learning, dynamic programming is computationally expensive, and requires a veridical understanding of the world's structure to guarantee success.

RL methods overcome the credit assignment problem by selecting actions via state value estimates that reflect both the immediate reward in the next state and the average amount of future reward predicted by that state going forward. As with RL methods for value learning, models of instrumental action are implementable without recourse to complete knowledge of the world, but still converge on the same optimal action selection policies as dynamic programming approaches (Watkins and Dayan, 1992).

Q-learning

The Q-learning model (Watkins, 1989) solves the problem of action selection by learning the value of state-action pairs (i.e. $Q(s, a)$, the cumulative future reward predicted by taking action a in state s). The procedure for updating Q-values follows the logic of TD value learning:

$$Q(s_t, a_t)_{new} = Q(s_t, a_t)_{old} + \alpha * \delta(t) \tag{1.20}$$

The delta term is computed with respect to the agent's estimate of the most valuable

action (of all possible actions A) that will be available in the next state:[9]

$$\delta(t) = r_t + \max_A[\gamma * Q(s_{t+1}, a)] - Q(s_t, a_t) \tag{1.21}$$

Action value estimates greatly simplify the problem of instrumental behavior. In each state, choosing the correct course of action requires only a means of comparing the values associated with the available choices. An obvious strategy would be choosing the action with the greatest estimated value. However, care must be taken in using such "greedy" strategies; as action values are learned iteratively by trial and error, selecting only the best-seeming action at a given moment might lead to suboptimal behavior if Q-value estimates do not yet fully reflect the true reward structure of the world.

The problem of selecting the best possible action while ensuring that all actions are sampled with sufficient frequency to ensure accurate action value estimates is referred to as balancing exploration and exploitation. A common solution is choosing actions stochastically, weighted by their value, via a "soft-max" activation function (Sutton and Barto, 1998). The probability of choosing action a from a set of n possible actions is computed as follows:

$$p(a) = \frac{\exp^{Q(a)/\tau}}{\sum_{i=1}^{n} \exp^{Q(i)/\tau}} \tag{1.22}$$

The parameter τ is referred to as the temperature, and sets the balance of exploration of the space of possible actions and exploitation of the action with the greatest estimated value. High τ reduces the agent's value sensitivity, causing indiscriminate selection of high and low value actions. As temperature limits towards zero, agents approach the greedy strategy of always choosing the action with the greatest Q-value. Between these extremes, it is possible to construct agents that usually choose the best possible action, but probabilistically select actions with lower value estimates, ensuring that action values are learned appropriately for all options in each state.

[9] A similar RL model of action selection, SARSA (state-action-reward-state-action), computes prediction errors with respect to the action the agent will actually select in the next state, rather than the best possible action available (Rummery and Niranjan, 1994). This approach learns action values for a particular, pre-defined policy, while standard Q-learning learns action values for a policy that balances exploration and exploitation. The issue of exploration and exploitation is discussed further below.

1.4.4 Extensions of RL models

The previous sections described RL models for passively learning state values (as in the case of Pavlovian conditioning) and learning the value of state-action pairs (as in instrumental behavior). These basic RL formulations have served as a starting point for modeling a variety of more complex learning and decision making situations. Here, we review several useful extensions of RL relevant to studies of intertemporal choice.

Outcome-based RL

Thus far, our discussion of RL models has been restricted to those that assume agents operate without explicit knowledge of the environment's state space and transition function. Such agents represent the average predicted value of actions, but lack a representation of the outcome each action engenders. This poses problems for modeling phenomena such as state-dependent valuation. For instance, imagine an environment with two sources of reinforcement, water and food. For a thirsty rat, the value of water should exceed the value of food. However, an RL agent that operates with only a scalar value signal cannot discriminate between these two reinforcers, and thus cannot be act to remedy specific need states like hunger or thirst. A similar problem arises in cases of devaluation (Dickinson and Balleine, 1994; Balleine, 2001), where a previously valuable outcome suddenly changes for the worse. In laboratory studies, this situation is achieved by pairing a palatable food reward with a toxin to induce illness. While animals quickly learn to avoid the poisoned food option, outcome-insensitive RL agents must update value iteratively, and consequently require many experiences with the now aversive choice before they cease selecting it (Daw et al., 2005; Glimcher et al., 2008).

Outcome-specific RL agents function much the same as their simpler counterparts, but are additionally granted access to knowledge about the world's state and transition structure (Sutton and Barto, 1998). This information can be learned through experience, or gifted outright. In either case, outcome-specific agents can use forward search through the structure of the world to predict the long-term consequences of their actions; such agents represent outcomes explicitly, allowing them to achieve state-dependent decision making. The trade-off for the flexibility afforded by the outcome-specific algorithm is one of computational complexity; the ability to perform forward search through the state

space is limited by the agent's memory capacity, and is slower than simply selecting action values in each state.

Outcome-sensitive and insensitive algorithms need not be mutually exclusive. Daw and colleagues (2005) proposed a dual-controller framework in which both algorithms compete for control of the agent's behavior. Uncertainty about the value of actions arbitrates between the controllers, and the system that is most certain of action values prevails. This formulation captures the fact that animals in decision making experiments sometimes appear to use outcome-insensitive (i.e. stimulus-response) strategies, but in other situations seem to employ outcome-specific (i.e. stimulus-outcome) reasoning. The dual-controller scheme allows the agent to take advantage of the speed and computational simplicity of outcome-insensitive systems when the task is well-learned and uncertainty is low, but revert to the slower, more flexible outcome-sensitive algorithm when changes in the structure of the world render stimulus-response predictions inadequate. The authors suggested a neural implementation of this dual-controller model, identifying dorsolateral striatum with the outcome-insensitive system and prefrontal cortex with the outcome-based component.

Motivational effects in RL models

One of the primary strengths of the RL modeling approach is its abstract nature and pared down framework. The MDP structure captures enough detail to model realistic decision making scenarios, and eliminates enough complexity to remain tractable. However, behavior is more than just prediction and action. For instance, both the actions that animals select and the speed with which they carry our those actions are important features of behavior. It is well established that animals perform free-operant tasks at a faster rate when they are more motivated (i.e. their deprivation state is greater). RL accounts of decision making successfully predict the ends towards which behavior is directed, but until recently, did not account for the rate at which behaviors occur.

Niv and colleagues (2006a) developed an RL model of free-operant behaviors that includes a mechanism for motivation to influence behavior. The model is based around a Q-learning agent that maximizes its long-term average rate of reinforcement (Daw and Touretzky, 2002; Daw, 2003). This formulation differs slightly from previously described RL models, in that it deals with the question of temporal horizon not by discounting

reward, but by optimizing the long-term average rate of reinforcement, akin to foraging theory models discussed previously (§1.3; Stephens and Krebs, 1986; Kacelnik, 1997; Daw, 2003). In this model, decisions are made with respect to the estimated average rate of reward options entail rather than by comparison of the discounted action values of each.

Agents in the model (Niv et al., 2006a) select both actions and the latency with which those actions are executed to maximize the rate of reinforcement. Motivational state (i.e. hunger) is controlled by directly increasing the value of reinforcers, and the model is constructed such that the optimal latency of actions depends on the average rate of reward the agent measures from the environment. In this way, the average rate quantifies the opportunity cost of time on a moment-by-moment basis. When the average reward rate is high, opportunity cost is great, and the optimal latency of actions decreases. Without some factor to balance the effect of opportunity cost on action latency, optimal behavior would default to always performing actions as quickly as possible. Accordingly, the model imposes a cost for taking actions that is proportional to the speed with which actions are executed, thereby checking the influence of opportunity cost on behavioral "vigor."

Because opportunity and action costs determine how quickly behavior evolves, the model (Niv et al., 2006a) reproduces the increased vigor with which animals respond for high-rate reward schedules. The model also offers a computational account of why deprived (i.e. motivated) animals respond for reinforcement more quickly: because hunger increases the value of food reward, the agent's estimates of action values increase with hunger, causing an increase in the predicted average rate of reward in the environment, and a consequent decrease in the best latency for actions. This mechanism explains the somewhat puzzling fact that hungry animals perform *all* actions more quickly, including actions such as grooming, which are not directed towards earning food. Because the influence of opportunity cost extends to all possible actions, when estimates of average reward are high, optimal action latency decreases even for behaviors that do not result in food delivery.

Niv and colleagues (2006a) suggest a convenient and plausible neural implementation for the motivational component of their RL model. As noted previously, it is thought that phasic dopamine release by midbrain neurons encodes a TD prediction error. Integrating

this signal over time results in a measure of average reward, a key component of the motivational RL model. One possibility, then, is that phasic dopamine drives learning via prediction errors and the more slowly-varying, tonic dopamine signal encodes an average reward rate/opportunity cost signal useful for guiding behavior. Pharmacological and lesion data are broadly consistent with this idea (Niv et al., 2006a,b; Niv, 2007).

Hyperbolic discounting in an RL framework

Temporal discounting in RL models has historically been implemented with exponential formulations, largely for computational convenience; with a constant drop in discounted value for each additional unit of delay, exponential formulations discount consistently across an arbitrary number of intervening states. Hyperbolic discounting functions, on the other hand, decrement discounted value at different rates for different length delays. While one can simply substitute hyperbolic discounting directly into the previously-described equations (e.g. eqs. 1.16 & 1.21), the overall pattern of discounting that such models exhibit depends on the granularity of the task's state space.[10] If the delay period is modeled as a single state in which the agent must dwell before reward delivery, "one-step" hyperbolic discounting performs as expected (Redish, 2004; Kalenscher and Pennartz, 2008). However, if the delay period is broken down into multiple units (e.g. unique states are created for each second, or half second, or quarter second, etc. of the delay), the agent's net discounting diverges from a hyperbolic curve (Kurth-Nelson and Redish, 2009; Redish and Kurth-Nelson, 2010).

Kurth-Nelson and Redish (2009) developed a Q-learning model that correctly implements hyperbolic discounting over arbitrary state space configurations. Their model distributes variables across a population of "micro-agents" (μAgents). Each μAgent operates independently, computing estimates of action values according to its own unique, exponential discounting parameter value. The "macro-agent" can ultimately only take a single action; action selection is therefore based on Q-value estimates averaged across the population of μAgents.

Although individual μAgents discount exponentially, the average of a population of exponential curves is hyperbolic. In consequence, with a sufficiently large population of

[10] RL models of intertemporal choice typically employ continuous-time, semi-Markov worlds in which states specify the duration for which agents must remain in the state before a new transition is allowed.

μAgents, the macro-agent exhibits hyperbolic discounting as an emergent property of the distributed representation. The delay tolerance of the macro agent is adjusted by altering the distribution of discounting rates in the population of μAgents. The authors tested their model on a simulation of Mazur's (1987) adjusting-delay task and found that it conformed to the predictions of hyperbolic discounting (Kurth-Nelson and Redish, 2009). The μAgent model was also used to demonstrate the necessity of hyperbolic discounting functions for Ainslie's precommitment effect (§1.2.3; Kurth-Nelson and Redish, 2010). It is worth noting that the μAgent model is effectively a generalization of previously-described two-exponential, quasi-hyperbolic discounting models (Phelps and Pollak, 1968; Laibson, 1997); the discounting predicted by such models could be approximated in the μAgent framework by setting the number of μAgents to two.

Neuroimaging data suggests that groups of neurons in the striatum might provide an anatomical substrate for a distributed discounting system like the μAgent model. As humans performed a decision making task involving delayed rewards, Tanaka and colleagues (2004) observed a gradient of value-correlated fMRI BOLD signals corresponding to different rates of exponential discounting across the striatum, suggesting that anatomical modules in the striatum computed exponentially-discounted values independently, with different discounting rates. As in the μAgent model, action selection achieved by averaging value estimates across striatal modules could give rise to the hyperbolic discounting humans often exhibit (Tanaka et al., 2004; Schweighofer et al., 2006).

1.4.5 Summary

RL models provide an iterative, computationally-tractable method of estimating the value of different world states and the actions they might afford. Rooted in both animal learning theory and machine learning traditions, RL models offers an abstract but powerful means of simulating decision making in both simple and complex settings. Correlates of basic RL model variables have been identified in the brain, which, along with behavioral considerations, suggests that temporal difference methods of value estimation capture something of the underlying neural bases for these processes. With relatively straightforward extensions to basic RL approaches, agents can be engineered to exhibit complex features of behavior, including state-specific valuation, latent learning, and motivational influences on operant responding. Distributed discounting approaches suggest

a way in which true hyperbolic discounting can be implemented in RL agents.

Chapter 2

The intertemporal foraging task

We designed a behavioral task to test intertemporal choice in rats. The primary aim of the *intertemporal foraging task* was to examine subjects' temporal preferences in a setting that mimicked the topology of foraging decisions in natural settings, allowing us to compare behavioral models of intertemporal choices from economics, psychology, and behavioral ecology. In this chapter, the structure of the task is described, and the rationale for its design is detailed. We present data demonstrating the face validity of the task for measuring rats' delay preferences. Portions of the data included in this chapter were published in Wikenheiser et al. (2013).

2.1 Task design and rationale

2.1.1 Apparatus

Rats performed the foraging task on an elevated circular track (track width = 10 cm; track diameter = 80 cm). The track was surrounded by distal visual cues that remained stable throughout the course of the experiment. Three automated food pellet dispensers (Med associates) were mounted to small platforms spaced evenly along the track's perimeter (fig. 2.1a). An overhead camera recorded subjects' position via a light-emitting diode (LED) "backpack," a cloth strip containing a battery-powered LED fastened around the rats' bodies with velcro. Online position tracking was processed with a Cheetah 160 data acquisition system (Neuralynx). The task was controlled using

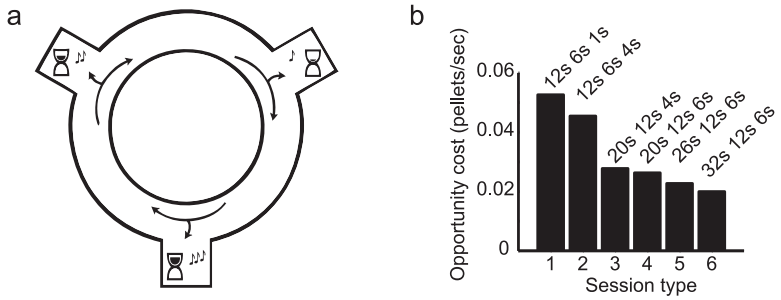

Figure 2.1: **The intertemporal foraging task.** (a) Rats foraged for food on a circular track equipped with three food pellet dispenser sites. Tones cued subjects to the delay associated with each feeder location. (b) Task session types differed in the lengths of delays. Together, the delays determined the session's environmental richness, or opportunity cost. Figure reproduced with permission (Wikenheiser et al., 2013).

custom Matlab (MathWorks) software.

2.1.2 Subjects

All experimental and animal care procedures complied with National Institutes of Health guidelines for animal care and were approved by the University of Minnesota Institutional Animal Care and Use Committee. Male, Fisher-Brown Norway hybrid rats ($n = 10$ for the behavioral portion of the experiment) aged 10–14 months were subjects for the experiment. Rats were maintained on a 12 hour light-dark cycle. Behavioral sessions occurred at the same time each day ($\pm$ 1 hour). Rats were handled for 7 days prior to beginning task training to familiarize them with the experimenter. During the handling period subjects were acclimated to consuming 45 mg sucrose pellets (Test Diet) which were used as food reward in the foraging task. During task performance rats were food deprived to no less than 80% of their unrestricted feeding weight. Water was always freely available in the home cage.

2.1.3 Training and task

Before beginning the experiment rats performed a training task to familiarize them with the environment and apparatus. In daily, 30 minute behavioral sessions, subjects were trained to run unidirectional laps around the track and collect food reward at each of the feeder sites. Attempts to run backwards were physically blocked by the experimenter;

rats quickly learned that running backwards was futile, so blocking subjects was never necessary after the training phase of the experiment. As subjects entered a 15 cm zone centered on each feeder site, two food pellets were dispensed. The feeder site then became inactive (i.e. would not dispense further food pellets) until subjects returned to the site on the next lap around the track. After rats ran more than 30 laps for 3 consecutive sessions the training phase was complete and task performance began.

Task sessions occurred daily, for a duration of 30 minutes. Rats earned sucrose pellets by distributing their time between the three feeder locations, each of which dispensed food pellets only after a delay period passed. Sessions featured a long, medium, and short delay, each of which were associated with a particular feeder site; this association remained fixed within a behavioral session, but delays were counterbalanced across feeder locations across sessions. All feeder sites delivered the same amount of food (2 sucrose pellets, each 45 mg).

The delay period began when rats entered a 15 cm zone centered on each feeder site. Entry into this zone was signalled by a tone sequence (200 ms pulses, repeated once per second). The tone's frequency was proportional to the duration of the feeder delay at that site. At this point, subjects made a stay or go decision. If rats remained in the feeder zone for the duration of the delay period, food pellets were dispensed. Otherwise, rats were free to leave at any time. Following either food pellet delivery or premature exit from the zone, the tone sequence ceased, and the feeder site became inactive until rats arrived there on their next lap around the track.

Six sets of delays were selected (fig. 2.1b) for testing, producing six session types. Because the delay sets ranged in duration from short to long, session types varied in opportunity cost. Session types with relatively short delays simulated a "rich" environment, in which food was generally easy to come by, and the opportunity cost of time high. Session types with longer delays approximated "leaner" environments, in which the average rate of reward was lower. Throughout the course of the experiment, subjects experienced each session type four times, in pseudo-random order (the same session type was never repeated on consecutive days).

2.1.4 Features of the task

The foraging task poses an intertemporal choice problem (i.e. trading off value and delay) in a topology akin to natural foraging decisions. Instead of selecting between two mutually-exclusive options, subjects made a sequence or stay/go decisions; choosing to "accept" a particular feeder delay in no way precluded accepting the next delays in the sequence. However, the fixed session length, and the "closed economy" forced subjects to allocate their session time wisely. Because the task was entirely self-paced, factors such as the inter-trial interval were controlled by how quickly subjects left feeders and traveled to the next site. The free-operant design entails losing a measure of experimental control, but is also more realistic, as encountering a food item in nature does not necessarily impose a fixed interval in which foragers cannot find another food item.

While the task structure has features which more closely approximate natural foraging than other choice paradigms, the task contingencies are simplified compared to those in natural environments. For instance, statically assigning delay lengths to feeder sites for the duration of each session greatly simplifies the problem of searching for resources, as subjects could quickly learn which items were deterministically available at particular locations in the environment, and could count on them to remain stable for the duration of a session, a luxury not often afforded to foragers in the wild. Similarly, the tone cues provided subjects with another source of information about the delays at each location. Fixing delays in a spatial location and cuing them with tones reduced ambiguity about the values of the feeder sites to a perhaps unrealistic extent. However, this feature also made for a stronger test of foraging theory models, which (in their simplest formations) assume that foragers have complete information about the environment, and that recognition of different food items is instantaneous and certain.

2.2 Validity of the task

If the foraging task measured delay preferences, we would expect behavior on the task to show several characteristics. First, subjects should make binary choices (i.e either leaving a site shortly after encountering it, or waiting for the entire delay to pass). The alternative possibility is that subjects employed some delay-based heuristic, such as waiting at each site for a fixed number of seconds and then leaving if reward is not

delivered.

After learning the locations of delays on the task, we would expect subjects to construct a strategy for earning reward given the set of available delays, and then use that strategy for the remainder of the session. Alternatively, if subjects were incapable of associating spatial locations or tones with delay durations, behavior might never settle into a stable strategy, instead varying erratically on every encounter with a feeder site.

Finally, we would expect delays to exert a consistent influence on subjects' behavior. The vast majority of previous work has shown that delay generally decreases the value of outcomes in some way; if subjects did not understand the structure of the task or the contingencies of food delivery, however, delay may have failed to influence preferences in a way that we can interpret.

2.2.1 Rats made binary decisions

Across all behavioral sessions, rats ran on average 36.2 laps per session (range: 24 – 85 laps). A one-way ANOVA found an effect of session type on number of laps ($F_{5,234} = 7.60$, $p < 0.0001$), and Bonferroni post-hoc testing reveled that subjects ran significantly more laps during type one sessions (pairwise comparisons between all other types were not significant). The effect size of this difference was not substantial, however, with type one sessions comprising on average 44.3 laps, approximately 8 more laps than other session types.

For each visit to a feeder location, we computed the amount of time subjects waited at the site, beginning when rats entered the feeder zone and ending either when the delay period passed and food pellets were delivered or subjects exited the zone in pursuit of a different location before the delay had expired. Figure 2.2 shows histograms of waiting durations for each delay length. Columns of the histogram are normalized to sum to one, and thus display the proportion of encounters with each delay that subjects waited for a given duration.

Across all delay lengths, rats generally either left shortly after entering the feeder zone or waited for the full feeder delay to pass. The median time spent waiting on skip trials was 0.61 s. To place this value in context, with a typical running speed of approximately 50 cm/s, it would take rats a minimum of approximately 0.2 s to pass through the 12 cm feeder zone. A Kruskal-Wallis test revealed significant differences in

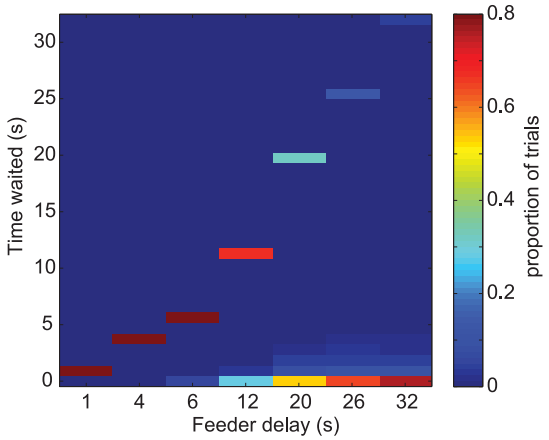

Figure 2.2: **Rats made binary choices.** Normalized histograms of waiting durations are plotted for each of the delay lengths subjects encountered on the task. Subjects tended to either leave early during the delay or wait out the entire delay period.

the median wait duration on skip trials at the seven different delay values ($\chi^2_{6,7403} = 13.90$, $p = 0.03$); however, post-hoc Bonferroni testing revealed no significant pairwise differences across groups. Together, these data suggest that rats made binary accept or reject choices on encountering feeder sites.

2.2.2 Rats showed stable, non-random behavioral strategies

Rats performed session types of the task in pseudo-random order, and delay locations were randomized across feeders to ensure that subjects did not develop idiosyncratic preferences for particular feeder locations. Consequently, rats were forced to spend the initial portion of each session learning which delays were located at which feeder sites. Efficiently harvesting reward from the task depended critically on rats' ability to remember which delays were available in the current session and where they were located in space. An important question, then, is whether rats developed consistent, stable behavioral strategies once they learned the feeder-delay pairings for a particular session. An alternative possibility is that learning of delay-location pairings did not occur, or occurred very slowly, precluding rats from planning behavioral strategies, and leading to more stochastic behavior.

Beginning from any feeder site, in choosing where to go next subjects effectively had three options: they could stop at the next feeder site in the sequence, they could skip the

next site and stop at the feeder after that, or they could skip the next two locations and run an entire lap around the track, returning to wait for food at the site they departed from.[1]

The structure of the task suggests that behavioral strategies could be captured by a transition matrix that records the probability of journeys beginning from one feeder and ending at another feeder. If rats used consistent behavioral strategies on the task, probability in the transition matrix would be concentrated over a few particular trajectories. If, however, rats' behavior never settled into a consistent strategy, probability in the transition matrix would be spread more evenly over many trajectory types. Examining the variability (i.e. entropy) of this transition matrix over the course of behavioral sessions would reveal whether subjects ran consistent patterns of visits between feeder sites, or whether behavior was largely unstructured.

We constructed 3 feeder $\times$ 3 feeder transition matrices for each behavioral session. Matrices were updated following each visit to a feeder site in which subjects waited for the delay to pass and received food reward. Rows of the matrix denoted the feeder at which journeys began and columns of the matrix represented the feeder at which journeys ended. For example, if a rat departed from feeder one and then ran to and waited at feeder three, the matrix cell representing a journey from feeder one to feeder three was incremented by one. Values in the matrix were normalized by the total number of journeys between feeders; consequently, each entry in a matrix represents the probability of trajectories starting and ending at particular locations. Transition matrices were updated cumulatively following each wait visit in the session, allowing us to examine the behavioral entropy at each time point within sessions.

We computed the entropy of transition matrices over the course of all behavioral sessions (fig. 2.3). For each behavioral session, we also randomly permuted rats' skip or wait decisions and computed a transition matrix in the same way for this randomized pattern of choices, allowing us to compare rats' behavioral strategies to random behavior

[1] This description of possible choices is a simplification. There was nothing that stopped subjects from skipping more than two sites in a row. For instance, subjects might have departed from feeder one and skipped the next five feeders in the sequence, eventually stopping at feeder one again, after running two entire laps without waiting for food delivery. Such behavior would not represent a particularly efficient food gathering strategy, but may have occurred if rats did not understand that delays were fixed at spatial locations. Rats never showed such behavior; across all behavioral sessions, the maximum number of consecutive skips was two, suggesting rats understood that bypassing a site would not result in it offering a different delay on the next visit.

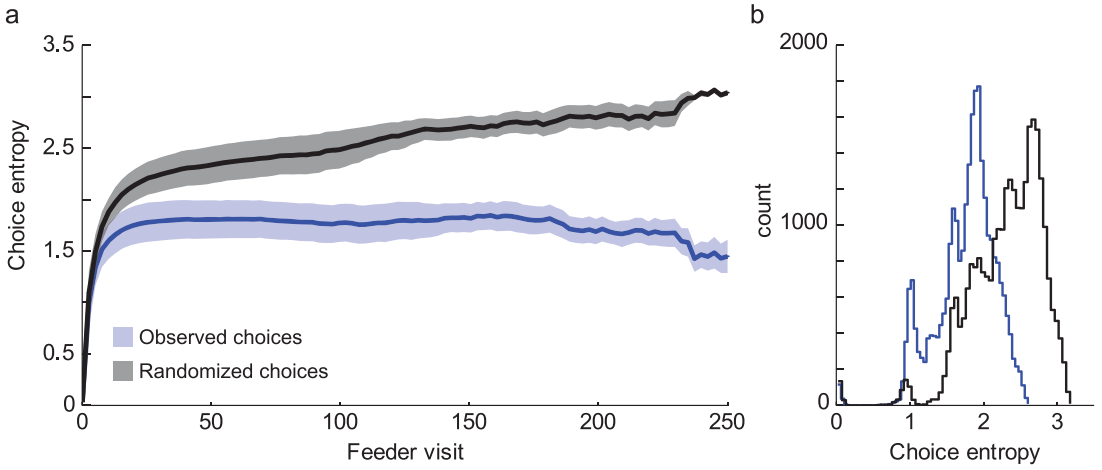

Figure 2.3: **Rats used stable, non-random strategies.** (a) The entropy of rats' decisions reached a steady value early within behavioral sessions, after approximately 5 exploratory laps, and remained stable there after. The entropy of randomly-permuted decisions increased over the course of behavioral sessions. Shaded regions indicate standard deviation (n = 56,737 feeder encounters across 240 sessions from 10 rats). (b) Observed rat behavior was significantly more orderly (i.e. less entropic) than randomly-permuted decisions.

that preserved the relative frequency of skip and wait choices.

The choice entropy of observed behavior settled to a stable level after approximately 15 feeder visits (i.e. five laps; fig. 2.3a). The choice entropy of randomized behavior was significantly greater than that of observed behavior (fig. 2.3b; $U_{56737} = 5.15x10^8$, $p < 0.0001$; Mann-Whitney's U-test).

2.2.3 Rats were sensitive to feeder delays

Temporal preferences were measured by computing the fraction of encounters with each feeder site in which rats waited for the delay period to expire and received food (probability of waiting; p_{wait}). Although rats varied to some extent in their delay tolerance, subjects were generally less willing to wait for food pellets at longer delays. Figure 2.4a shows p_{wait} values across all delays for four individual subjects. Data were fit with sigmoid curves. Curves fit for all subjects are shown in figure 2.4b.

We also measured p_{wait} separately for individual session types (fig. 2.5). Across all session types, p_{wait} decreased as a function of feeder delay.

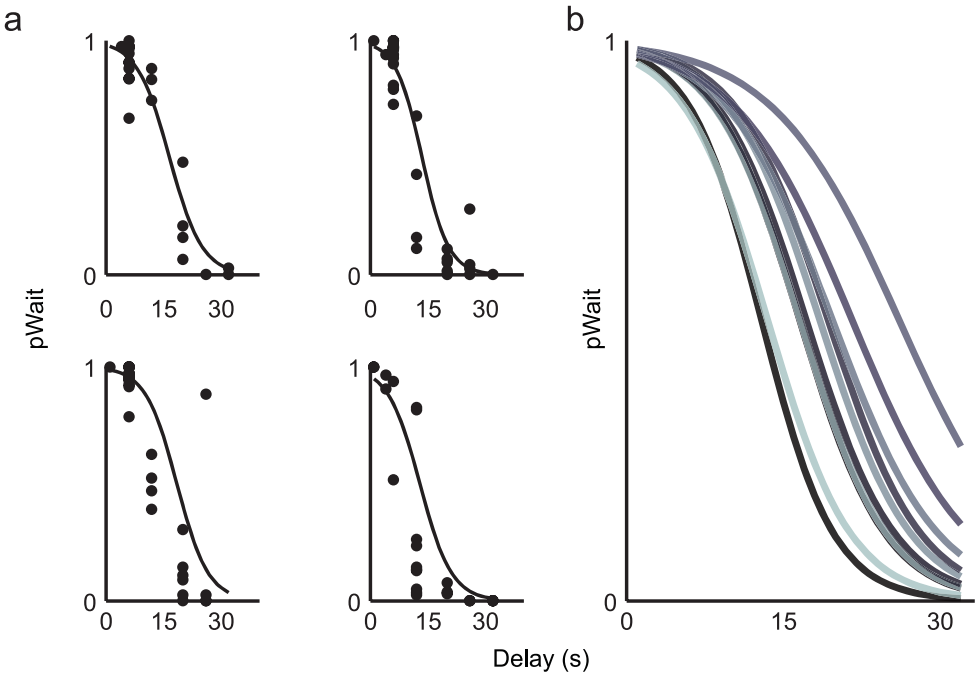

Figure 2.4: **Rats were sensitive to feeder delays.** (a) The fraction of trials in which subjects waited for food delivery (p_{wait}) is plotted against delay duration. Data are from four individual rats, pooled across all session types ($n = 24$ sessions/rat). (b) Sigmoid curves fit for all rats ($n = 10$) show that while rats varied in their delay tolerance, p_{wait} generally decreased with increasing delay. Figure reproduced with permission (Wikenheiser et al., 2013).

2.3 Summary

Together, these data suggest that the intertemporal foraging task successfully elicited delay-based decision making from rats. Rats' decisions were consistent with binary choices, ruling out the possibility that behavior was guided by a simple waiting-threshold rule. Within each behavioral session, subjects settled into consistent, non-random behavioral strategies. In agreement with a large body of previous research, delay influenced subjects' preferences in a predictable manner, with long delays decreasing the likelihood of subjects accepting offers.

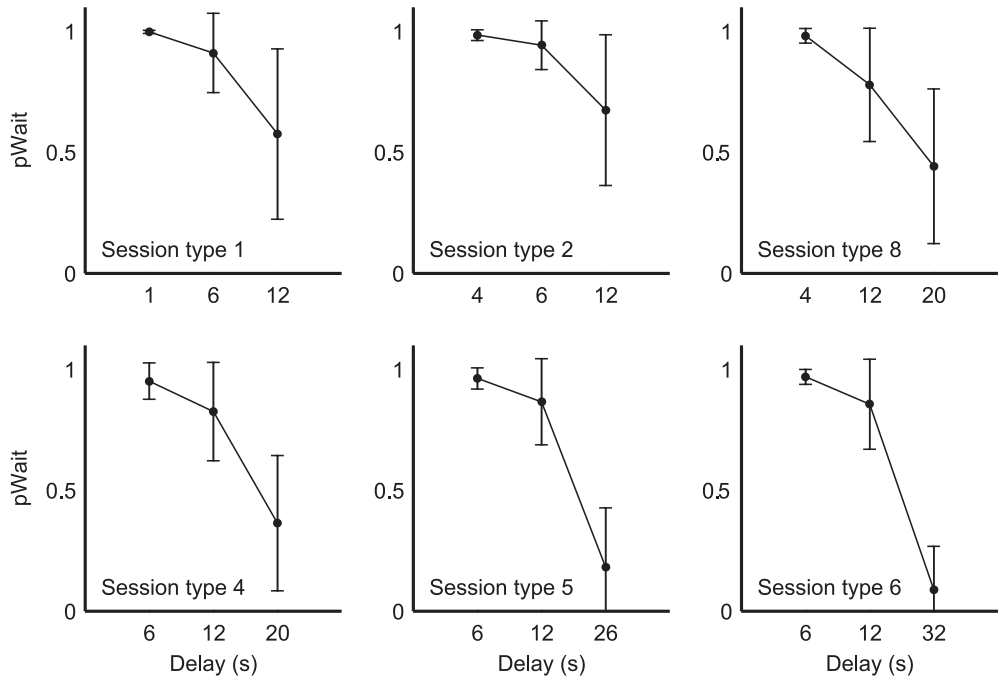

Figure 2.5: **Delay sensitivity across session types.** Rats were sensitive to feeder delays in all session types; p_{wait} values decreased with increasing delay duration. Error bars indicate the standard deviation ($n = 40$ sessions for each session type). Figure reproduced with permission (Wikenheiser et al., 2013).

Chapter 3

Behavior on the intertemporal foraging task

In this chapter, we compare rats' behavior on the foraging task to three behavioral models pertinent to intertemporal choice: the matching law, temporal discounting, and foraging theory rate maximization. Portions of the data included in this chapter were published in Wikenheiser et al. (2013).

3.1 Matching law

The matching law (Herrnstein, 1970) states that animals adjust their level of behavioral investment in an option to match the fraction of their total earnings (income) that option provides. Although developed to account for choice between concurrent VI schedules, matching strategies have been observed in a wide variety of circumstances (Herrnstein, 1997). Matching is a near-optimal strategy in situations where subjects choose between time-dependent outcomes that, once primed, remain set to deliver food until the subject's next response (§1.2.2). The foraging task does not contain such structure, as every visit to a feeder required rats to endure the entire delay period before food was delivered. Nevertheless, given evidence that matching may be an innate, "pre-programmed" resource-gathering strategy (Gallistel et al., 2007), it is plausible that subjects may have used a matching strategy on the foraging task.

We computed subjects' fractional investment in each option (the proportion of time

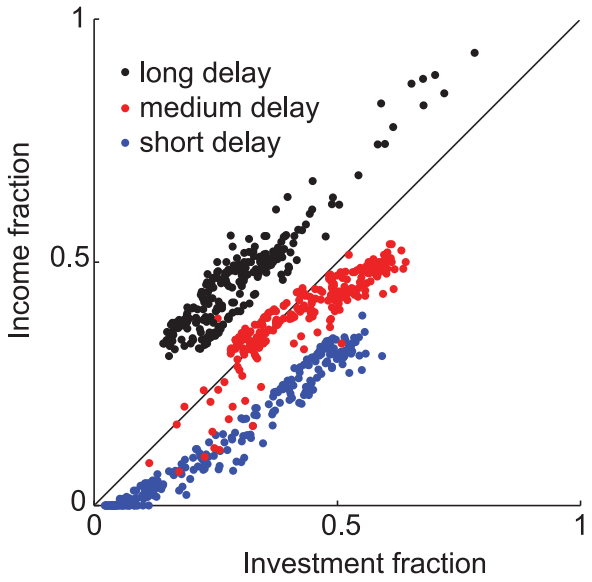

Figure 3.1: **Rats did not match behavioral investment to income.** Matching law predicts that the behavioral investment fraction for a given option should equal (match) the fraction of income earned from that option. When income and investment fractions are plotted against each other, matching predicts that points should fall along the unity line. The behavior we observed deviates substantially from this prediction. Figure reproduced with permission (Wikenheiser et al., 2013).

spent waiting at a site) and plotted it against the fractional pellet income subjects earned from that option (fig. 3.1). The matching law predicts that points should be distributed along the unity line. Most observations did not conform to this prediction. While some points fell close to the matching prediction (particularly those of the medium delay option), the majority lie off the diagonal. This suggests that subjects did not match their behavior to the level of income gained from feeder sites in the task.

We examined several other aspects of behavior for characteristics of the matching strategy. Matching strategies typically result in exponentially-distributed durations of time spent at each option. We computed visit durations (the amount of time subjects spent in the feeder zone) for each encounter with a feeder site. Because feeder sites in our task were associated with different delay lengths, we normalized visit durations by each site's delay to view the data on the same scale (fig. 3.2a). In our task, subjects sometimes waited for the delay period to expire, but in other cases left the site before the delay passed. Accordingly, the overall distribution of visit durations had two peaks—one resulting from trials in which subjects waited for food delivery and another resulting from

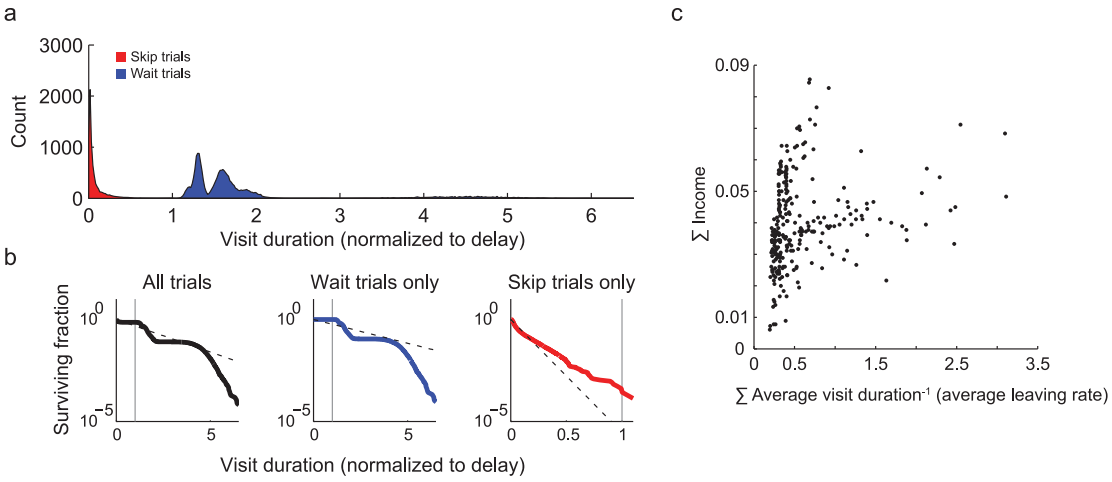

Figure 3.2: **Visit durations were inconsistent with matching law.** (a) Matching strategies are characterized by exponentially-distributed visit durations. Visit durations (normalized to the delay length at each site) are plotted, with distributions split based on whether subjects waited for food delivery (blue) or left the site before the delay period expired (red). (b) Survivor curves for normalized visit durations for all visits, wait trials only, and skip trials only. The vertical grey line marks the time of reward delivery. The dashed black line is the survival curve of an exponential distribution with its mean matched to the visit duration distribution. (c) In matching tasks, the sum of the average leaving rates (the reciprocal of the average visit duration) at all potential food sites increases in proportion to the sum of the income across all food sites. Figure reproduced with permission (Wikenheiser et al., 2013).

trials in which subjects left the site early.

To fairly test whether visit durations were distributed exponentially, we considered the complete set of visits (all trials), and also separated trials based on whether subjects waited for food delivery (wait trials only) or left before earning food (skip trials only). We computed survivor curves for visit durations separated in this way, and compared them with survivor curves calculated for an exponential distribution with the mean matched to the mean visit duration of the sample in question (fig. 3.2b). Examining these plots shows substantial divergence between the expected exponential survivor curves (dashed lines) and the visit duration survivor curves (solid lines), suggesting that visit durations were not distributed exponentially.

To ensure that normalizing and pooling data did not affect our analyses, we tested distributions of visit durations from each site within each session for exponentiality using a Lilliefors test of the null hypothesis that sample data derive from an exponential distribution with an unspecified mean. Again, we separately considered visit durations

from all trials, skip trials only and wait trials only. To ensure that enough samples were present to accurately assess the distribution's shape, only duration distributions with at least 10 or more samples were tested. We found that in the vast majority of cases there was sufficient evidence to reject the null hypothesis that data were distributed exponentially ($P < 0.05$; all trials: 99%; skip trials only: 96%; wait trials only: 100%).

Finally, we examined the relationship between the sum of the average leaving rates (the reciprocal of the average visit duration for each site in a session) and the sum of incomes for each site. Matching predicts a linear relationship between these quantities; however, we observed no clear relationship in our data (fig. 3.2c).

Taken together, these analyses suggest rats did not use a matching strategy while performing the foraging task.

3.2 Temporal discounting models

Exponential and hyperbolic discounting models are often thought of as the gold standard for modeling preferences in intertemporal choice scenarios. We used a RL (§1.4; Sutton and Barto, 1998) framework to test whether rats' behavior was consistent with exponential or hyperbolic discounting. We implemented Q-learning models that performed exponential (Watkins, 1989) and hyperbolic discounting (Kurth-Nelson and Redish, 2009), and tested these models on a simulation of our foraging task. Simulated agents decided whether to wait for food delivery by comparing the Q-value associated with staying at the current feeder with the Q-value of proceeding to the next site. We chose the RL approach to model the foraging task because simpler, static models require assumptions about the potentially infinite ways in which subjects might have compared options (i.e. comparing feeder_n vs. feeder_{n+1}, feeder_n vs. feeder_{n+1} and feeder_{n+2}, etc.). Because RL models learn the expected value of actions (i.e. stay vs. go) rather than the value of feeder sites themselves, these models offer a simpler, computationally-tractable means of predicting the behavior that would result from delay discounting on our task.

The behavior of the Q-learning agents depended largely on two parameters: the discounting rate (γ) and the action selection temperature (β; eq. 1.22). The γ parameter controlled the rate with which value fell off as a function of delay, while the β parameter dictated the agent's value sensitivity (Sutton and Barto, 1998). Small β values resulted in

a strongly value-sensitive agent that tended to exploit its knowledge of the environment by strictly choosing actions with the largest Q-value. Larger β values favored exploration, leading agents to occasionally select actions with lower Q-values to ensure that their action value estimates were correct. We tested models on the task by systematically varying γ and β parameters over a wide range of values, and recording the behavior (i.e. the p_{wait} values for each feeder site) that resulted. Each parameter combination was simulated 10 times, and model behavior was averaged across repetitions. By comparing model behavior with subjects' actual behavior, this approach allowed us to ask whether any combination of β and γ values could reproduce the behavior that rats showed on the task.[1]

To change the delay tolerance of the macro-agent, we altered the distribution of micro-agent discounting rates by raising each value in the population to an exponent ranging from 0.01 to 100; distributions skewed toward high γ values resulted in slower discounting, while a preponderance of low γ values resulted in faster discounting. Consequently, the macro-agent, unlike the exponential model, is not characterized by a single discounting rate.

Model behavior across the parameter space is shown in figure 3.3 for exponential and hyperbolic models. For every pairwise combination of β and γ parameters, we computed the mean squared error (MSE) between observed and model-predicted p_{wait} values for each behavioral session (fig. 3.4). MSE values were constrained between zero (indicating a perfect match between model and observed behavior) and one (indicating the maximum possible difference between model and rat behavior). Error levels were generally high across the parameter space, indicating a poor match between model-predicted and actual behavior. Moreover, the lowest MSE values occurred in an extreme region of the parameter space, corresponding to agents that seem unrealistic from a behavioral perspective; models with such large γ values show little sensitivity to delay, in contrast with rats' behavior on the task (fig. 2.4). These data show that discounting models fit rats' observed behavior best only when discounting is effectively absent, suggesting that temporal discounting (be it exponential or hyperbolic) does not provide a compelling

[1] A final model parameter, the learning rate, controlled the speed with which simulated agents updated their estimates of action values. For both exponential and hyperbolic discounting models, the learning rate was fixed at a moderate value (0.15), and behavior was examined after simulating 2000 transitions, when Q-value estimates had stabilized.

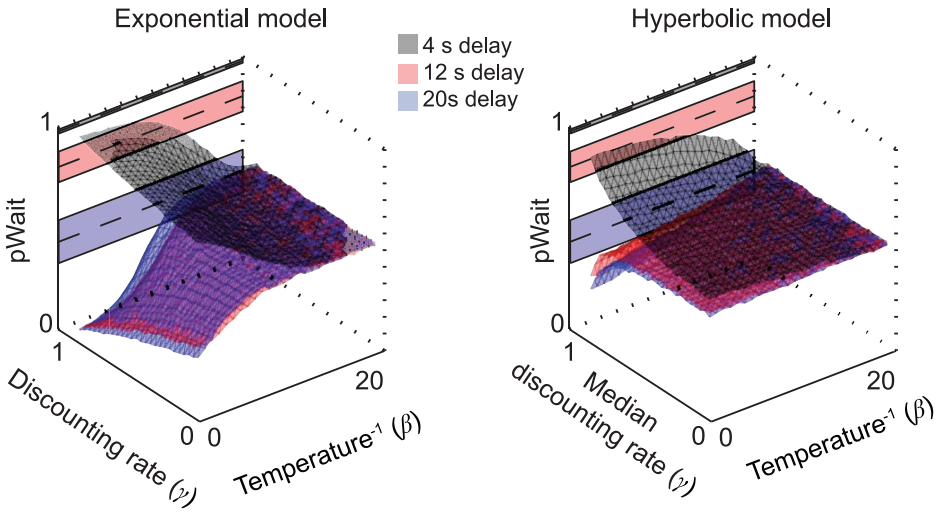

Figure 3.3: **Temporal discounting model behavior.** The behavior of exponential (left) and hyperbolic (right) temporal discounting agents for one session type of the foraging task is plotted across the parameter space of β and γ values. For comparison, subjects' actual behavior is projected behind the surfaces (mean p_{wait}; shaded regions indicate 95% confidence intervals around the mean). Because the hyperbolic discounting micro-agent model is not characterized by a single discounting rate (delay tolerance being determined instead by the distribution of discounting rates across the population of micro-agents), the median discounting rate of the micro-agent population is plotted instead. Figure reproduced with permission (Wikenheiser et al., 2013).

explanation of rats' strategies on the task.

3.3 Rate maximization models

Foraging theory rate maximization models suggest that subjects should tailor their behavior to maximize the long-term rate of food intake. As discussed in §1.3.5, experimental evidence suggests that foraging animals in some cases maximize their long-term rate of intake, but in other decision making tasks use short-term rate strategies that may or may not result in long-term rate maximization, depending on precisely how subjects parse the inter-trial interval. We therefore considered both long- and short-term rate models.

Both rate models were derived following Stephens and Krebs' (1986) development of Charnov's (1976a; 1976b) prey model formulation. Each feeder location was modeled as a unique prey type, with the location's delay as a proxy for handling time. Models differed in their treatment of travel time. The short-term rate model did not include a

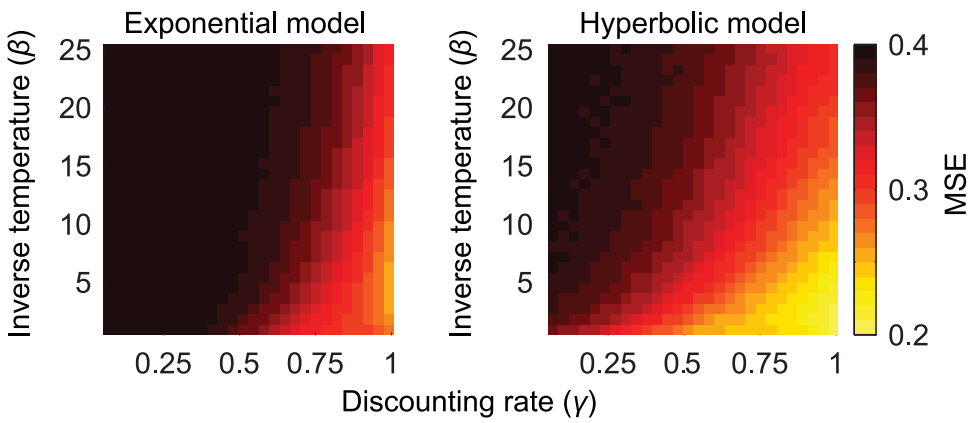

Figure 3.4: **Comparing models and observed behavior.** To test whether any parameter combinations produced behavior similar to that of subjects, we computed the mean squared error (MSE) between model predictions and observed behavior for all parameter combinations and averaged across sessions ($n = 240$ sessions). For both models, the best model fits were achieved with parameter combinations that correspond to little or no temporal discounting. Figure reproduced with permission (Wikenheiser et al., 2013).

handling time term, computing rates of food intake over only the feeder delay time (i.e. the interval between arriving at a feeder site and receiving food pellets):

$$R_{short} = \frac{\sum_{i=1}^{3} p_{\text{wait}_i}}{\sum_{i=1}^{3} p_{\text{wait}_i} \text{delay}_i} \tag{3.1}$$

The long-term rate model was identical, but for the inclusion of the travel time, T_{travel}, in the denominator:

$$R_{long} = \frac{\sum_{i=1}^{3} p_{\text{wait}_i}}{T_{\text{travel}} + \sum_{i=1}^{3} p_{\text{wait}_i} \text{delay}_i} \tag{3.2}$$

The mean travel time between individual feeders was 2.34±0.16 s (mean±standard deviation; $n = 28{,}106$ journeys between feeders). Thus, the lap travel time included in the long-term rate model was 2.34 s × 3 feeders ≈ 7 s.

For each session type, we used equations 3.1 and 3.2 to compute the maximum possible rate obtainable by finding the permutation of p_{wait} values that maximized R. Then, subjects' actual p_{wait} values were substituted into each equation to determine the rate of intake associated with their behavior under each model. The rates of intake that rats earned were normalized to the maximum possible rate to compute the achieved

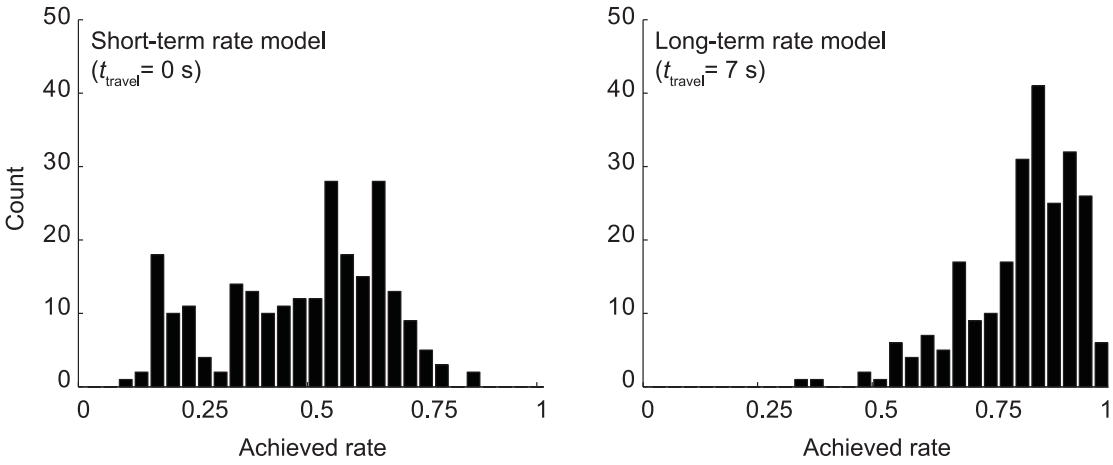

Figure 3.5: **Achieved rates.** Distributions of achieved rates (the fraction of the maximal possible rate that rats earned) for all sessions ($n = 240$) under the short-term (left panel) and long-term (right panel) rate models.

rate (i.e. how well rats performed on the task compared to how well they could have performed by adopting the rate-maximizing strategy; Pompilio and Kacelnik, 2010) for each behavioral session. Histograms of achieved rates for short- and long-term rate models are plotted in figure 3.5.

The rate-maximizing strategy for session types under the short-term rate model (eq. 3.1) was to accept only the short-delay item in each session type, and to ignore long- and medium-length options. Rats did not follow this strategy; subjects nearly always waited for both the short- and medium-length options, and often accepted a sizeable fraction of long-delay items (fig. 2.5), resulting in a decreased rate of intake. Accordingly, with respect to the short-term rate model, subjects' achieved rates were substantially less than the maximum possible rate (mean achieved rate±standard deviation = 48±18%; $n = 240$ sessions).

The long-term rate formulation (eq. 3.2) predicted that subjects should accept short- and medium-delay sites and reject long-delay feeders. Subjects' behavior was much closer to following this behavioral strategy, and under the long-term rate formulation rats earned a greater fraction of the maximum possible rate (mean achieved rate ± standard deviation = 83±17%; $n = 240$ sessions). These data suggest that subjects chose behavioral strategies to approximately maximize food intake rates computed over both feeder delays and travel time between feeders, as predicted by the foraging theory

long-term rate maximization framework.

3.4 Modeling deviations from rate maximization

Rats did not fully rate maximize under either of the rate models, as in many session types rats accepted long-delay feeder sites, contrary to the long-term rate-maximizing strategy of only accepting short- and medium-length delay options. To better understand how subjects arrived at the strategies they selected, we developed a version of the foraging theory rate model that predicted the deviations from rate maximization we observed. Rats' propensity to wait for long delays was the fundamental discrepancy between observed behavior and the long-term rate model (eq. 3.2) predictions. One potential interpretation of this behavior is that rats perceived some cost associated with rejecting feeder sites. To capture this notion quantitatively, we modified equation 3.2 to include an aversion parameter, A, representing an unwillingness to reject potential food options upon encounter:

$$R_s = \frac{\sum_{i=1}^{3} p_{\text{wait}_i} - \sum_{i=1}^{3}(1 - p_{\text{wait}_i})A}{T_{\text{travel}} + \sum_{i=1}^{3} p_{\text{wait}_i}\text{delay}_i} \tag{3.3}$$

Here, the subjective rate of food intake (R_s) decreases with instances of skipping feeder locations, in proportion to the A parameter. Larger A values result in the model exhibiting greater aversion (i.e. paying a greater cost) for skipping feeder sites. Hereafter, we refer to equation 3.3 as the "rejection-averse" rate equation, to contrast it with the standard long-term rate equation (eq. 3.2).

The addition of the A parameter strongly impacted the task's reward structure. We used the rejection-averse rate equation to calculate R_s for all possible strategies in all session types. Figure 3.6 shows how R_s varied across strategies with increasing values of A. As A increased, a larger volume of strategy space produced relatively high rates of food intake. With A fixed at zero, subjects' behavior (marked with black dots; fig. 3.6) fell well outside of the high intake rate regions of strategy space; however as A increased, the profitable region of strategy space shifted to encompass the behavior rats showed on the task. This suggests that subjects may have been behaving in accordance with rate maximization as defined by the rejection-averse equation. To test this idea, we assumed

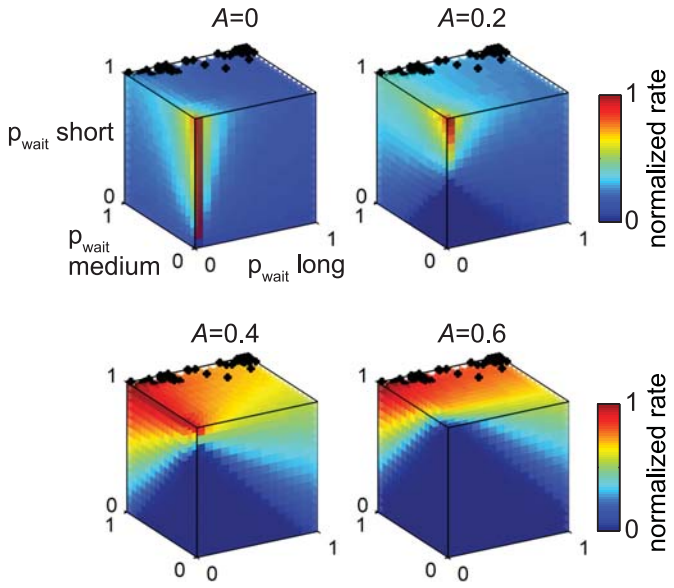

Figure 3.6: **The A parameter influenced the reward structure of the task.** We calculated the subjective rate of reward (R_s) for all possible behavioral strategies using equation 3.3, and colored each region of strategy space with the corresponding rate, normalized to the maximum possible value. With increasing values of A, high intake rates shifted to different regions of the strategy space. When $A = 0$, subjects' behavior (marked with black dots, $n = 40$ sessions) fell far from the strategies that earned high rates of food intake. However, as the region of high-rate strategies shifted with increasing values of A, subjects' behavior fell in more profitable regions of strategy space. Data in this figure were computed with T_{travel} set to zero (as in equation 3.1) to more clearly depict the influence of A on achieved intake rates. Figure reproduced with permission (Wikenheiser et al., 2013).

that subjects chose strategies according to equation 3.3, and found the value of A that maximized the achieved rate of food intake for each session. Rate-maximizing A values were significantly greater than zero ($p < 0.0001$, $n = 240$ behavioral sessions; ranksum test). Behavior governed by the rejection-averse model earned a substantial fraction of the maximum possible rate (mean achieved rate$\pm$standard deviation $= 97 \pm 5\%$).

Demonstrating that subjects' behavior nearly maximizes R_s (with the appropriate A parameter) shows that there exists a subjective valuation system (i.e. eq. 3.3) that can account for rats' behavior on the task. This suggests that rats might have monitored their rate of food intake in light of a subjective aversion to abandoning potential food sources (eq. 3.3) rather than the cost-free perspective (eq. 3.2).

The A parameter fit to subjects' behavior varied across session types. The best-fitting A value correlated positively with the opportunity cost of time ($p < 0.0001$, R^2

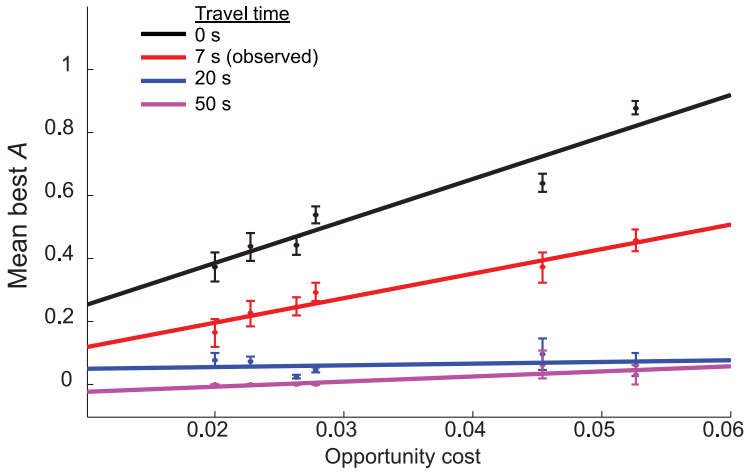

Figure 3.7: **The best fitting A parameter value varied with opportunity cost.** We found the best-fitting A parameter for each behavioral session ($n = 240$ sessions) by maximizing R_s in equation 3.3 with different travel time values. The best-fitting A value was positively correlated with opportunity cost for short travel times (including the observed travel time of approximately 7 s). At longer travel time values, subjects' behavior approached rate maximization without inclusion of the A parameter, so best-fitting A values were generally low, and did not vary appreciably with opportunity cost. Error bars indicate standard error of the mean.

$= 0.22$; $n = 240$ behavioral sessions, $T_{travel} = 7$ s), which differed across session types (fig. 2.1). This correlation suggests that rats' sensitivity to rejection costs was not a static quantity, but was instead influenced by the reward statistics of the environment. With increasing opportunity cost, rats grew more averse to abandoning potential food items.

3.5 Discussion

These data suggest that rats performing the intertemporal foraging task did not use matching or delay discounting strategies, but instead made choices that approximately maximized their long-term rate of food intake. Delay discounting models, for many years the dominant means of evaluating and modeling intertemporal decisions, were unable to account for the preferences subjects exhibited. This result contributes to a growing body of work that finds temporal discounting models inadequate for explaining temporal preferences in certain contexts (Stephens and Anderson, 2001; Pearson et al.,

2010; Hayden et al., 2011; Blanchard et al., 2013). Although data from two-alternative, forced-choice tasks are well fit by discounting curves, these results cast doubt on the psychological reality of temporal discounting in more complex decision making tasks, especially those involving sequences of decisions and distinct foreground and background components.

Rats performing the foraging task approached rate maximization under the long-term rate model described by equation 3.2. The long-term rate model computes the rate of food intake over all relevant delays in the task, including time spent traveling between feeder locations. Interestingly, previous laboratory studies have consistently demonstrated that animals are insensitive or under-sensitive to task delays analogous to travel time (§1.3.5). In our task, rather than simply imposing a delay to simulate travel time, rats were forced to physically travel between feeder sites. Requiring physical travel may have made travel time a more salient aspect of the task, causing subjects to consider long-term, rather than short-term rates in choosing their strategies. In addition to time, physical travel between sites also entails energy costs. Our findings suggest that at least in some cases, virtual search costs simulated by delays may not be equivalent to the actually moving through an environment in pursuit of food.

While our data suggest a rate-maximization strategy guided rats' behavior on the foraging task, as noted above, a vast body of experimental work has demonstrated that discounting curves account for intertemporal choice data in many contexts. An interesting question, then, is whether decision makers can flexibly switch between discounting- and rate-based strategies depending on the context of the decision. Two-alternative, one-shot decisions between delayed outcomes might well be solved by mentally discounting and then comparing outcome values to arrive at a decision. It is possible, however, that sequential choice problems involving more than two outcomes are too difficult or time consuming to solve via discounted value comparisons. Such scenarios might instead favor strategies that maximize food intake over either short or long timescales, depending on how the task is cognitively parsed by decision makers.

Subjects did not fully maximize their intake rates under either model, earning on average approximately 20% less food than they might have by adopting the rate-maximizing strategy. Several factors may have contributed to this. An incomplete understanding of the task structure and contingencies may have influenced behavior in unpredictable

ways. For instance, although food delivery was deterministic (i.e. always occurred if subjects waited for the entire delay period to pass), it is possible that animals nevertheless considered reward delivery somewhat uncertain. This may have resulted in subjects waiting at long delay locations to gain further information about the likelihood of reward delivery, leading to a more accurate understanding of the task's reward structure. Other species actively seek reward-related information in a behavioral task, even when this information cannot influence their earnings (Bromberg-Martin and Hikosaka, 2009). A similar effect in rats might account for their over-sampling of long delays.

Perceived uncertainty could also have influenced preferences in other ways. If, for instance, rats perceived an offer of food at their current location as somehow more likely to be delivered than food at future sites, such beliefs could have given rise to a heuristic where "a bird in the hand" is valued more highly than hypothetical future reward, despite differences in delays. It is also possible that subjects were simply not sensitive to reward rate, and were instead tracking their food earnings and adjusting their behavior according to some other, unknown metric.

A further possibility is that some subjective, psychological factor influenced rats' valuation of feeder sites on the task. To explore such an effect, we modified the foraging theory prey selection rate equation to include a cost for rejecting potential prey items. This "rejection-averse" model closely matched rats' behavior, suggesting that it captured an aspect of the decision making process rats used on the foraging task.

What is the nature of the perceived cost that influenced rats' behavior on the foraging task? One possibility is that A represents some general movement or movement initiation cost. However, because the physical dimensions of the foraging task apparatus were unchanged across session types, any parameter related generally to movement costs should have remained constant across session types, in contrast to our findings (fig. 3.7). This suggests that A cannot be fully explained by movement costs or other general energetic expenditures.

We considered whether the A parameter might map on to some bias known to affect human decision makers (Freidin and Kacelnik, 2011). One such bias is the sunk cost fallacy, an economic error in which willingness to continue pursuit of an option is influenced by past investment in that option, rather than anticipated future returns (Aronson, 1961; Arkes and Blumer, 1985). In economic terms, considering sunk costs is

irrational, as investments committed to a course of action cannot be recovered. Nevertheless, in many cases decision makers show a stronger preference for options they have invested resources in, despite the poor long-term consequences this might entail (Arkes and Blumer, 1985; Arkes and Ayton, 1999).

How might the sunk cost effect manifest on our task? Consider the decision a subject faces upon arriving at the long-delay feeder site: a forward-thinking, rate-maximizing rat would skip that feeder, proceeding to a shorter-delayed site instead, and thereby enjoying a greater overall rate of food earnings. In contrast, a rat making its decision with respect to the cost already spent in getting to its current position would be more likely to wait out the delay, despite the consequent decrease in overall food intake rate. Thus, in our task, the extent to which subjects were sensitive to sunk costs is indexed by how frequently they accepted feeder sites that decreased their overall food intake. The A parameter in equation 3.3, then, could be considered a session-by-session measure of sunk cost sensitivity on the foraging task, where larger A values indicate a stronger influence of sunk costs on decision making.

The A parameter fit to subjects' behavior varied systematically across session types, and was positively correlated with the opportunity cost of the session. Opportunity cost quantifies the average value of, given the density of reinforcement available in the environment (Niv et al., 2006a). When environmental resources are abundant, opportunity cost (i.e. the cost of allocating time to an activity) is high. When resources are scarce, less reward per unit time is at stake, and opportunity cost is low. Our data conflict with the normative prediction that increasing opportunity costs ought to "invigorate" behavior (Niv et al., 2006b). In sessions where the value of time was greatest (high opportunity cost), subjects were most willing to wait out long-delay options (high A values). Conversely, when opportunity cost was low (and subjects stood to lose little by accepting low-rate items), rats' aversion to rejecting feeder sites was lower.

The influence of sunk costs can explain why A was positively correlated with opportunity cost. In addition to energy, rats also invested time in traveling between and waiting at feeders. Although the distance between food options was fixed, perceived differences in the value of time across session types would result in rats' subjective valuation of a consistent time investment increasing as opportunity cost grew. When opportunity cost was high, the subjective behavioral investment (i.e. the sunk cost) would have seemed

greater. For decision makers succumbing to the sunk cost fallacy, reluctance to abandon potential food items would grow with increasing opportunity cost, consistent with our observations (fig. 3.7).

An influential account of why humans are swayed by retrospective investment suggests that sunk costs affect behavior because decision makers inappropriately overgeneralize an aversion to wasting valuable resources (Arkes, 1996; Arkes and Ayton, 1999). While it is generally a good strategy to avoid squandering valuable resources, misapplication of the waste-aversion heuristic could lead to continued investment in a doomed venture, because sticking with a losing option subjectively validates or justifies previous investment. Complementary theories have suggested that humans attend to sunk costs in order to maintain their reputation (either to themselves or others) as self-consistent decision makers that avoid wasting retrospective allocations, leading to the puzzling scenario in which a misguided attempt to appear rational leads to demonstrably suboptimal behavior (Staw, 1976; Staw and Fox, 1977; Brockner, 1992). One feature all these theories share is a critical role for complex social, cognitive, or metacognitive explanatory mechanisms; accordingly, these theories explicitly predict that animals should be immune to biases induced by sunk costs (Arkes and Ayton, 1999).

Our findings provide evidence that retrospective considerations might influence the behavior of rats in a manner consistent with the sunk cost fallacy. Recent data from rats performing a similar foraging-like decision making task also identified a retrospective factor (namely, regret over making a bad decision in the past) that influenced future decisions (Steiner and Redish, 2014). These findings suggest that sensitivity to retrospective influences might have deeper evolutionary roots than previously suspected.

Chapter 4

The neuroscience of the rodent hippocampus

The hippocampus is a brain region that plays a critical role in cognitive processes such as learning, memory, and decision making. This chapter reviews the anatomy and physiology of the hippocampal formation. Our review focuses on the organization and function of the rodent (primarily rat) hippocampus, but parallels to other species will be noted.

4.1 Anatomy of the hippocampus

The hippocampal formation is a bilateral collection of structures that lies beneath the cerebral cortex within the temporal lobe, situated with its long axis roughly perpendicular to the long axis of the brain as a whole (fig. 4.1). Although positioned beneath the cortex, the hippocampus is not truly a sub-cortical structure; rather, it is a section of three-layered "archicortex" enfolded by neocortex. Its name derives from the similarity of a dissected hippocampus to the sea horse (*Hippocampus sp.*); however, within an intact brain the hippocampus more closely resembles a banana, curving caudally and ventrally from the septal nuclei near the midline of the brain into the temporal lobe. The long dimension of the hippocampus is accordingly referred to as its septal-temporal axis, while the orthogonal, short aspect is termed the transverse axis.

Slices made along the transverse axis of the hippocampal formation reveal an orderly, laminar collection of cells that segments into several distinct subregions differentiated by

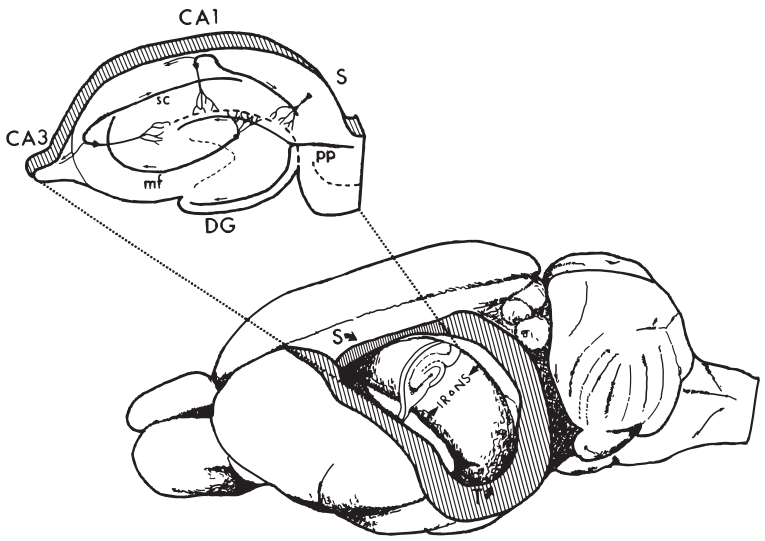

Figure 4.1: **Three dimensional arrangement of the hippocampal formation.** With overlying cortex removed, the hippocampus can been seen to extend from the midline septal region (S) to deep within the temporal lobe (T). A transverse slice (cut perpendicular to the septal-temporal axis; inset) reveals the hippocampal sub-fields CA1, CA3 and dentate gyrus (DG). The perforant path (PP), mossy fiber (MF), and Schaffer collateral (SC) projections are labelled. Figure reproduced with permission (Amaral and Witter, 1989).

connectivity patterns, cellular properties, and gene expression profiles. Although several naming conventions have been proposed for the hippocampal subregions, de No's (1934) nomenclature, which divides the hippocampal formation into the dentate gyrus and CA fields 1–3, is most commonly used today (fig. 4.1, inset). Most anatomists classify the CA fields as hippocampus proper, with the inclusion of dentate gyrus, subiculum and perhaps entorhinal cortex comprising the hippocampal formation (Amaral, 1987; Amaral and Witter, 1989; Shepherd and Koch, 1990).

The principal neurons of the CA regions are glutamatergic pyramidal cells (Amaral and Lavenex, 2007). The cell bodies of hippocampal pyramidal neurons lie within a compact layer that is divided into three segments (CA1, CA2, and CA3). Neurons in all CA fields are situated orthogonally to the layer created by their cell bodies, with apical dendrites extending towards the center of the hippocampal formation and basal dendrites radiating outwards. Pyramidal neurons in the CA2 and CA3 regions tend to be slightly larger than those in CA1 (Amaral and Witter, 1989).

The principle neurons of the dentate gyrus are granule cells. Granule cell bodies are

significantly smaller than hippocampal pyramidal neurons, and lack basal dendrites, but share a similar laminar organization. The granule cell layer encloses a less-organized region of dentate gyrus referred to as the polymorphic zone, which contains a variety of cell types whose projections remain confined within the dentate gyrus (Amaral and Witter, 1989; Shepherd and Koch, 1990; Amaral and Lavenex, 2007).

All regions of the hippocampal formation contain an overwhelming diversity of GABAergic interneurons, including both "classic" interneurons with spatially-restricted, local projections and also GABAergic neurons with diffuse, long-range connections (Shepherd and Koch, 1990; Freund and Buzsáki, 1996; Amaral and Lavenex, 2007).

The major excitatory pathway through the hippocampus, often referred to as the tri-synaptic circuit, traces a unidirectional loop through the hippocampal subfields. This system of projections is stereotyped along the septal-temporal axis of the hippocampus, and is so orderly that Ramon y Cajal (1911) was able to infer the basic flow of information through the hippocampal formation based on anatomical studies alone.

The tri-synaptic loop begins with projections from neurons in layer two of the entorhinal cortex, which are the major source of external input to the hippocampal formation (Andersen, 1975; Shepherd and Koch, 1990). This projection, the perforant pathway (so named because it passes through, or perforates, the subiculum enroute to the hippocampal formation), terminates primarily on granule cells of the dentate gyrus. Each dentate granule cell projects to multiple pyramidal neurons in CA3. The axons of granule cells form the mossy fiber pathway, so named for the distinctive, irregular appearance of these projections. CA3 pyramidal cells contact two major targets: other pyramidal neurons in area CA3 (recurrent connections), and pyramidal neurons in area CA1 (via the Schaffer collaterals; Schaffer, 1892). Finally, CA1 pyramidal cells complete the loop through the hippocampal formation, projecting axons to layer five of the entorhinal cortex and to the subiculum.

The subiculum lies between CA1 hippocampus and the entorhinal cortex, and serves as a major output of the hippocampus proper, projecting subcortically to thalamic and hypothalamic nuclei, amygdala, ventral striatum, and brain stem regions. Additionally, the subiculum contacts cortical targets, including entorhinal, perirhinal, pre-limbic, and medial orbitofrontal cortices (Groenewegen et al., 1987; van Groen and Wyss, 1990; O'Mara et al., 2001; Jay and Witter, 2004).

It is important to note that while the tri-synaptic loop is prominent, it is not the only path of information flow through the hippocampus. For instance, in addition to the Schaffer collateral inputs from CA3, neurons in the CA1 region receive a direct, excitatory projection from layer 3 of the entorhinal cortex (Amaral and Witter, 1989), and in turn project directly back to the deep layers of entorhinal cortex (Swanson et al., 1980), thus defining a more succinct loop connecting entorhinal cortex and CA1.

Finally, an important source of subcortical input to the hippocampal CA fields arises from the septal nuclei and the nucleus of the diagonal band of Broca (Lewis and Shute, 1967; Amaral and Lavenex, 2007). This projection includes both an excitatory, cholinergic component that impinges on hippocampal pyramidal cells and interneurons (Benardo and Prince, 1981), and an inhibitory component that primarily influences hippocampal interneurons (Freund and Antal, 1988). In concert with the complement of voltage-sensitive ion channels present in hippocampal neurons and local inhibitory circuitry, the septal projection gives rise to the prominent oscillatory voltage signals characteristic of the hippocampal formation, which are detailed in the next section (Green and Arduini, 1954; O'Keefe and Nadel, 1978; Nerad and McNaughton, 2006; Colgin, 2013).

4.2 Electrophysiology of the hippocampus

Recording the electrical activity of the brain offers a powerful means of understanding brain function. Over the past decades, neural recording methods have improved dramatically, to the extent that it is now routine to monitor the activity of tens to hundreds of individual neurons in freely-behaving animals (Stevenson and Kording, 2011).

4.2.1 Extracellular recording techniques

Extracellular voltage recordings of brain activity are composed of two separable component signals, action potentials and the local field potential. When an electrode tip is sufficiently close to a neuron, it is possible to record action potentials, the discrete, fast, voltage "spikes" produced by individual cells in the brain. Because action potential waveforms are fairly stereotyped within cells of a particular type, if the recording electrode is close to multiple neurons, spikes from different cells may not be discriminable.

The advent of tetrodes (Wilson and McNaughton, 1993; Gray et al., 1995) has allowed for the simultaneous recording of action potentials from many neurons. Tetrodes consist of four micro-wires wound together to form a compact bundle of electrodes. When implanted in a brain region, each wire records electrical activity, and the signal recorded by each of the four tetrode channels varies slightly depending on its precise position relative to the surrounding neurons. Consequently, comparing action potential waveforms recorded simultaneously across tetrode channels allows for *post-hoc* clustering of spikes belonging to individual neurons.

Besides action potentials, extracellular recordings also exhibit slower, broad-band voltage fluctuations that result from the summed activity of many neurons in the portion of brain tissue surrounding the electrode tip. This signal is termed the local field potential (LFP), and is thought to arise primarily from currents across dendrites (the primary site of excitatory input to neurons), although any currents across neuronal (or glial) membranes likely contribute to the LFP in some manner. The composition of the brain's extracellular space acts as a low-pass filter, allowing the slowly-fluctuating LFP signal to propagate spatially to a greater extent than the fast action potentials of individual neurons. In many brain regions, LFPs have an oscillatory character (Buzsáki et al., 2012).

Single-unit and LFP recordings provide complementary information about brain function. Spiking data allows us to examine how individual neurons in the brain represent and process information, while LFP recordings provide information about the activity of brain networks on a larger scale. In addition, important interactions between spiking and on-going LFP activity have been identified.

In the following sections, we review hippocampal network states defined by patterns of LFP activity, the characteristic spiking representations of single neurons in the hippocampus, and the interplay between unit spiking and the LFP.

4.2.2 Hippocampal network states

Neural recordings of hippocampal activity were performed as early as the 1930s (Jung and Kornmüller, 1938). Early work was limited to acute LFP recordings taken from anesthetized animals, but as recording technologies developed, it became possible to

measure LFP activity in awake, behaving subjects (Buzsáki, 2005; Buzsáki, 2006). Consequently, a large body of previous work extending to the present has been directed toward understanding the behavioral correlates and possible functional roles of distinctive hippocampal LFP patterns (Green and Arduini, 1954; Vanderwolf, 1971; O'Keefe and Nadel, 1978; Buzsáki et al., 1983; Colgin et al., 2009). These experiments have identified at least two *network states*, or information-processing modes that the hippocampus operates in, termed the *theta state* and the *large, irregular activity state*.

The theta network state is characterized by strong, rhythmic 6–12 Hz LFP oscillations (Vanderwolf, 1969, 1971; O'Keefe and Nadel, 1978; Buzsáki et al., 1983). Two distinct forms of theta oscillation (type I and type II) have been described. Type I theta occurs during "voluntary," primarily movement-related behaviors, and occupies the upper portion of the theta frequency band (e.g. 7–12 Hz). Behaviors such as running, walking, swimming, and climbing are all accompanied by type I hippocampal theta oscillations. Type II theta is a slightly slower oscillation (e.g. 5–7 Hz), and accompanies non-translational, attentive behaviors such as fearful freezing (Seidenbecher et al., 2003), preparation of a motor response (Lenck-Santini et al., 2008), attending to a conditioned stimulus (Holmes and Adey, 1960), or pauses preceding an important decision (Johnson and Redish, 2007).

In addition to differences in behavioral correlates and dominant frequencies, type I and II theta are further distinguishable by pharmacological manipulations and the brain structures responsible for generating each (Kramis et al., 1975). The theta network state also occurs during rapid eye movement sleep (O'Keefe and Nadel, 1978).

When theta oscillations are absent, the hippocampal LFP exhibits arrhythmic voltage fluctuations, referred to as large, irregular activity (LIA; Vanderwolf, 1971; O'Keefe and Nadel, 1978). Behavioral correlates of the LIA network state include inattentive wakefulness, consummatory behaviors, and grooming. Slow-wave sleep is also accompanied by LIA (Vanderwolf, 1971; O'Keefe and Nadel, 1978; Buzsáki, 1982; Buzsáki, 1989). The disorganized voltage fluctuations characteristic of LIA are punctuated by transient, population bursts of spiking that engage large numbers of neurons in the hippocampus. These short bursts of heightened synchrony are called sharp wave ripple (SWR) complexes, because of the distinctive LFP oscillation associated with them (O'Keefe and Nadel, 1978; Buzsáki et al., 1983; Buzsáki, 1989). Apart from firing during SWR events,

the activity of hippocampal neurons during LIA is generally low.

4.2.3 Hippocampal place cells

Based on a long history of lesion and pharmacological studies, it is well established that the hippocampus is involved in cognitive processes related to learning, memory, and decision making (O'Keefe and Nadel, 1978; Cohen and Eichenbaum, 1993; Redish, 1999). In rodents, damage to or inactivation of the hippocampus profoundly impairs spatial decision making, certain forms of delay-based conditioning, and tasks that require animals to track relationships between sequences of cues. Humans with damage to the hippocampus often become greatly limited in their ability to form new long-term, episodic memories. Additionally, hippocampal damage impairs the ability of humans to use previously-established memories to episodically imagine potential future events.

The commonality between these deficits is an inability to link together disparate features of the environment in a sequential order that reflects the structure of the world (Cohen and Eichenbaum, 1993; Eichenbaum et al., 1999; Eichenbaum, 2004). Successful performance of spatial memory tasks requires that subjects understand connections between disparate locations in the environment in order to plan routes between their current position and areas that reward might be available. Conditioning tasks require animals to form temporal links between predictive stimuli and their consequences. Similarly, episodic memory involves constructing a narrative that connects people, places, events, etc. in a sequential, temporal context (e.g. I met a friend at the coffee shop, we enjoyed tasty espresso, and then we went for a walk in the park). In the same vein, imagination can be thought of as using previously-learned linkages between features of the environment to generate plausible, but never-experienced episodes based on the way the world has worked in the past (Schacter et al., 2007, 2008; Buckner, 2010).

Together, the behavioral deficits that result from hippocampal dysfunction in both humans and other species strongly suggest that the hippocampus functions as a relational memory system that both learns connections between important stimuli, but also reconfigures, or inverts these learned relationships to construct predictions about the future based on previous experience.

The discovery of spatially-tuned pyramidal neurons in the CA fields of the rat hippocampus (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978) has provided a

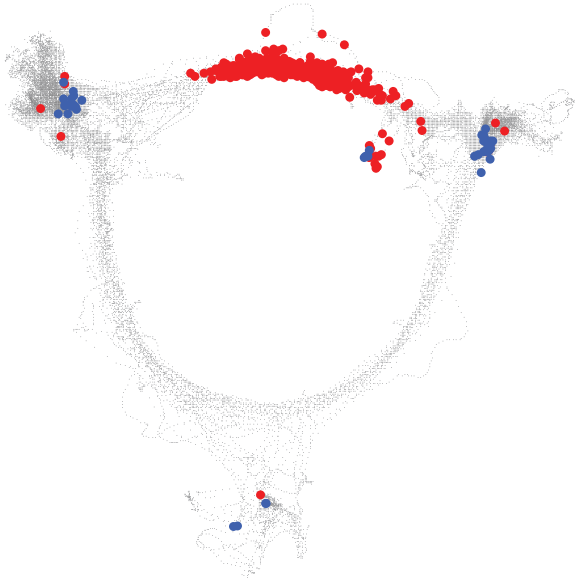

Figure 4.2: **A hippocampal place cell.** The activity of a single pyramidal neuron in the hippocampus recorded while a rat performed the foraging task is plotted. The rat's direction of movement was counter-clockwise. Position tracking data taken over the course of the behavioral session is plotted in black, and the animal's location each time the cell fired an action potential is marked with a colored dot. The majority of spikes occurred as the rat passed through the portion of the track connecting the upper left and upper right feeders, while the hippocampus was in the theta network state (red dots). Additional spikes occurred at each of the three feeder locations, when the hippocampus was in the LIA network state (blue dots).

neural substrate for understanding how the hippocampus supports cognitive function. An example of a hippocampal *place cell* is shown in figure 4.2. Most of the neuron's action potentials occur as the rat passes through a particular region of the environment (the cell's *place field*).

Place cells are found in both the CA1 and CA3 regions of hippocampus. While the majority of place cells exhibit a single place field in a given environment, some cells have multiple place fields (O'Keefe and Nadel, 1978; McNaughton et al., 1996; Redish, 1999). In most circumstances, place fields are distributed relatively uniformly across space (McNaughton et al., 1983, but see Hollup et al., 2001). The spatial firing patterns of place cells are so reliable that by using statistical techniques (e.g. Zhang et al., 1998), it is possible to decode an animal's position in the environment on a moment-by-moment basis from the population activity of a a sufficiently large ensemble of simultaneously-recorded place cells (fig. 4.3).

In an unchanging environment, cells maintain their place fields in the same location from day to day for effectively as long as it is possible to record their activity (Muller et al., 1987; Thompson and Best, 1990; Lever et al., 2002; Kentros et al., 2004; Ziv et al., 2013). While evidence suggests that any pyramidal neuron in the hippocampus is capable of having a place field in some environment, not all neurons are active in every environment (O'Keefe and Conway, 1978; Thompson and Best, 1989; Epsztein et al., 2011; Lee et al., 2012). Elegant work using immediate early gene expression to determine which cells were active as rats explored their surroundings has established that approximately 40% of hippocampal pyramidal cells have place fields in a typical experimental enclosure (Guzowski et al., 1999).

Pyramidal cells that lack a place field in an environment are usually completely silent as the rat moves through its environment and the hippocampus operates in the theta network state. However, all pyramidal neurons (even those that lack a place field in the current environment) may fire action potentials during SWR events that occur in the LIA network state (Buzsáki, 1989; Ylinen et al., 1995). Spiking during SWR events departs from the spatial tuning characteristic of place cells during the theta state (i.e. spikes can occur when the animal is outside of the neuron's place field). Examples of such *extra-field spikes* can be seen at the feeder locations in figure 4.2. The significance of SWR-associated place cell spiking will be addressed later in this chapter.

When an animal moves from one environment to the next, the place cell ensemble is said to "re-map," meaning that a different set of hippocampal neurons will exhibit place fields in the new location (Muller and Kubie, 1987; Quirk et al., 1990; Wilson and McNaughton, 1993). In sufficiently discriminable environments, the sets of neurons that express place fields are statistically independent (Guzowski et al., 1999; Redish, 1999). Further, if one cell expresses a place field in multiple environments, the fields that it exhibits are unrelated to one another (e.g. a neuron with a place field in the upper right hand corner of environment A is no more likely to have a place field at a similar position in the other environments for which it is active). Interestingly, there is no discernable anatomical topography in the hippocampus; neurons nearby one another in the brain are no more likely to be active in the same environment or to represent analogous portions of different environments (Redish et al., 2001).

As noted previously, a major input to the hippocampal formation arises from the

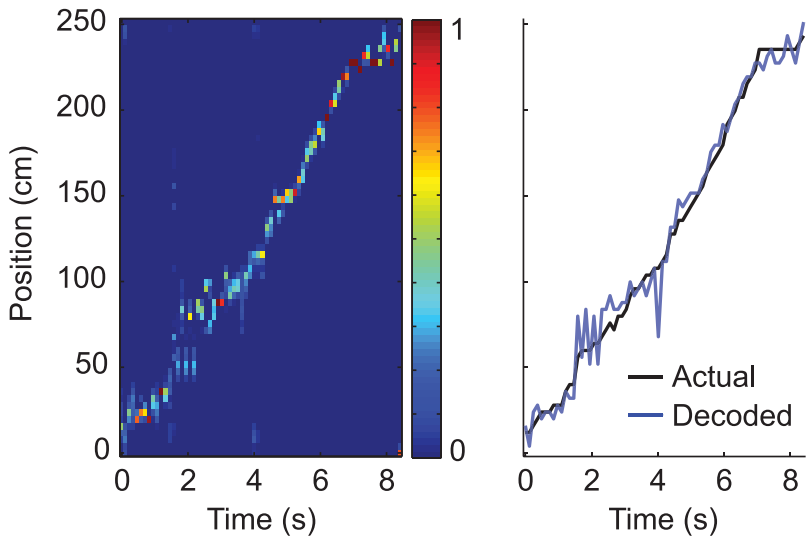

Figure 4.3: **Decoding position from place cell spiking.** A Bayesian decoding algorithm was used to estimate a rat's location as it ran a trajectory around the track while performing the foraging task. The left panel shows the probability distribution of the subject's most likely location over linearized space for each 100 ms time step (color intensity indicates the amount of probability in each spatial bin; columns on this plot sum to one). The peak of the decoded probability distribution for each time step was taken as the decoded position. Decoded position matched the animal's actual position well (right panel). Figure reproduced with permission (Wikenheiser and Redish, 2013).

entorhinal cortex. The entorhinal cortex is the recipient of highly-processed sensory information from all sensory modalities (Amaral and Lavenex, 2007). As such, an interesting and important question concerns how various types of sensory inputs influence place cell firing.

Distal visual cues (i.e. stable features of the surroundings approximately 1–3 meters from the rat in typical experimental settings) play a large role in establishing place fields. When distal visual cues are rotated around a stationary recording arena, place cells rotate along with the distal cues (O'Keefe and Conway, 1978; Muller and Kubie, 1987; Hetherington and Shapiro, 1997; Redish, 1999) as though they are constructed with respect to the configuration of prominent visual features of the environment. Similarly, rotating a symmetrical maze within an environment while leaving distal cues unchanged minimally affects the location of place fields (e.g. Miller and Best, 1980). However, place cells are not solely driven by visual inputs. Normal place fields are expressed in the complete absence of visual cues, and place fields initially established with respect to visual information are largely maintained in complete darkness (Hill and Best, 1981;

McNaughton et al., 1989; Quirk et al., 1990; Save et al., 1998).

In the absence of visual input, proprioceptive, self-motion signals drive place field expression, updating estimates of the rat's current location in the environment by *dead reckoning*. Over a long enough timescale, place cells maintained solely by self-motion information are prone to accumulation of position estimation error, and place fields consequently drift from their initial locations. This suggests that while the hippocampal system can maintain and update place cell activity using only self-motion information in the absence of other sensory inputs, stable visual cues (when available) serve to "reset" position estimates, maintaining a stable configuration of place fields across the environment (McNaughton et al., 1989; Touretzky and Redish, 1996; Samsonovich and McNaughton, 1997; Redish, 1999).

In addition to visual and proprioceptive information, other sensory modalities influence place cell firing. For instance, stable configurations of discrete olfactory cues can support place fields in the absence of other information (Zhang and Manahan-Vaughan, 2013), and auditory cues modulate place cell responses (Moita et al., 2003, 2004). Tactile features (detected either by whisker or paw) can also organize place field locations (Young et al., 1994; Shapiro et al., 1997; Tanila et al., 1997; Redish, 1999; Gener et al., 2013). Even the horizontal slope of an environment can play a role in orienting place fields (Jeffery et al., 2006).

The balance of evidence suggests that under the appropriate conditions, nearly any spatially inhomogeneous stimuli that subjects are equipped to detect can exert some measure of control over place field organization. It should be noted that the relative weighting of the different sensory modalities that influence place cell firing is not a static quantity, but can instead be modulated to best serve the animal's momentary needs. For instance, in an experiment where distal and proximal visual cues (i.e. extra- and intra-maze cues, respectively) were independently manipulated, the ability of cues to influence place fields depended on which set of cues subjects needed to use in order to correctly solve a navigation task; thus, when distal cues were important, distal cues controlled place fields, and vice versa (Zinyuk et al., 2000).

Based on the description above, it may be tempting to think of place cells as multimodal sensory integrators that combine information from useful environmental cues with proprioceptive signals in order to generate stable representations of space. However, even

this conception is incomplete, as hippocampal representations are known to depend on a number of non-sensory, cognitive variables (Redish, 1999). For instance, as discussed above, different environments are represented by independent sets of hippocampal place cells. However, subtle changes to a subset of the cues in an environment (e.g. removing or changing only one of many salient visual landmarks) often induce a fraction of neurons to lose or gain place fields without inducing a full re-mapping as would result from entering an entirely different environment (Muller and Kubie, 1987; Redish, 1999; Colgin et al., 2008). The extent of such "partial re-mapping" is not necessarily correlated with the extent of environmental change, and can vary between individual rats, suggesting that partial re-mapping does not result directly from a simple change in sensory experience. In addition, some environmental changes do not affect the location of place cells, but instead induce long-term changes in place cell firing rates (Leutgeb et al., 2005; Fyhn et al., 2007; Leutgeb et al., 2007; Colgin et al., 2008).

Changes in an animal's behavior can alter place cell representations within an identical sensory environment. Requiring rats to perform a novel behavioral task in a familiar environment can affect both the location of place fields (Markus et al., 1995; Jeffery et al., 2003; Kentros et al., 2004), and the moment-to-moment variability of place cell spiking (Fenton and Muller, 1998; Olypher et al., 2002; Jackson and Redish, 2007; Wikenheiser and Redish, 2011). Similarly, place cell representations can be modified as subjects learn the motivational significance of stable features of their surroundings (Hollup et al., 2001; Moita et al., 2003). These data suggest that multiple stable hippocampal maps are created for the same physical space (Touretzky and Redish, 1996; Redish, 1997; Samsonovich and McNaughton, 1997; Redish, 1999), and that an attentional process determines which map is instantiated on a moment-by-moment basis, allowing animals to dynamically select the representation best suited to current behavioral demands (Fenton et al., 2010).

4.2.4 Theta phase precession

Many early studies recognized that the spiking of hippocampal neurons was organized with respect to concurrent LFP rhythms. Action potentials fired by place cells during the theta network state are concentrated within cycles of the theta frequency LFP oscillation,

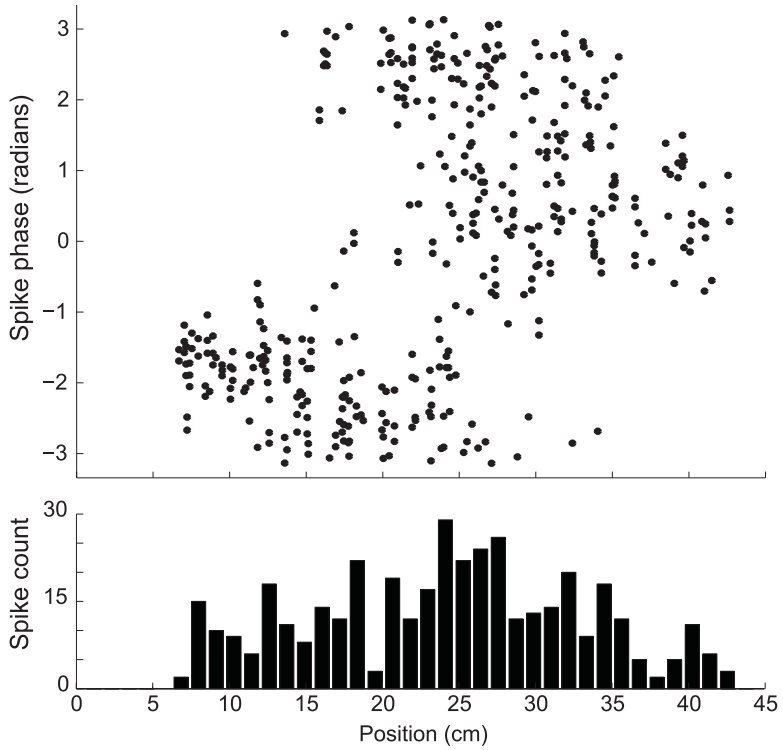

Figure 4.4: **Theta phase precession.** The phase precession of a single hippocampal place cell, recorded while a rat performed the foraging task, is depicted. The top panel plots the theta phase at which action potentials occurred against the animal's location in space at the time of each spike (data from all passes through the place field in a single behavioral session are plotted). The histogram in the bottom panel shows the number of spikes that occurred at each spatial position. The rat's direction of movement was from left to right. As the rat traversed the place field, action potentials showed a systematic relationship with the phase of the on-going theta rhythm, occurring increasingly earlier in the theta cycle as the rat progressed through the field.

with most spikes occurring near the trough of theta cycles.[1] The functional importance of this relationship, however, remained unclear.

Instead of examining the average theta phase at which place cells fired action potentials, O'Keefe and Recce (1993) examined the phase of individual place cell spikes as animals passed through place fields, and noted a consistent relationship between action potentials and theta phase. As animals entered a place field, spikes occurred late in the

[1] Theta oscillations can be detected throughout the hippocampal formation, but are greatest in amplitude near the hippocampal fissure. Concurrent recordings of theta across the depth of the hippocampus have revealed that theta is phase shifted at different depths. For consistency, when discussing theta phase in this thesis we refer to the oscillation as measured at the fissure.

theta cycle, but as animals progressed though the place field, spikes occurred at increasingly early theta phases. The authors termed this phenomenon *theta phase precession*, and suggested that it might serve as a temporal code for position within the place field. Such a temporal coding scheme could allow place cells to represent spatial locations with a greater precision, with the cell's overall firing rate indicating that the animal is somewhere within the place field, and the phase at which spikes occur indicating which portion of the field the animal currently occupies. Phase precession occurs in the vast majority of place cells (both in CA1 and CA3 regions), and is observed as rats traverse linear tracks and two-dimensional environments (O'Keefe and Recce, 1993; Skaggs et al., 1996; Dragoi and Buzsaki, 2006; Huxter et al., 2008).

Implications of phase precession: sequence learning and place field expansion

It is important to note that phase precession imposes a layer of the temporal organization on spiking which exceeds the precision necessary for the canonical spatial tuning of hippocampal neurons; place cells could in principle show identical spatial tuning without showing phase precession. Thus, understanding what purpose phase precession might serve is important to understanding hippocampal function.

As discussed above, phase precession increases the precision of spatial representations via a temporal coding strategy; statistical methods that consider both the firing rates of place cells and the phase at which spikes occur are better able to decode rats' positions in the environment (Jensen and Lisman, 2000). However, it remains unclear whether the brain is able to extract spike phase information and decode hippocampal activity in the same way. Further, this mechanism is compromised in two-dimensional environments, as spike phase always shifts from late to early during place field traversal, regardless of which direction the rat passes through the field (Skaggs et al., 1996; Maurer and McNaughton, 2007).

Nevertheless, phase precession could be a useful coding strategy for other reasons. For instance, it has been argued that phase precession ensures that spikes across neurons are appropriately aligned to induce spike timing-dependent plasticity (STDP; Bi and Poo, 1998). STDP is a Hebbian (1949) learning rule by which the strength of synaptic connections between a pair of neurons is increased or decreased depending on the relative timing of their action potentials. If a spike from neuron A occurs immediately before a

spike from neuron B (a "causal" ordering of action potentials), the synapse connecting A to B is strengthened. Conversely, an "acausal" ordering of spikes yields synaptic depression. The STDP mechanism operates on a timescale of < 80 ms; pairs of spikes separated by longer intervals do not alter synaptic connectivity. One line of reasoning argues that phase precession ensures the ordering of spikes on a fast timescale reflects the relative locations of place fields in space. In combination with STDP, phase precession could result in experience-dependent links developing between place cells, such that entry into one place field would eventually trigger an earlier than usual, predictive activation of the next place field in sequence. It should be noted that while phase precession organizes spikes in a manner conducive to STDP, phase precession itself does not depend on plasticity mechanisms (Ekstrom et al., 2001).

The experience-dependent backward expansion of place fields has been proposed as evidence that such a sequence-learning mechanism operates in the hippocampus. Throughout the course of a recording session, many place cells activate progressively earlier as the rat approaches their place fields (Mehta et al., 1997, 2000, 2002). This effect manifests as place fields stretching backward in a direction opposite that of the animal's movement with increased experience in an environment. Place field expansion is asymmetric; while the early portion of place fields progressively enlarge, the ends of place fields remain largely unchanged. Infusion of NMDA receptor antagonists into the hippocampus prevents place field expansion without affecting phase precession (Ekstrom et al., 2001), consistent with a Hebbian/STDP mechanism for place field expansion.

Implications of phase precession: theta sequences

As noted by Skaggs and colleagues (1996), the phase precession of individual place cells implies that the relative timing of place cell spikes within a theta cycle should recapitulate the ordering of place fields in space, giving rise to orderly sequences of place cell spiking on a theta cycle timescale. During the initial portion of theta cycles, neurons with place fields predominately behind the rat are active, while at later phases, most spiking results from the predictive activation of fields immediately in front of the animal (fig. 4.5). In consequence, over the course of individual theta cycles, the ensemble representation traces out a temporally-compressed trajectory through the region of space around the rat's current position. These representations are called *theta sequences*

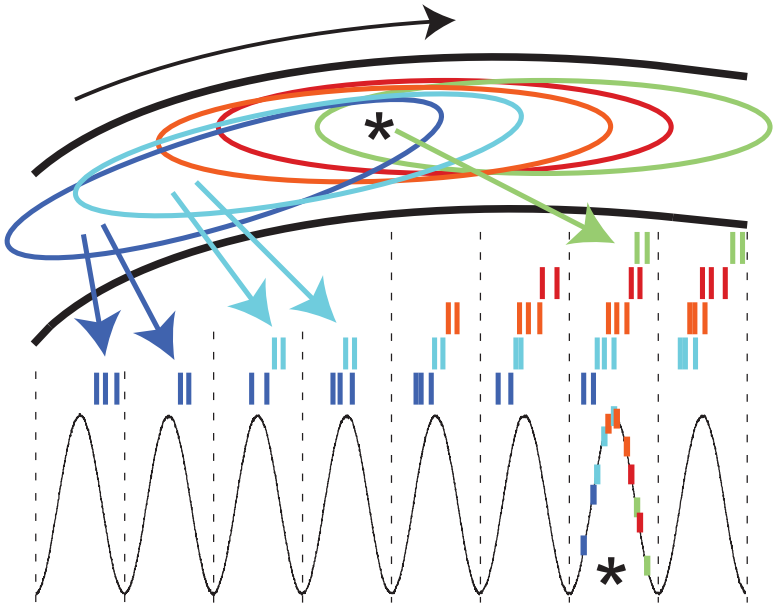

Figure 4.5: **Theta sequences.** The phase precession of individual neurons implies a sequential pattern of place cell activation at the ensemble level. The cartoon depicts a raster plot of place cell spiking, with time along the x-axis and each row displaying the activity of a single place cell. Vertical tick marks indicate times that action potentials occurred, and are colored to match their corresponding place fields depicted in the segment of track above the raster plot. The asterisks indicates the rat's position. Tracing along each row of the raster, phase precession results in a systematic change in the phase at which place cells spike relative to the theta oscillation as the rat progresses through place fields. Vertical dashed lines denote the boundaries of theta cycles; within theta cycles, ensemble spiking occurs in an order that reflects the organization of place fields in space. Figure courtesy of Nathan Schultheiss.

(Foster and Wilson, 2007).

One prominent theory (Jensen and Lisman, 1996; Lisman and Redish, 2009; Lisman and Jensen, 2013) argues that sequence learning and sequence read-out are the primary functional principles around which hippocampal representations are constructed during the theta network state. This view suggests a revision to the traditional conception of spatial coding in the hippocampus; instead of assuming that place fields represent a region of space, a cell's "true" representation is taken to be much smaller, corresponding to the end of a traditional place field. In this view, the spikes that form the first portion of a traditional place field are actually predictive representations of the forthcoming location that the cell represents. Several experimental findings are consistent with this idea. As described previously, place field expansion primarily affects the early portion of

place fields, leaving the ends of fields (the "true" location represented by the cell) intact (Mehta et al., 1997, 2000, 2002). In the sequence coding framework, place field expansion results in better (i.e. earlier) prediction of upcoming locations. Spike-time correlations between pairs of place cells within theta cycles are more reliable than the correlation between spike phase and position (Dragoi and Buzsaki, 2006). Similarly, theta sequences are more precisely patterned than would be expected if phase precession alone organized place cell spike timing (Foster and Wilson, 2007, but see Chadwick et al., 2014).

The sequence-based line of reasoning suggests that any temporal coding for position arising from phase precession is incidental to theta cycle-paced sequence expression accompanied by movement through the environment; together, these two factors are sufficient to produce phase precession of individual place cells. Understanding the mechanism of phase precession could speak to the question of whether ensemble-level sequence expression or single cell-level spatial representations are the best description of place cell firing patterns. Proposed single cell mechanisms of phase precession depend on a interaction between cells' intrinsic membrane properties and the temporal patterning of inhibitory and excitatory inputs neurons receive. Hypothetical network mechanisms postulate asymmetric synaptic connectivity architectures (pre-wired, established through experience-dependent synaptic learning rules, or some combination of the two) in the hippocampal network (Maurer and McNaughton, 2007). Despite extensive experimentation, modeling, and debate, the mechanisms of phase precession in the hippocampus remain unresolved. Regardless of precisely how phase precession and theta sequences arise, however, it is increasingly clear that theta sequence representations are important for behavior on decision making tasks involving both both spatial and non-spatial elements. In the next section, we review evidence linking theta sequences to adaptive behavioral performance; Chapter 5 presents experimental data that further support this view.

Theta sequences and decision making

In one of the first studies to examine the content of hippocampal ensemble representations on a sub-theta cycle timescale, Johnson and Redish (2007) analyzed the activity of CA3 hippocampal neurons recorded as rats performed a multiple-T decision making task using a fast timescale Bayesian decoding algorithm (Zhang et al., 1998; Brown

et al., 1998). While rats paused to ponder a high-cost decision (go left or right at the choice point), hippocampal activity ceased to represent the animal's actual location on the maze, and instead projected forward along the maze ahead of the rat, tracing out trajectories along the possible future paths the rat was choosing between. The LFP recorded during these non-local representations exhibited strong theta oscillation, suggesting that theta sequences represented paths ahead of the animal. The expression of forward-shifted trajectories could be useful for decision making, generating a representation of possible future actions that could be evaluated by structures downstream of the hippocampus, such as the ventral striatum (van der Meer and Redish, 2009, 2010), to arbitrate between choices.

The discovery of non-local hippocampal representations on a theta cycle timescale was surprising; when averaged over many theta cycles, theta sequence representations appear to begin slightly behind the rat, and project forward only a small distance beyond the rat's current position (Skaggs et al., 1996; Foster and Wilson, 2007; Maurer et al., 2012). However, subsequent work has revealed that the expression of individual theta sequences is more variable than initially appreciated. For instance, Maurer and colleagues (2012) demonstrated that the "look-ahead" representation found at the end of the theta cycle is modulated by behavior in a manner consistent with predictive function. By carefully examining average ensemble representations across the theta cycle, they found that the extent of space represented within a cycle scaled with running speed, resulting in representations that extended farther forward as animals moved more quickly and were arguably in greater need of predictive representations extending farther along their immediate future path.

Gupta and colleagues (2012) examined individual theta sequences in detail, on a cycle-by-cycle basis, as rats performed a spatial decision making task, and observed considerable heterogeneity in the trajectories traced out by theta sequence spiking. Some sequences began behind the rat and ended at its current location, other representations were centered around the rat, and still others began near the rat and projected forward to varying extents. The expression of theta sequences was modulated in a way that suggests theta sequences parsed or segmented the environment; as rats approached regions of the maze imbued with motivational or informational salience (i.e. turns, food delivery sites, etc.) theta sequences shifted from starting near the rat and projecting forward to starting

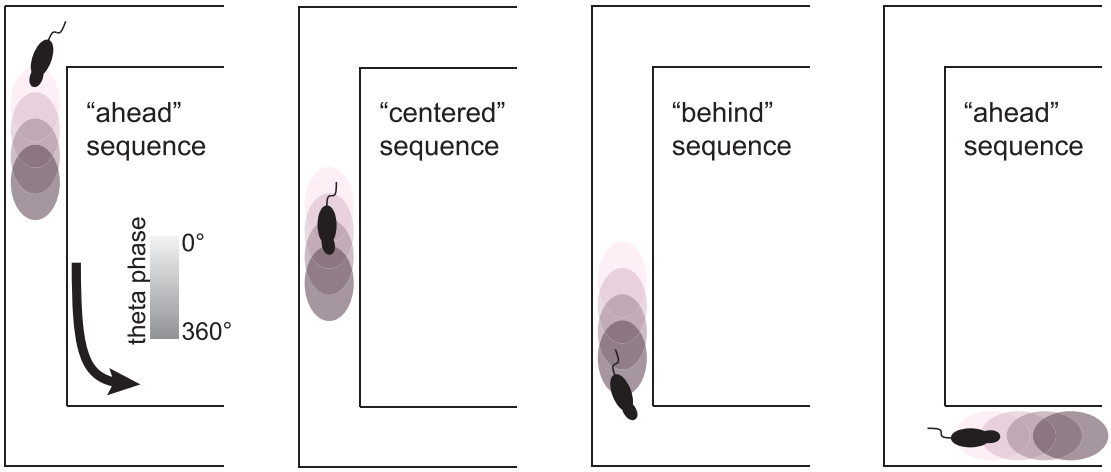

Figure 4.6: **Theta sequences form "chunked" spatial representations.** The content of theta sequences is modulated by salient features of the environment in a manner that gives rise to a distinctive, segmented representation of space. Immediately after rounding the first corner on a maze, sequences begin at the rat's current location and extend asymmetrically, such that a greater portion of space in front of the rat is represented. Midway between turns, sequences tend to be approximately centered on the rat. As the rat nears the second turn, sequences shift backward, beginning some distance behind the rat and extending up to its current position. Finally, once past the second turn, representations are again largely ahead of the animal. Modulation of theta sequence content in this way leads to an increased density of representation covering the regions between landmarks. Cartoon modeled after data from Gupta et al., 2012.

behind the rat and projecting up to its actual position. Thus, as a rat approached a prominent intra-maze landmark, theta sequences representations shifted from predictive and forward-directed to more retrospective or backward-looking.

This shift in sequence content around landmarks imposed a distinctive organization on the hippocampal representation of space, in which semi-discrete, landmark-bounded "chunks" of the environment emerged (fig. 4.6; Gupta et al., 2012). A recent study of rats performing a linear track task reported a similar result, with CA1 place cell activity appearing more forward-directed as subjects left feeder sites at the ends of the track, and more backward-lagging on approach to feeders (Bieri et al., 2014). These results suggest that, rather than passively encoding features of the environment as they exist, representations within theta cycles actively segment space, effecting a sort of spatial information compression..

Although spatial representations are often the best, most obvious correlate of hippocampal neuron activity, representations of non-spatial information by hippocampal

ensembles have been observed on tasks that hinge on non-spatial factors. Non-spatial hippocampal representations share many of the properties of place cell activity. For instance, phase precession of hippocampal neurons has frequently been observed during non-spatial behaviors. Just as spatial phase precession results from the read-out of spatial sequences during forward movement, phase precession during non-spatial behaviors might be thought of as non-spatial theta sequences expressed during imagined, mental progression through the representational dimension of the sequence (Lisman and Redish, 2009). Representations like these are consistent with processes such as mental time travel (Suddendorf and Corballis, 2007, 2010; Wikenheiser and Redish, 2012) or imagination (Buckner, 2010), and could be useful for mentally exploring possible outcomes associated with future actions.

Early work established that phase precession can occur even when rats are not traveling through space by recording hippocampal activity as rats ran in a stationary running wheel (Harris et al., 2002). Pastalkova and colleagues (2008) extended this work, providing evidence that theta sequence-like representations include non-spatial information. In their experiments, rats performed a hippocampus-dependent delayed spatial alternation task. During the delay periods punctuating spatial alternation trials on a T-maze, rats were forced to spend a fixed duration running in a wheel, a behavior that is accompanied by strong theta oscillation in the hippocampus. During alternation trials, hippocampal neurons showed normal spatial tuning on the T-maze. During wheel-running epochs, however, hippocampal neurons (in some cases the same cells that exhibited place fields on the maze) showed reliable tuning to the time elapsed in the running wheel, consistent with previous theoretical work (Levy et al., 2005). Different subsets of time-encoding neurons were activated during delays that preceded leftward or rightward alternation trials, and much like place cells, these temporally-tuned cells showed clear phase precession (Pastalkova et al., 2008). Hippocampal neurons with temporal tuning have since been found in rats performing other behavioral tasks that require planning (Gill et al., 2011; MacDonald et al., 2011; Kraus et al., 2013; Eichenbaum, 2013). Together, these findings are strongly suggestive of coordinated, sequential representations within theta cycles, in this case encoding a temporal, rather than spatial, sequence.

Takahashi and colleagues (2009) recorded hippocampal cells as rats performed an operant alternation task. Subjects initiated a trial by maintaining a nosepoke in a

central port until the fixation period ended, and then alternated nosepokes in ports to the right or left. Reminiscent of Pastalkova and colleagues' (2008) findings, unique ensembles of neurons were active during the fixation period prior to leftward or rightward alternation trials. Spiking during the fixation period phase precessed, consistent with theta sequences (Takahashi et al., 2014). Note that the neurons active during the fixation delay were not necessarily place cells that would subsequently activate as subjects moved to the left or right nosepoke ports; instead, it seems that neurons active during fixation were forming a purely temporal representation related to forthcoming behavior.

Lenck-Santini and colleagues (2008) measured hippocampal activity as rats were engaged in a shock avoidance task. Subjects were dropped on to the metal floor of the test arena and required to jump out of the arena within a fixed time window to avoid an electric shock delivered through the arena floor. Hippocampal pyramidal cells spiked before self-initiated leaps to safety, and these spikes precessed relative to ongoing theta oscillation, suggesting that cells encoded the temporal intervals around an important, task-related event. Although executing the jump out of the arena was, of course, a movement through space, jump-responsive neurons began phase precession seconds before jumps were actually executed.

The experiments described thus far show that spatial and non-spatial information relevant to behavior is represented by hippocampal neurons, that cells encoding non-spatial information phase precess, and that forthcoming behavior can influence the content of theta sequences. While these results are all suggestive of a role for theta sequences in planning future behaviors, establishing a causal link between theta sequences and behavior has proven challenging. An ideal manipulation to get at this question would disrupt the sequential arrangement of spiking without altering other properties of hippocampal neuron representations.

Cannabinoid agonists exert a surprisingly specific effect on the timing of place cell spikes. Robbe and colleagues (2006; 2009) administered a cannabinoid agonist to rats trained to perform a delayed spatial alternation task. The drug had a clear behavioral effect, reversibly reducing task performance to chance levels. Surprisingly, the spatial firing patterns of CA1 place cells were almost entirely unaffected by the drug. Temporal coordination between cells, however, was effectively absent. Theta phase precession and other characteristics of theta sequences were severely attenuated by drug administration.

Thus, a manipulation that disrupted theta sequences (while mostly sparing other place cell properties) had drastic effects on behavior in a hippocampus-dependent task.

Because cannabinoid receptor activation impaired performance of a learned task (Robbe and Buzsaki, 2009), these data support the idea that theta sequences play a role in on-line planning for immediately forthcoming behavior. Close inspection of position tracking data before and after drug administration (Robbe and Buzsaki, 2009; their figure 1) reveals that cannabinoid agonism altered fine-scale behavioral dynamics in a suggestive manner. Following drug delivery, the rat spent much more time pausing on the central stem of the maze and at the choice point. Additionally, the rat appears to have spent more time peering over the edge of the maze, and generally scanning his surroundings. This is suggestive of an increase in vicarious trial and error (VTE), a behavior Tolman (1938; 1939) and others (Johnson and Redish, 2007; Johnson et al., 2007; Papale et al., 2012) have associated with deliberative decision making. Usually, VTE is a fairly transient event (Johnson and Redish, 2007; Papale et al., 2012); a fascinating possibility is that with dysfunctional sequence expression rats were impaired at planning a suitable course of action, and therefore remained deliberative and indecisive for longer than usual, resulting in sustained VTE. However, because cannabinoid agonism also caused some amount of motor side effects (not to mention its well-documented cognitive effects in human subjects), this hypothesis, while intriguing, remains speculative.

4.2.5 Hippocampal sequences during the LIA network state

As mentioned previously, hippocampal firing patterns during the LIA network state are quite different than those of the theta state. In contrast to the place fields and theta sequences that characterize the theta state, pyramidal neurons are largely silent during LIA, except during sharp-wave ripple (SWR) events, brief population bursts of spiking accompanied by distinctive, high frequency oscillations in the LFP. Although the LIA state occurs both during moments of awake quiescence and slow wave sleep, early work on the topic focused on LIA during sleep.

Pavlides and Winson (1989) showed that hippocampal neurons active as rats explored an environment were more likely to fire action potentials during post-behavior sleep. Later work extended this result, demonstrating that correlations in the spiking of pairs of place cells co-active during behavior were preserved and enhanced during sleep periods

following behavior (Wilson and McNaughton, 1994). These data suggest that spiking during sleep-associated LIA states reflects recent behavioral experience. It has since been established that place cell spiking around SWRs during sleep represents spatial trajectories through previously-visited environments (Skaggs and McNaughton, 1996; Kudrimoti et al., 1999; Nádasdy et al., 1999; Lee and Wilson, 2002). Additionally, similar sequential trajectory representations accompany SWRs during awake LIA (Foster and Wilson, 2006; O'Neill et al., 2006; Jackson et al., 2006; Diba and Buzsáki, 2007).

Sequences during LIA differ in several ways from theta sequences (fig. 4.7). Foremost, the content of LIA sequences can be completely divorced from the animal's current location in space. While theta sequences represent trajectories through regions of space near the animal's actual location, LIA sequence trajectory representations can be local (beginning or ending near the animal's position) or remote (traversing a region of space that does not cross the animal's current location). Compared to theta sequences, LIA sequences tend to represent longer trajectories, which can in some cases span the entire length of the environment, recruiting most of the place cells active in that context. Finally, the temporal compression of LIA sequences is greater than that of theta sequences; where as theta sequences represent spatial trajectories approximately 10 times faster than the animal travels through the same region of space (Skaggs et al., 1996), LIA sequences can occur at speeds more than 20 times that of actual behavior (Nádasdy et al., 1999; Davidson et al., 2009).

LIA sequences and memory consolidation

A popular idea, dating back at least to theories proposed by Marr (1971), is that memories encoded by the hippocampus are later re-instantiated in neocortical brain regions for long-term storage. This process is known as *memory consolidation* (Squire and Zola-Morgan, 1991; Sutherland and McNaughton, 2000). Others have suggested that consolidation involves an intra-hippocampal transfer of information, between CA3 and CA1 regions (Buzsáki, 1989). Although determining the brain locus of a memory's final resting place remains an area of active research (Cohen and Eichenbaum, 1993; Dudai and Eisenberg, 2004), models of consolidation share a requirement for some mechanism of inducing long-term changes in synapse function to store previous experience.

Hippocampal sequences in the LIA state have many features expected of a memory

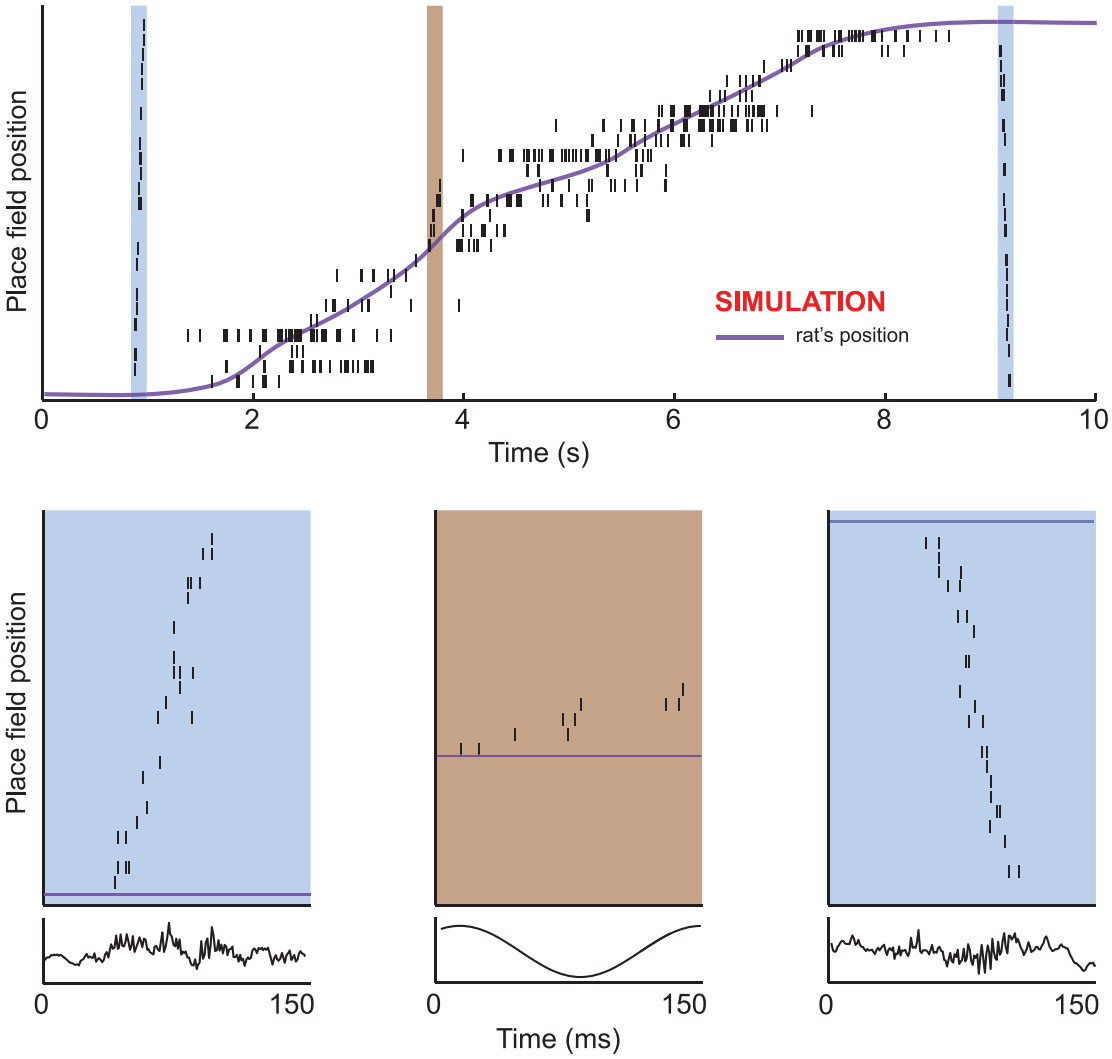

Figure 4.7: **LIA sequences.** Simulated activity in an ensemble of place cells is plotted as a rat runs through an environment. Each row represents the activity of a single neuron, and cells are sorted based on where their fields occur in space. Sequences are highlighted and reproduced (with the accompanying LFP activity) on an expanded timescale in the lower panels of the figure. As the rat runs, spiking within cycles of the theta rhythm is organized into sequences (red shaded region). When the rat stops running, the hippocampus enters the LIA network state, and LIA sequences occur (blue shaded regions). LIA sequences show both backward (right lower panel) and forward (left lower panel) temporal ordering. Unlike theta sequences, LIA representations can span large swaths of the environment, and exhibit greater temporal compression.

consolidation mechanism. Sequences occur during sleep (Wilson and McNaughton, 1994; Skaggs and McNaughton, 1996), which has long been recognized as a privileged time for learning and memory improvement, as reduced incoming sensory information allows internal brain dynamics to dominate information processing. The temporal compression inherent to LIA sequences ensures that spiking occurs at a fast timescale appropriate for the induction of synaptic plasticity. Further, patterns of spiking that occurred during behavior can be repeated many times during sleep-associated LIA, further promoting activation of plasticity mechanisms (Sutherland and McNaughton, 2000; O'Neill et al., 2010; Carr et al., 2011; Girardeau and Zugaro, 2011). These features of LIA sequences have led to the idea that trajectory representations during SWRs "reactivate" or "replay" behavioral experience in the service of memory consolidation.

Recent experiments have gone beyond correlational approaches to establish a causal role for sleep LIA sequences in memory consolidation. By monitoring the LFP while animals slept after performing a behavioral task, electrical stimulation of the hippocampus could be triggered on SWR events, disturbing the temporally-patterned sequence spiking that would otherwise occur. Disrupting sequence spiking in this way impaired learning, strongly supporting a causal role for SWRs in consolidation (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010). While the electrical stimulation used to perturb hippocampal representations in these experiments is somewhat non-specific in that it may have affected aspects of hippocampal function beyond spiking sequences, these data nevertheless provide direct, causal support for the importance of sleep LIA sequences in memory consolidation.

4.2.6 Awake LIA sequences and the construction of representations

While it seems likely that sleep LIA sequences function at least partially in memory consolidation, sequences also occur during waking quiescence (Foster and Wilson, 2006; Jackson et al., 2006; Diba and Buzsáki, 2007; Wikenheiser and Redish, 2013), and the properties of awake LIA sequences are suggestive of a broader functional role. For instance, awake sequence representations and actual behavior can diverge substantially, a fact that is hard to account for if sequences function solely to consolidate memories of recent experience. As noted previously, sequence representations do not necessarily begin near the animal, nor do they necessarily cross through the position the rat currently

occupies (Karlsson and Frank, 2009; Davidson et al., 2009; Gupta et al., 2010). In fact, when animals are trained in multiple, distinct environments, sequences representing previously-experienced contexts (which recruit place cells that are not active in the current location) have been described, intermingled with "local" sequences representing paths through the animal's current surroundings (Jackson et al., 2006; Karlsson and Frank, 2009).

Even when considered over a fairly long timescale (i.e. the duration of an entire behavioral session), the frequency with which portions of the environment are included in awake LIA sequences often does not match cumulative behavioral experience. For instance, in rats performing a multiple-T decision making task, the probability of a location being included within awake LIA sequences was sometimes inversely related to how often the rat visited that area. In a session where only left-side laps were rewarded (and rats consequently made few visits to the right loop of the maze), sequence representations to the unrewarded side were actually more frequent than those representing the path the rat traveled for the majority of the session (Gupta et al., 2010).

Further challenging the memory consolidation perspective, sequences during awake LIA can occur in the opposite order of experience (backward sequences; Foster and Wilson, 2006; Diba and Buzsáki, 2007; Davidson et al., 2009; Gupta et al., 2010). While the bulk of trajectory representations during sleep match the direction that rats traveled through space while awake (e.g. Lee and Wilson, 2002, but see Wikenheiser and Redish, 2013), the occurrence of backwards trajectories during waking LIA seems problematic for consolidation, as it would lead to the storage of memory episodes in the wrong temporal order. Consolidation of both backward and forward sequences would be vulnerable to memory interference, as forward and reverse trajectories represent equally-plausible experiences, but the rat may have traveled in only one direction.

Finally, cognitive factors seem to have a strong influence over the content of awake LIA sequences, suggesting that awake sequence representations are more than a passive reflection of past experience, but instead are biased by motivationally relevant information. For instance, when animals encounter new environments for the first time, awake LIA sequences preferentially represent recently-explored portions of space (Cheng and Frank, 2008). Similarly, reinforcement sculpts awake LIA sequence content. Sequences are more likely to occur during quiescence following reward delivery (Singer and Frank,

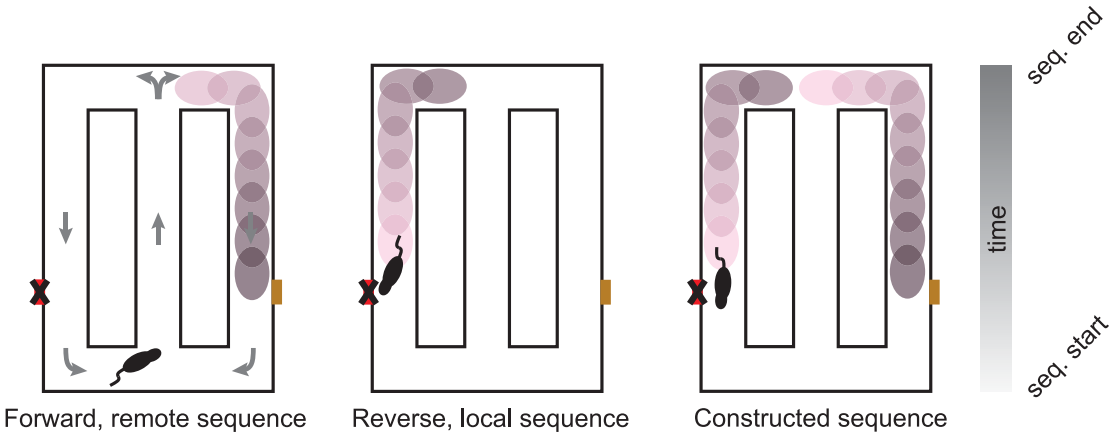

Figure 4.8: **Construction of novel representations by LIA sequences.** Combinatorial expression of LIA sequences can generate trajectory representations never directly experienced by the subject. In this example (modeled after the results in Gupta et al., 2010) a rat is performing a T-maze decision making task. Food delivery sites are marked with rectangles, and only visits to the right-side feeder are rewarded. Arrows in indicate the possible directions the rat is allowed to travel at each location on the maze. A forward sequence spanning the region between the choice point and the right feeder (left panel), preceded by a backward sequence originating at the left feeder and ending at the choice point (middle panel) could be used to represent the unexperienced trajectory from left-side feeder to right-side feeder, a shortcut between potential reinforcers (right panel). Gupta and colleagues (2010) observed constructive representations like this more frequently than would be expected due to the chance occurrence of each component sequence.

2009).

Together, the work reviewed here suggests that awake LIA sequences do more than simply recapitulate recent experience, but are instead involved in synthesizing representations of the world by assembling bits of previous spatial experience and motivational information. In this way, the hippocampus could connect disparate pieces of experience together in a novel fashion to adaptively guide behavior. Consistent with this idea, Gupta and colleagues (2010) observed that novel (i.e. never-experienced) trajectories can be represented by awake LIA sequences. During performance of a spatial decision making task, the authors discovered sequences connecting spatially-contiguous portions of the maze to trace out a path the rat had never actually traversed. Forward and backward ordered sequences occurred in equal proportion during performance of this task, and *de novo* trajectories were constructed by linking backward and forward representations of neighboring maze segments (fig. 4.8). Representations like these could subserve short-cut behavior or other cognitive processes, and suggest a mechanism by

which animals could extrapolate beyond actual, physical experience (Samsonovich and Ascoli, 2005; Derdikman and Moser, 2010). Neuroimaging data indicates that the human hippocampus is similarly involved in constructing novel representations from memory (Barron et al., 2013; Schacter and Addis, 2009; Buckner, 2010; Addis et al., 2011).

Direct links between awake LIA sequences and forthcoming behavior have been shown in several experiments. In rats trained to shuttle back and forth between food delivery sites placed at both ends of a linear track, awake LIA sequences expressed before and after completion of trials contained representations consistent both with planning and memory-like functions (Diba and Buzsáki, 2007). While paused at a feeder site, before initiating a new trial, forward-ordered sequences beginning at the rat's current location and extending to the opposite feeder were detected. After arriving at a feeder, sequences re-tracing the recently-completed journey in reverse occurred (Diba and Buzsáki, 2007). Representations preceding trial start might have played a role in planning upcoming trajectories, while reverse-ordered sequences occurring after the completion of a trial may have had some memory role, perhaps in associating past behavior with reward in a process reminiscent of the credit assignment problem of reinforcement learning (Johnson and Redish, 2005; Foster and Wilson, 2006; Foster and Knierim, 2012). More recent work by Pfeiffer and Foster (2013) showed that awake LIA sequences recorded while rats performed a goal-directed navigation task were biased to end in the spatial location that the rat would next travel to.

The results of a study by Jadhav and colleagues (2013) suggest a causal role for awake LIA sequences in online planning. The authors probed the function of awake SWRs in rats learning to perform a hippocampus-dependent decision making task (Kim and Frank, 2009). In these experiments (Jadhav et al., 2013), electrical stimulation disrupted spiking during SWRs, following the same paradigm used to probe the function of SWR sequences during sleep (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010). In the behavioral task, rats were required to visit the arms of a W-shaped maze in a particular order. Inbound trials required rats to run from an outer arm to the central stem. In contrast, on outbound laps, rats departed the central stem and headed to the outer arm opposite that they had visited most recently. Although hippocampal lesions slow the learning of both inbound and outbound trials, real-time disruption of awake SWRs specifically impeded acquisition of outbound trials. Further, in animals pre-trained to

proficiency on the task, SWR disruption resulted in mildly degraded performance on outbound choices (Jadhav et al., 2013). Although subject to the same caveats discussed above in reference to sleep SWR disruption (Girardeau et al., 2009; Ego-Stengel and Wilson, 2010), these data provide strong evidence that awake SWRs play some role in coordinating behavior in real time, in addition to whatever learning function they might fulfill.

The recent report of *pre-play*, LIA sequences that represent trajectories through regions of an environment that subjects can view but not physically enter, suggests that the hippocampus can build representations of regions of space the animal has not yet encountered (Dragoi and Tonegawa, 2011, 2013). This finding is in line with observations that the hippocampus plays a role in processing visible, but inaccessible, objects (Levcik et al., 2013). While observation of a space is not sufficient for the establishment of stable place cells covering the unvisited portion of the environment (Rowland et al., 2011), it is possible that pre-play establishes a rough, approximate spatial representation that is subsequently refined and bound to prominent environmental features after direct, physical experience. Although the behavioral implications of pre-play are not yet clear, this sort of representation has clear utility in planning future behavior.

4.2.7 Hippocampal sequences and the cognitive map

O'Keefe and Nadel (1978), based on lesion, behavioral, and (limited) electrophysiological evidence presciently suggested that the hippocampus functions as a cognitive map. Tolman developed the concept of a cognitive map based on behavioral studies of rats, envisioning a proactive, predictive learning system that could flexibly manipulate and retrieve information to guide surprisingly complex behaviors (Tolman, 1932, 1938, 1939; Tolman et al., 1946). Tolman's conception of the cognitive map hinged on an internal representation of the environment, constructed by animals in the absence of explicit reward or punishment, and used to generate expectancies, or predictions about the cause and effect structure of the world (Tolman, 1948; Johnson and Crowe, 2009).

The cognitive map theory ascribed unprecedented mental abilities to rats; needless to say, these ideas were not without controversy (Hull, 1943; Guthrie, 1952; Tulving and Madigan, 1970). Nevertheless, subsequent work has validated many of Tolman's conjectures, and his cognitive map framework is increasingly accepted. The properties of

hippocampal sequence representations during LIA and theta network states fulfill both the "cognitive" and "map" components of Tolman's construct, flexibly manipulating complex representations of the environment that are rooted in space, but modulated by relevant non-spatial information.

Sequence representations like the ones described in this chapter are likely neural mechanisms by which the hippocampus supports synthetic, prospective processes like mental time travel (Suddendorf and Corballis, 2007, 2010; Wikenheiser and Redish, 2012), prediction (Lisman and Redish, 2009; Addis and Schacter, 2012), and imagination (Samsonovich and Ascoli, 2005; Hassabis et al., 2007b,a; Buckner, 2010). Strong tests of this idea require causal manipulations, and some effective attempts in this direction have been described (Robbe et al., 2006; Robbe and Buzsaki, 2009; Girardeau et al., 2009; Ego-Stengel and Wilson, 2010; Jadhav et al., 2013).

Nevertheless, much remains to be learned by correlating hippocampal activity with behavior. A notable gap in our current understanding of rodent hippocampal function concerns goal-directed behavior. The majority of previous hippocampal electrophysiology experiments have tested animals on memory-guided decision making tasks, in which rats are rewarded for making correct choices under an arbitrary reward contingency defined by the experimenter. Although informative, these experiments do not speak to how the hippocampus functions when animals' choices are driven by their own, internal valuations.

The intertemporal foraging task elicits value-guided decisions from rats. When choosing between feeder sites, there are no explicitly correct or incorrect responses; instead, subjects are free to harvest reward as they see fit, with few constraints on behavior imposed externally. Examining hippocampal activity while subjects perform the foraging task allows us to ask whether and how rats' self-determined intentions are reflected in hippocampal representations. This topic is taken up by experiments described in the next chapter.

Chapter 5

Hippocampal representations on the intertemporal foraging task

In this chapter, we examine hippocampal representations recorded as rats performed the intertemporal foraging task. Portions of the data presented in this chapter are currently under review for publication.

As discussed in the previous chapter, the hippocampus is a key component of the brain's decision making system (Cohen and Eichenbaum, 1993), and is critical for the episodic simulation of imagined future possibilities and flexibly using past experiences to make predictions about the future (Lisman and Redish, 2009; Schacter and Addis, 2009; Buckner, 2010; Addis et al., 2011; Barron et al., 2013). In humans, it is well established that the hippocampus plays a key role in goal-directed decision making (Peters and Büchel, 2010; Viard et al., 2011; Wimmer and Shohamy; Bornstein and Daw, 2013; Shohamy and Turk-Browne, 2013). Less is known about how the rodent hippocampus supports behavior in such situations.

Assuming the hippocampus mediates similar cognitive functions across species in the performance of goal-directed behavior, planning-related signals in the rodent hippocampus would likely occur within the framework of spatial representations (O'Keefe and Nadel, 1978). Theta sequences are a promising candidate neural mechanism for prospective planning (Skaggs et al., 1996; Tsodyks et al., 1996; Dragoi and Buzsaki, 2006; Foster and Wilson, 2007; Maurer et al., 2012). As implied by hippocampal phase

precession phenomenon (O'Keefe and Recce, 1993; Maurer and McNaughton, 2007), the average theta sequence representation begins slightly behind the rat's current position in space, and projects forward some distance beyond the rat (Skaggs et al., 1996; Foster and Wilson, 2007; Maurer et al., 2012). However, given the moment-to-moment variability in theta sequence expression (Maurer et al., 2012; Gupta et al., 2012), modulation of theta sequence content suggests a means by which the hippocampus could use a spatial framework to support goal-directed decision making (Johnson et al., 2007; Lisman and Redish, 2009; Penny et al., 2013).

If theta sequences are computations of prospective plans, ensemble spiking should in some way reflect rats' currently-active goals. To test this idea, we examined theta sequences as rats performed a value-guided decision making task. We measured the theta sequence look-ahead distance (the extent to which theta sequence spatial representations extended forward in front of the rat's actual position) and tested whether this property of theta sequence representations was related to rats' goals. Because behavior was guided by subjective preferences rather than rats' attempts to determine and match reward contingencies set by the experimenter, the foraging task offered a unique opportunity for testing how the hippocampus contributes to volitional navigational decisions akin to those made in natural settings.

5.1 Neural recording and analysis methods

5.1.1 Behavioral training

Four male, Fisher-Brown Norway hybrid rats aged 6–14 months, (Harlan; Indianapolis, IN) were subjects for the experiment. Each of these rats first completed the behavioral portion of the experiment described in Chapters 2 and 3. Before beginning behavioral training rats were handled for 7 days and acclimated to eating the food pellets that would be delivered during the behavioral task (45 mg sucrose pellets, Test Diet; St. Louis, MO). Rats were maintained on a 12 hour light-dark cycle, and behavioral sessions occurred at the same time daily, during the light phase. Subjects were food restricted to maintain their weight at $\geq 80\%$ of their free-feeding weight; water was always available in the home cage. All experimental and animal care procedures complied with National Institutes of Health guidelines for animal care and were approved by the Institutional

Animal Care and Use Committee at the University of Minnesota.

Rats performed the foraging task on an elevated, circular track (diameter = 80 cm) with three food pellet dispensers (Med Associates; St. Albans, VT) positioned evenly around the perimeter. An overhead camera recorded subjects' position via a light-emitting diode affixed around the rat's body (before electrode implantation) or light-emitting diodes mounted to the headstage (post electrode implantation). Data were recorded with a Cheetah 160 acquisition system (Neuralynx; Bozeman, MT). Custom Matlab software (MathWorks; Natick, MA) controlled the task.

The behavioral sequence was described previously (chapter 2). Briefly, subjects first performed a training task in which they ran unidirectional laps around the track to earn food at each feeder. Pellets were delivered as soon as rats arrived at each feeder site, and attempts to run backwards were blocked by the experimenter during training sessions (during neural recording sessions, rats were well trained on the task and only ran forwards). After rats ran 30 or more laps for three consecutive sessions, the training phase was considered complete and task performance began. During daily, 30-min sessions of the foraging task, rats earned food pellets from the three feeder locations, each associated with a delay period. The delay began when the subject approached within 7.5 cm of a feeder site. Entry into this zone was indicated by a tone sequence (200 ms pulses, repeated once per second). The tone's frequency was proportional to the site's delay. Six sets of delays were used, defining six unique session types. Rats experienced session types in pseudorandom order (the same type was never repeated on consecutive days). Delays were counterbalanced across feeder sites to ensure that delay distributions at each location were equivalent across sessions. Subjects performed each session type four times.

5.1.2 Surgery and recordings

Following completion of the behavioral sequence (24 sessions) subjects were allowed *ad libitum* access to food for at least 24 hours, and then implanted with tetrode arrays targeting the dorsal CA1 region of hippocampus of the right hemisphere (-3.8 mm anteroposterior, 2.0 mm lateral from bregma). Surgical procedures have been described in detail elsewhere (Jackson et al., 2006; Wikenheiser and Redish, 2011). Electrode arrays (Kopf; Tujunga, CA) containing 12 tetrodes and two reference electrodes were used.

Tetrodes were advanced slowly over approximately 1 week until estimates of electrode depth and electrophysiological signatures were consistent with the CA1 pyramidal layer. One reference was placed in the corpus callosum above hippocampus and one reference was placed in the hippocampal fissure. Data were recorded by a 64 channel Neuralynx Cheetah system (Bozeman, MT). The voltage on each tetrode channel was monitored at 32 kHz and filtered from 600–6000 Hz. When the voltage on any channel exceeded a preset threshold (100 μV), 1 ms of activity on each channel was saved to disk and time-stamped. Spikes were manually sorted into putative single units off-line using MClust 3.5 (Redish and Schmitzer-Torbert, 2008). Local field potentials were recorded from one channel per tetrode, sampled at 2 kHz and filtered from 1–425 Hz. Daily recording sessions continued for as long as large ensembles of well-isolated hippocampal neurons were recorded, with subjects cycling pseudorandomly through session types.

5.1.3 Data analysis

Only well-isolated unit recordings were included for analysis. We analyzed 1263 cells recorded across 26 sessions; the median isolation distance of clustered units was 32.65, and the median L-ratio was 0.02 (Schmitzer-Torbert et al., 2005). Ensemble sizes ranged from 25–62 simultaneously recorded neurons. All analyses were conducted using Matlab (MathWorks; Natick, MA).

Detection of theta state. Because our analyses focused on hippocampal representations during the theta network state, we were careful to exclude data from non-theta epochs (e.g. during the large, irregular activity network state, when sharp-wave ripples are prominent; Carr et al., 2011). All analyses of hippocampal theta were based on the LFP signal recorded from the hippocampal fissure, where theta amplitude is greatest. Sharp-wave/rippple analyses were based on recordings taken from the CA1 pyramidal cell layer. Recordings were first pre-processed to remove 0.5 s of data around any artifacts (instances of maximum/minimum voltage), and then bandpass filtered between 6–10 Hz to obtain the theta band signal, between 2–4 Hz to obtain the delta band signal, and between 140–220 Hz to obtain the sharpwave/ripple band signal. Instantaneous amplitude (spectral power) was estimated via the Hilbert transform. Theta cycles were defined as the time between peaks of the 6–10 Hz band-pass filtered fissure LFP. The

log-transformed ratio of theta to delta power was computed (Csicsvari et al., 1999; Jackson et al., 2006), and this quantity was averaged within each putative cycle of the theta rhythm. Only theta cycles with an average theta-delta ratio $>1\sigma$ the session average theta-delta ratio were included for analyses. Additionally, only theta cycles with a period corresponding to a 6–10 Hz oscillation were included. Spikes that occurred when ripple power was $>4\sigma$ the session average were excluded from analyses.

Place fields. Place fields detection followed described approaches (Gupta et al., 2010; Wikenheiser and Redish, 2011). Units with a session mean firing rate <0.05 Hz (non task-responsive cells) or >5 Hz (putative interneurons) were excluded from place field analyses. Spikes recorded when the animal's running speed was <5 cm/s were excluded. Similarly, spikes occurring during non-theta states were not used to compute place fields. The linearized maze was divided into approximately 2 cm bins and firing rate was computed within each spatial bin. Contiguous bins in which the firing rate was $\geq$15% the cell's session maximum firing rate were considered place fields. Fields detected in this way that were separated by $\leq$2 bins (approximately 4 cm) were merged. To avoid ambiguity about which spatial positions place cells were representing, neurons with multiple place fields were excluded from further analyses.

Theta sequence look-ahead. To measure theta sequence look-ahead distance, ensemble place cell activity was rotated around the animal's current position (as in fig. 5.3), and the look-ahead distance was taken as the average place field position of spikes that occurred in the final quarter of the theta cycle. In other words, the look-ahead distance of each theta cycle was the distance between the rat's location and the average of place field centers of cells active in the final quarter of that theta cycle, weighted by the number of spikes each cell fired. Using variations on this approach (e.g. taking the position of the final spike in each cycle; taking the position of the spike furthest forward of the animal, regardless of its timing within the theta cycle) produced similar estimates of look-ahead distance. In addition to the criteria described above for detection of theta states, look-ahead was computed only for theta cycles containing at least 3 spikes from a minimum of 2 cells.

Trajectories. Comparisons of look-ahead distance were made for data from journeys between feeders devoid of pauses or other behavioral irregularities. We isolated trajectories that occurred between rewarded visits to feeder sites (i.e. cases where subjects

left a site after receiving food and traveled to a site to await another food delivery). If running speed fell below 10 cm/s at any point after departure from the origin feeder site and before arrival at the destination feeder site the trajectory was excluded. Figure 5.1 shows examples of one-, two-, and three-segment trajectories initiated from each feeder site, and the relative frequencies of each trajectory type.

Movement controls. To rule out the possibility that subtle variations in rats' running speed or acceleration modulated look-ahead distance, we used a bootstrapping approach to construct model data sets under the assumption that movement parameters alone determined look-ahead. Every theta cycle used to construct fig. 5.4 was assigned a surrogate look-ahead distance, drawn randomly from a distribution of all look-ahead values measured when the animal was running at a similar speed or accelerating at a similar rate (fig. 5.8).

Place field size analyses. To test whether current spatial goals affected place field size (fig. 5.9a), we identified fields that rats traveled through primarily on one of the three trajectory types. For every spike a place cell fired, we determined which trajectory type the rat was completing at that moment; place cells with $\geq$90% of spikes occurring on a single trajectory type were included in this analysis.

To compute place field size on a trial-by-trial basis (fig. 5.9b), we measured the rat's location when the first and last spike occurred on each pass through the place field, relative to that place field's center, and determined the trajectory type that rats were completing. Mixed trajectory fields were defined as place cells with $\leq$60% of spikes accounted for by a single trajectory type, thus ensuring that a single mixed field contributed measurements to at least two trajectory types.

Place field expansion analyses. To measure backward expansion of place fields, we computed the average position of all the spikes a cell fired on a given pass through its place field. Mean spike positions for individual place field traversals were then normalized to fractional portion of the field, such the the beginning of the field corresponded to zero and the terminus of the field corresponded to one. For the analysis of the place field center of mass–lap number relationship (fig. 5.10), lines were fit only for trajectory types with $\geq$ 10 passes, to ensure reliable parameter estimates.

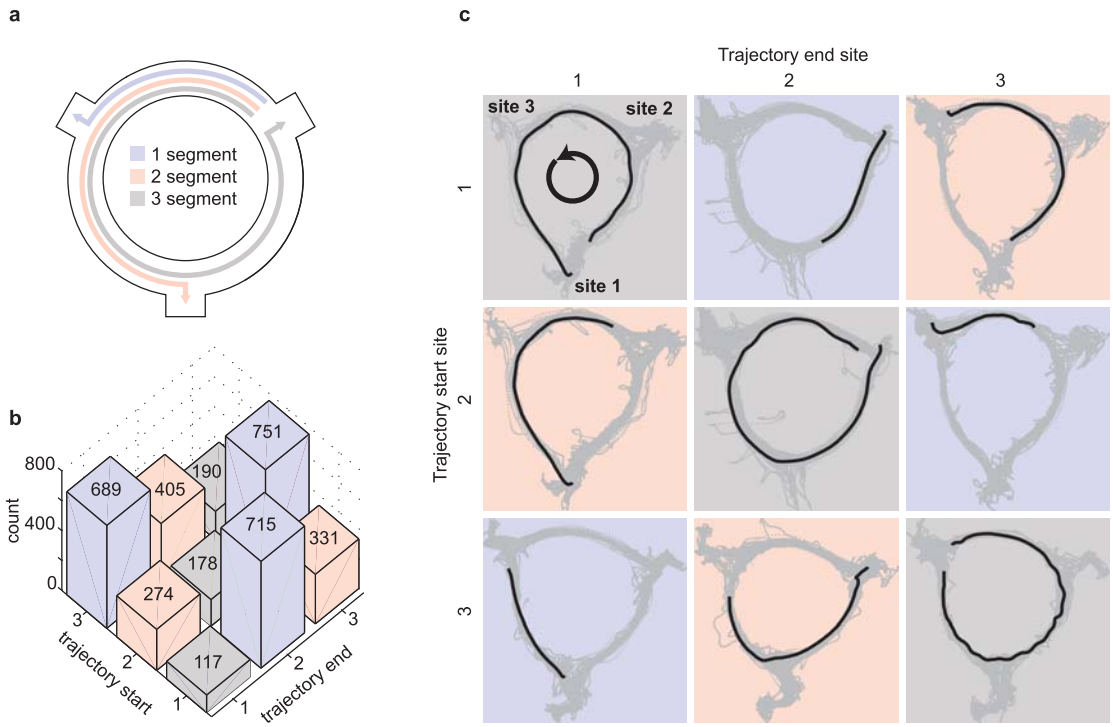

Figure 5.1: **Behavioral task. (a)** Rats allocated their time between three food delivery sites, each with unique delays that remained fixed within session, but varied across sessions. Rats ran unidirectional laps; thus, from any feeder site subjects could run a one-segment trajectory, a two-segment trajectory, or a three-segment trajectory. **(b)** The histogram indicates the frequency of trajectories beginning and ending at each physical feeder location (across all rats and sessions); trajectories are color-coded by type (one segment, two segment, or three segment). **c** Examples of trajectories beginning and ending at each feeder site are plotted. Position tracking data for the entire session is plotted in grey; data for a single trajectory in each square is plotted in black. Rats' direction of motion was counter-clockwise.

5.2 Results

5.2.1 Theta sequences reflect forthcoming behavior

Rats were trained to perform the intertemporal foraging task, in which they chose whether or not to wait for food delivered after varying amounts of delay. Rats ran unidirectional laps around a circular track with three food pellet dispensers spaced evenly around the perimeter, each associated with a fixed-length delay. If subjects remained at a feeder site until the delay period passed, food pellets (the same quantity at each feeder) were dispensed. Otherwise, the rat was free to move on to the next site. In either case, the feeder site became inactive until the rat approached it on the subsequent lap.

Within a session, the delay at each site was fixed; however, across sessions different sets of three delays were counterbalanced across the sites.

Rats ran different patterns of trajectories between sites, depending on the spatial arrangement of delays in the session and their willingness to wait for delayed reward (fig. 5.1). Beginning from a feeder site, subjects could perform any of three types of trajectories: running to and stopping at the next site completed a *one-segment* trajectory, skipping the next site and running to the feeder after that constituted a *two-segment* trajectory, and skipping the next two feeder sites, making a complete lap around the track, produced a *three-segment* trajectory.

Rats were implanted with tetrode arrays targeting dorsal, CA1 hippocampus. Consistent with previous reports (Dragoi and Buzsaki, 2006; Foster and Wilson, 2007; Gupta et al., 2012), place cell spiking was organized into theta sequences (figs. 5.2 & 5.3). We observed that in some cases, the trajectories represented by theta sequences were related to rats' forthcoming behavior on the task. For instance, figure 5.3a shows two sequences from different trials in the same session. In one trial (left panel) the rat ran to and stopped at the upper-left feeder, and the theta sequence that occurred as the rat traveled towards its destination represented a trajectory up to but not beyond the goal site. Later in the session (right panel), the rat skipped the upper left feeder site, and this decision was preceded by a theta sequence that traced a trajectory past the site. In figure 5.3b, theta sequences recorded at different positions along six trajectories, each beginning at the upper-right feeder and ending at the bottom-center feeder (i.e. two-segment trajectories skipping the upper-left site) are shown. Many of these sequences traced out trajectories extending beyond the rat's current location, ending near its eventual goal.

5.2.2 Theta sequence look-ahead was modulated by goals

To test whether theta sequences consistently related to upcoming behavior, we measured theta sequence look-ahead (the distance that theta sequences extended beyond the rat's current location; Gupta et al., 2012) as subjects performed the task. Data were divided into one-, two-, and three-segment trajectory cases, and theta look-ahead was examined as rats traversed the first portion of each trajectory (fig. 5.4a, shaded region), where behavior was consistent but rats' intended destination varied. Look-ahead

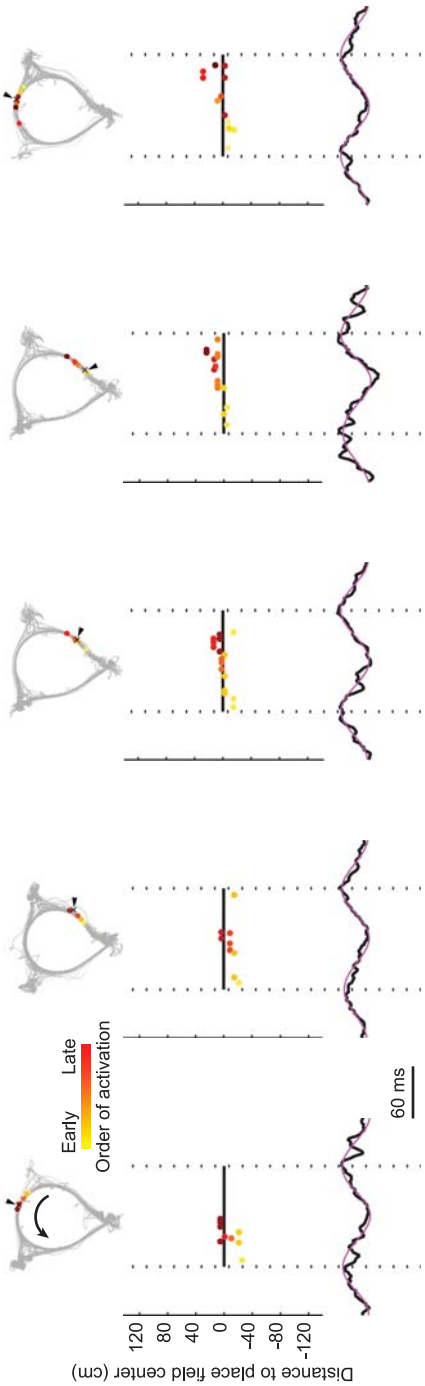

Figure 5.2: **Hippocampal ensemble spiking is organized into theta sequences.** Rasters of place cell spiking and two-dimensional projections of theta sequence trajectory representations are shown for five example theta sequences. In the raster plots, each row depicts the spiking of a single place cell, with position on the y-axis determined by the the location of the place field relative to the rat's location when the theta sequence occurred (rat's position is zero; positive numbers indicate the region of space in front of the rat; negative numbers indicate positions behind the rat). Spikes are color-coded to indicate the order in which they occurred. For 2-D spatial plots, each place cell that spiked during a theta sequence is represented with a dot at the center of its place field; dots are colored following the same convention as the rasters, to show the temporal evolution of the trajectory representation. The rat's location in space is marked with a black dot. The LFP (unfiltered, and filtered from 6–10 hz) is plotted below each raster.

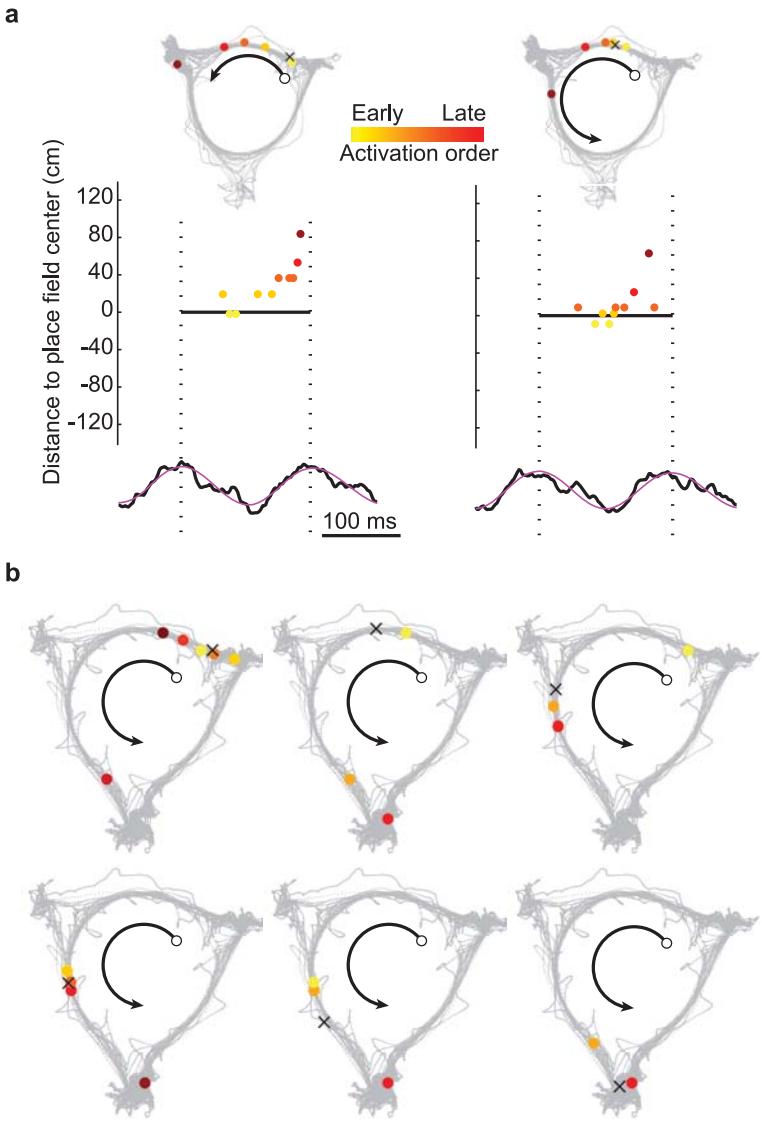

Figure 5.3: **Theta sequences reflect rats' future choices. (a)** When the rat was planning to stop and wait for food delivery at the upper left feeder (left panel) a theta sequence represented a trajectory up to that feeder. Later in the session, when the rat was about to skip the same feeder (right panel), the theta sequence representation instead traced a trajectory past the feeder. **(b)** Each box displays the spatial projection of a single theta sequence representation recorded during six separate journeys between the upper right and lower center feeders. Place cells near the goal destination were frequently active along with place cells near the rat's actual position.

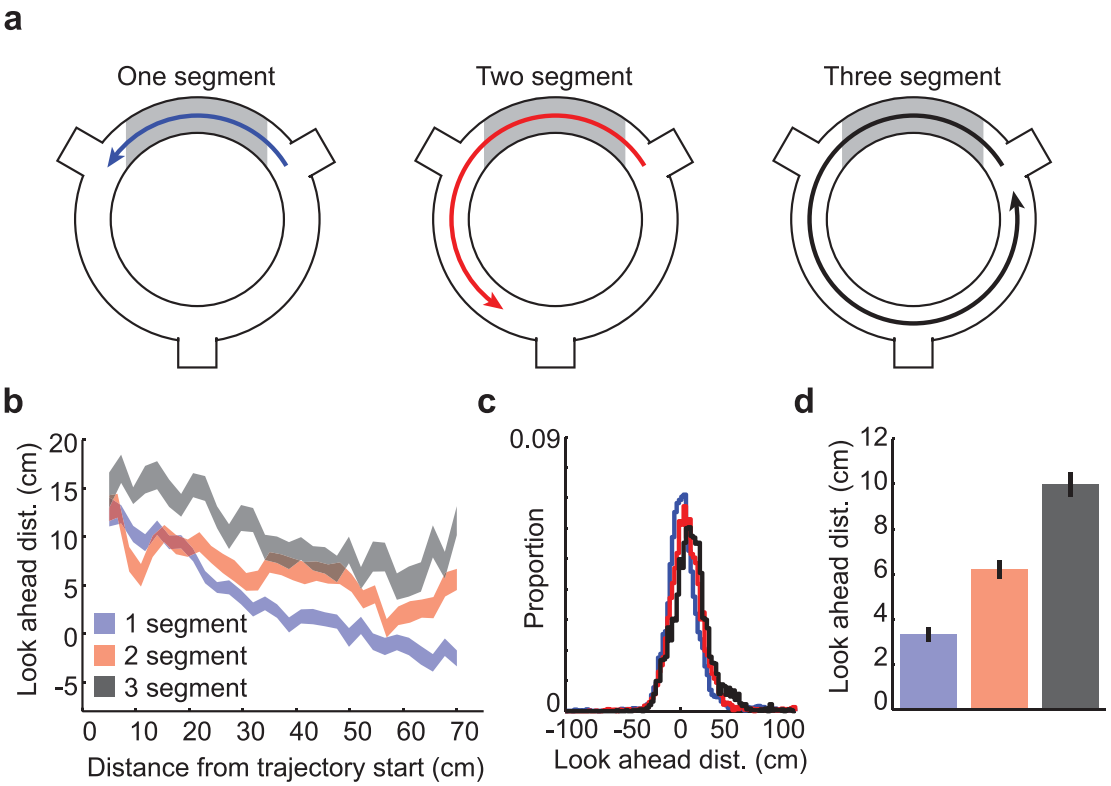

Figure 5.4: **Look-ahead distance varied with the length of planned trajectories.** **(a)** Data were aligned to trajectory initiation, split into three groups based on where the rat would stop, and examined over the initial limb of each trajectory (shaded region). For each trajectory type, plots display **(b)** the mean look-ahead distance across the initial portion of trajectories (± SEM; n = 20,271 theta cycles), **(c)** distributions of look-ahead distance, and **(d)** 95% confidence intervals.

distance was longest for three-segment trajectories and shortest for one-segment trajectories (fig. 5.4b). We performed a one-way ANOVA to test the effect of goal destination on theta look-ahead (fig. 5.4c). There was a significant effect of trajectory type on look-ahead distance ($F_{2,20268} = 172.09$, $p < 0.0001$), and post-hoc comparison via Tukey's HSD test indicated that look-ahead was greatest for three-segment trajectories, was intermediate for two-segment trajectories, and was shortest for one-segment trajectories (fig. 5.4d).

Theta sequence look-ahead showed a different pattern when aligned to arrival at a goal (fig. 5.5). As rats approached their goal destination at the end of a trajectory, look-ahead was not modulated by trajectory type (one-way ANOVA; $F_{2,20789} = 0.56$, *n.s.*). Thus, theta look-ahead depended on how far rats were from their goal destination, but

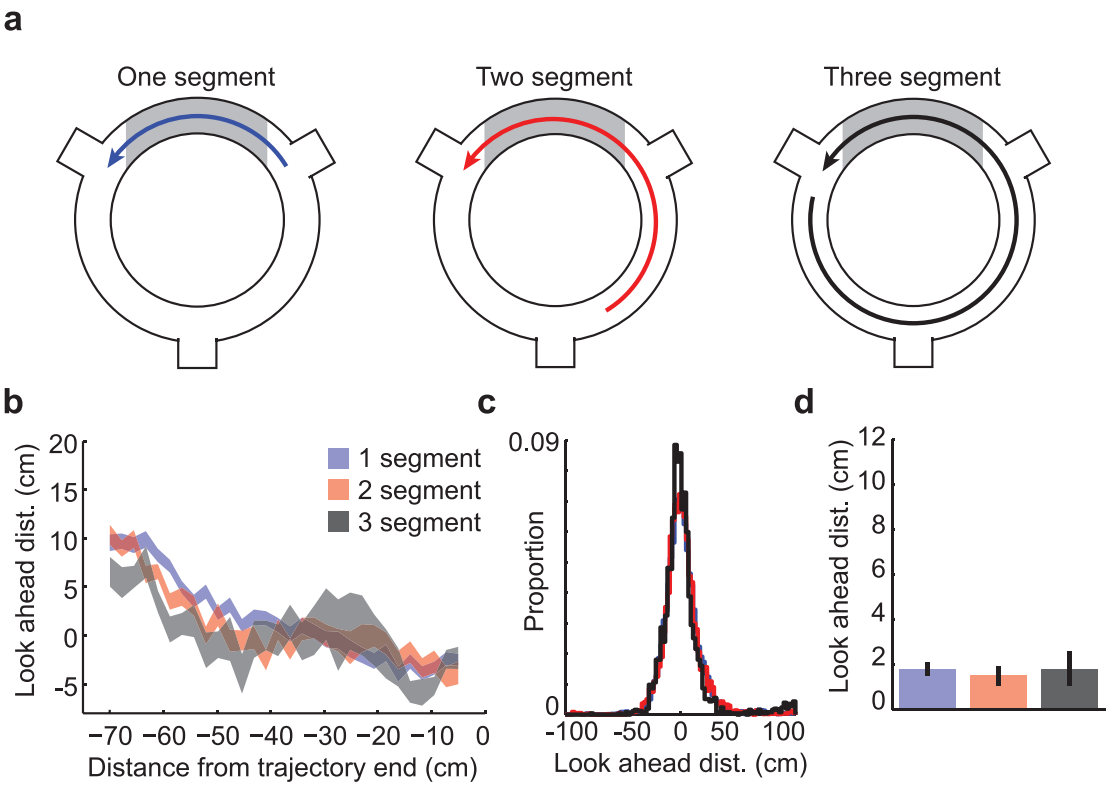

Figure 5.5: **Look-ahead did not vary on arrival to goal sites. (a)** Here, data were aligned to trajectory completion, as rats arrived at their goal destination, and examined over the final limb of each trajectory (shaded region). **(b)** Look-ahead distance did not differ on approach to the goal site (± SEM; $n = 20{,}792$ theta cycles), and distributions of look-ahead were overlapping across trajectory types **(c)**, as were 95% confidence intervals **(d)**.

not how far they had traveled to arrive in their current location, suggesting that theta sequences represented paths leading to goal destinations as rats performed the task.

This pattern of theta sequence look-ahead distance across trajectory types was similar for all individual rats in the study (fig. 5.6), and goal-dependent look-ahead modulation did not depend on the physical portion of the track at which trajectories began or ended (fig. 5.7).

Rats' running speed and acceleration profiles differed across trajectory types (speed: $F_{2,20268} = 310.69$, $p < 0.0001$; acceleration: $F_{2,20268} = 51.60$, $p < 0.0001$; one-way ANOVAs). To ensure that these variables did not account for goal-dependent differences in look-ahead distance, we used a bootstrapping procedure to generate surrogate data sets assuming that running speed or acceleration determined theta look-ahead (Gupta

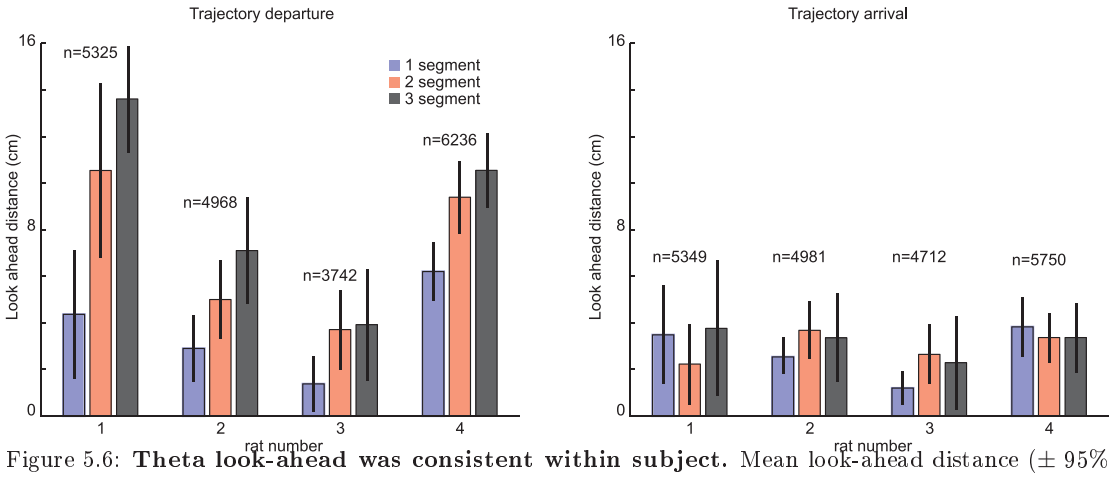

Figure 5.6: **Theta look-ahead was consistent within subject.** Mean look-ahead distance (± 95% confidence intervals) is plotted for individual rats on initiation of trajectories (left panel, cf. fig. 5.4) and on completion of trajectories (right panel, cf. fig. 5.5).

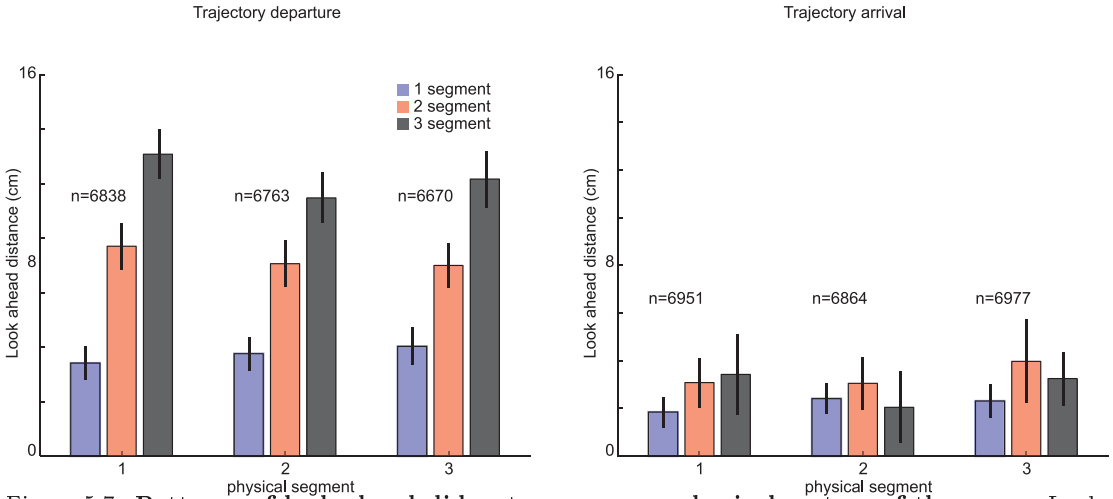

Figure 5.7: **Patterns of look-ahead did not vary across physical sectors of the maze.** Look-ahead distance was computed separately for one-, two-, and three-segment trajectories beginning or ending at each of the three physical feeder locations. Patterns were similar in all cases.

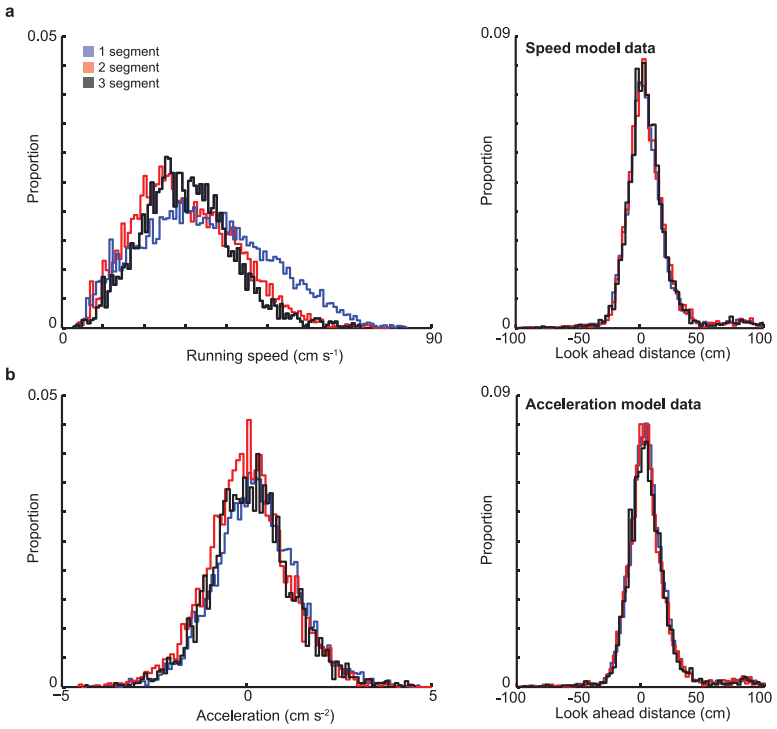

Figure 5.8: **Movement parameters did not account for differences in look-ahead on trajectory departure. (a)** Histograms display running speed (a, left panel) and acceleration (b, left panel) for theta sequences in figure 5.4. Because there were differences between the three trajectory types, we constructed surrogate theta look-ahead data sets (matched to the data used to construct fig. 5.4; $n =$ 20,269 theta cycles) assuming that speed (a, right panel) or acceleration (b, left panel) determined theta look-ahead. No goal-dependent effects were detected in these bootstrapped data sets.

et al., 2012). Neither of these model data sets showed a significant effect of goal location on look-ahead distance (fig. 5.8; speed model: $F_{2,20268} = 0.76$, *n.s.*; acceleration model: $F_{2,20268} = 0.33$, *n.s.*; one-way ANOVA). These results suggest that differences in theta sequence look-ahead were driven by representations related to rats' intended goal destination, rather than sensory or motor variables.

5.2.3 Single cell consequences of goal-dependent theta look-ahead

On a single cell level, increased theta look-ahead distance implies an earlier activation of neurons as subjects approached place fields, and consequently predicts that place fields lying primarily on longer trajectories should be larger than those on short trajectories. We identified place fields situated such that rats passed through them almost exclusively

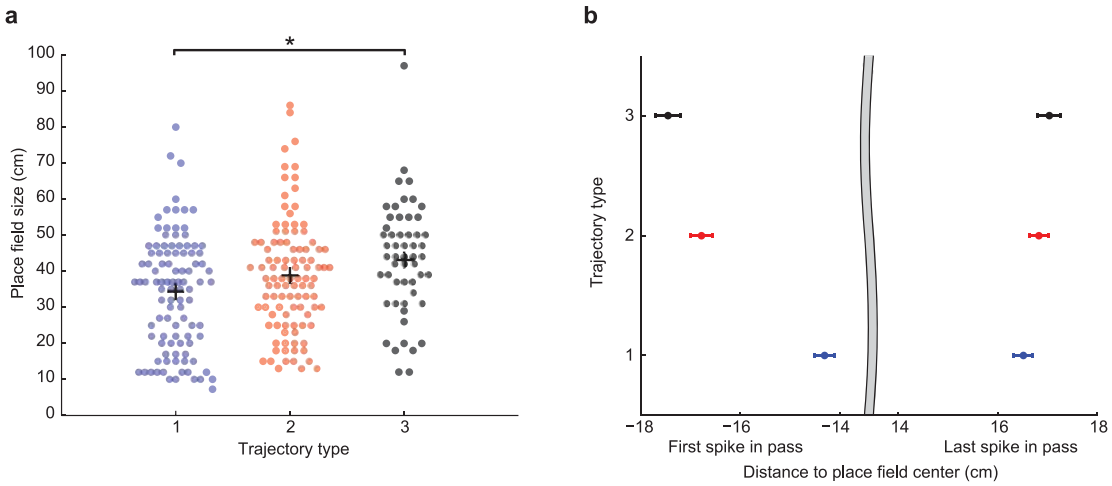

Figure 5.9: **Place field size was modulated by trajectory type. (a)** Consistent with trajectory-dependent modulation of theta look-ahead, place fields that rats traversed solely on three-segment trajectories were significantly larger than fields that rats passed through only during one-segment trajectories. Black crosses indicate the means of distributions. **(b)** Place field size varied similarly with goal location on single trials. The mean position of the first spike of passes through place fields during three-segment trajectories occurred significantly further in front of the place field center than on one-segment trajectory passes. The mean location of the last spike also varied across trajectory types, but to a lesser extent. Error bars indicate SEM.

during completion of only one trajectory type ($n = 258$ fields), and compared their sizes (fig. 5.9a). A one-way ANOVA identified a significant effect of trajectory type on place field size ($F_{2,255} = 5.53$, $p < 0.01$), and a post-hoc Tukey's HSD test showed that place fields on three-segment trajectories were significantly larger than those on one-segment trajectories (on average approximately 20%, or 10 cm, longer).

If theta look-ahead was modulated on a moment-by-moment basis, place fields that rats traversed en route to multiple, different goal sites would be expected to vary in size from trial to trial, depending on whether rats were bound for a nearby or distant goal site. To quantify place field size on a trial-by-trial basis, we measured the distance between each place field's center and the rat's position when the place cell fired its first and last spike on each pass through the field (fig. 5.9b). Consistent with goal modulation of look-ahead distance, the location of the first place cell spike on each pass varied significantly depending on rats' intended destination ($F_{2,10656} = 122.02$, $p < 0.0001$; one-way ANOVA), with Tukey's HSD post-hoc test showing that the initial spike on three-segment trajectory passes was shifted significantly towards the early portion of

the place field relative to the initial spike on one-segment passes. Importantly, this result held when only mixed-case place fields (fields rats traversed while completing more than one trajectory type; $n = 104$ fields) were included ($F_{2,2886} = 6.99$, $p < 0.001$; one-way ANOVA), strongly suggesting that the trajectory dependence of look-ahead distance we report here was due to modulation of place cell activation on single trials, and not due to group-wise differences in the place fields associated with particular trajectory types.

While the position of the last spike on passes through place fields also varied with goal destination ($F_{2,10656} = 3.49$, $p = 0.03$; one-way ANOVA), the difference across trajectory types was smaller than the difference in initial spike location (fig. 5.9b), suggesting that increased look-ahead distance primarily affected the early portion of the place field, leaving the end intact (Jensen and Lisman, 1996; Lisman and Jensen, 2013).

Previous work has shown that many place cells exhibit asymmetric expansion over time (Mehta and McNaughton, 1996; Mehta et al., 1997, 2000), with place field center of mass (COM) shifting in a direction opposite of the animal's movement, resulting in a negative relationship between COM and number of passes through the place field. To examine how asymmetric expansion interacted with look-ahead modulation, we computed the COM of place cell spiking for each pass through a place field, and fit lines to the relationship between lap number and COM, separately for each trajectory type (fig. 5.10a). Asymmetric expansion of place fields predicts that such lines would have a characteristic (negative) slope that is unaffected by current goal, while trial-to-trial look-ahead modulation predicts that line intercepts should vary with goal destination. Indeed, lines fit for many place cells had a slope less than zero, consistent with the backward expansion effect; however, a one-way ANOVA found no effect of trajectory type on line slope ($F_{2,488} = 0.10$, *n.s.*; fig. 5.10b), but showed that the intercepts of line fits for each goal destination varied significantly ($F_{2,488} = 18.72$, $p < 0.0001$). Post-hoc testing with Tukey's HSD test found that intercepts for three-segment trajectories were shifted significantly further toward the initial portion of the place field compared to one-segment trajectories. These results held when only mixed-case place fields (fields rats passed through en route to more than one goal site) were compared (slope: $F_{2,267} = 0.21$, *n.s.*; intercept: $F_{2,267} = 7.23$, $p < 0.0001$; one-way ANOVA), suggesting that look-ahead modulation occurred on a single-trial basis, in tandem with, but separable from the place field expansion phenomenon.

5.3 Discussion

The look-ahead distance of hippocampal theta sequence representations was increased as rats executed more lengthy spatial trajectories. In addition to representing animals' location as they performed the task, CA1 theta sequences were also embedded with information about current spatial goals. A previous study found that theta sequence properties were influenced by prominent features of the environment (Gupta et al., 2012), resulting in a cognitively-parsed, "chunked" representation of space. Here, we show that the motivational significance subjects attached to landmarks (i.e. whether or not they would skip a feeder site) similarly modulated the expression of theta sequences. These results bolster Gupta and colleagues' (2012) claim that modulation of theta sequences reflects a top-down, cognitive process; because the feeder locations used in this experiment were identical, a bottom-up process common to any salient environmental feature would have influenced theta sequence representations around all three goals equivalently, regardless of subjects' intentions. Instead, we found that sequence look-ahead and single cell place field firing properties reflected rats' goal destinations on a trial-by-trial basis, creating a processed representation of space that reflects on-going behavior.

These findings support models that ascribe a predictive role to hippocampal theta sequences. Although the function of theta sequences and phase precession remain unclear, a growing body of evidence suggests that hippocampal ensemble spiking performs a moment-by-moment prediction of upcoming spatial positions across the theta cycle (Jensen and Lisman, 1996; Lisman and Jensen, 2013). Such representations have clear utility in decision making situations: with a representation of immanent environmental features, subjects could adapt their behavior in real time based on upcoming features, or track progress towards some goal destination to ensure that a previously-devised plan is carried out correctly.

Consistent with a role for the hippocampus in actively coordinating ongoing behavior, we found that theta sequences reliably reflected rats' future trajectories; it remains unclear, however, whether such representations are critical for task performance, or simply a read-out of behavior. Robbe and colleagues (2006; 2009) used a pharmacological approach to show that disrupting the temporal organization of theta sequences (while minimally affecting other place cell properties) reversibly abolished correct performance

of a delayed-alternation T-maze task. Because rats in their study had achieved asymptotic performance before theta sequences were manipulated, a role for theta sequences in decision making (as opposed to facilitating or modulating learning-related processes) seems likely.

Previous studies have reported hippocampal representations related to goals or future behavior during quiescence, coincident with sharp-wave ripple (SWR) complexes (Diba and Buzsáki, 2007; Pfeiffer and Foster, 2013; Singer et al., 2013), and disruption of hippocampal spiking during SWRs impairs performance of a hippocampally-dependent, spatial memory task (Jadhav et al., 2013). Interestingly, both our results here and previous studies have shown that hippocampal representations related to upcoming behaviors also occur during the theta network state (Johnson and Redish, 2007; Pastalkova et al., 2008; Takahashi et al., 2009, 2014). The relationship between theta- and SWR-associated hippocampal planning signals remains untested (Schmidt and Redish, 2013). That hippocampal theta sequences continue to look ahead towards the target goal throughout a journey implies that the rodent hippocampus maintains plans online, as behavior is executed.

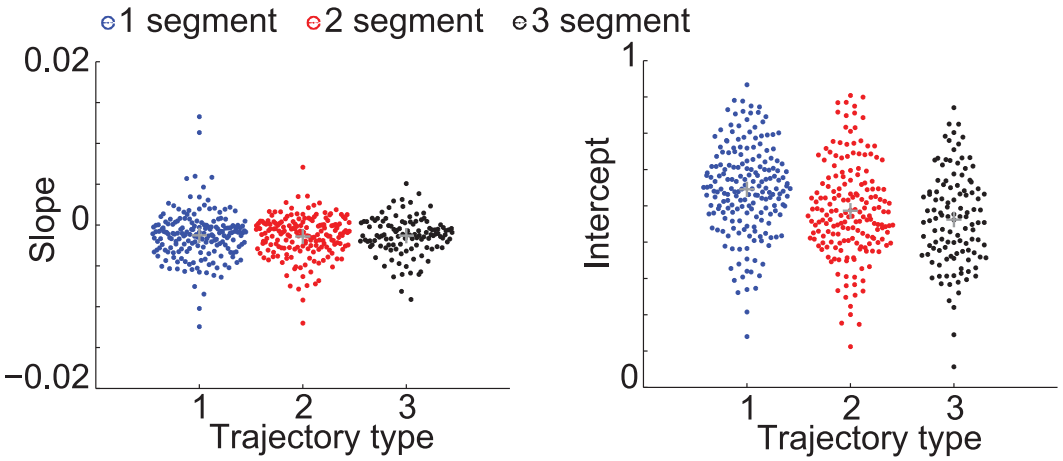

Figure 5.10: **Look-ahead was modulated by goal location on single trials. (a)** Scatter plots (left panels) show the spiking of two place cells that rats passed through on trajectories of different lengths throughout the course of the session. Circles mark the location of the rat when spikes occurred, and color indicates the trajectory type the rat was completing. Place field COM (normalized to span from 0 at the beginning of the field to 1 at the end of the field) is plotted for each lap. **(b)** We fit lines to the COM–lap number relationship as plotted in panel **a**, separately for each trajectory type. The slopes of these lines were not modulated by trajectory type, while line intercepts were significantly forward-shifted for longer, three-segment trajectories, consistent with trial-by-trial modulation of look-ahead distance. Grey crosses indicate the means of distributions.

References

D. Addis and D. Schacter. The hippocampus and imagining the future: Where do we stand? *Frontiers in Human Neuroscience*, 5, 2012.

D. Addis, T. Cheng, R. Roberts, and D. Schacter. Hippocampal contributions to the episodic simulation of specific and general future events. *Hippocampus*, 21:1045–1052, 2011.

G. Ainslie. Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, 82(4):463–496, 1975.

G. W. Ainslie. Impulse control in pigeons. *Journal of the experimental analysis of behavior*, 21(3):485–489, 1974.

D. Amaral and P. Lavenex. Hippocampal neuroanatomy. In P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe, editors, *The Hippocampus Book*, pages 37–114. Oxford, 2007.

D. G. Amaral. Memory: anatomical organization of candidate brain regions. In F. Plum, editor, *Handbook of Physiology; The Nervous System*, Progress in Brain Research, pages 211–294. American Physiological Society, Bethesda, 1987.

D. G. Amaral and M. P. Witter. The three-dimensional organization of the hippocampal formation: A review of anatomical data. *Neuroscience*, 31(3):571–591, 1989.

P. Andersen. Organization of hippocampal neurons and their interconnections. In *The hippocampus*, pages 155–175. Springer, 1975.

H. Arkes. The psychology of waste. *Behavioral Decision Making*, 9:213–224, 1996.

H. Arkes and P. Ayton. The sunk cost and Concorde effects: are humans less rational than lower animals? *Psychological Bulletin*, 125:591–600, 1999.

H. Arkes and C. Blumer. The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, 35:124–140, 1985.

E. Aronson. The effect of effort on the attractiveness of rewarded and unrewarded stimuli. *Journal of Abnormal and Social Psychology*, 63:375–380, 1961.

B. W. Balleine. Incentive processes in instrumental conditioning. In S. B. Klein and R. R. Mowrer, editors, *Handbook of Contemporary Learning Theories*, pages 307–366. LEA, 2001.

H. Barron, R. Dolan, and T. Behrens. Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature neuroscience*, 16(10):1492–1498, 2013.

M. Bateson and A. Kacelnik. Rate currencies and the foraging starling: the fallacy of the averages revisited. *Behavioral Ecology*, 7(3):341–352, 1996.

W. M. Baum and H. C. Rachlin. Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, 12(6):861–874, 1969.

H. M. Bayer and P. Glimcher. Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47:129–141, 2005.

R. Bellman. On a routing problem. *Quarterly Journal of Applied Mathematics*, 16(1): 87–90, 1958.

L. S. Benardo and D. A. Prince. Acetylcholine induced modulation of hippocampal pyramidal neurons. *Brain research*, 211(1):227–234, 1981.

G. S. Berns, J. Chappelow, M. Cekic, C. F. Zink, G. Panoni, and M. E. Martin-Skurski. Neurobiological substrates of dread. *Science*, 312(5774):754–758, 2006.

G.-Q. Bi and M.-M. Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neuroscience*, 18(10464-10472), 1998.

K. W. Bieri, K. N. Bobbitt, and L. L. Colgin. Slow and fast gamma rhythms coordinate different spatial coding modes in hippocampal place cells. *Neuron*, 2014.

T. C. Blanchard, J. M. Pearson, and B. Y. Hayden. Postreward delays and systematic biases in measures of animal temporal discounting. *Proceedings of the National Academy of Sciences, USA*, 110(38):15491–15496, 2013.

A. M. Bornstein and N. D. Daw. Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLoS computational biology*, 9 (12):e1003387, 2013.

J. Brockner. The escalation of commitment to a failing course of action: toward theoretical progress. *Academy of Management Review*, 17:39–61, 1992.

E. Bromberg-Martin and O. Hikosaka. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, 63(1):119, 2009.

E. N. Brown, L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *Journal of Neuroscience*, 18(18):7411–7425, 1998.

R. Buckner. The role of the hippocampus in prediction and imagination. *Annual Review of Psychology*, 61:27–48, 2010.

R. Bush and F. Mosteller. A mathematical model for simple learning. *Psychol. Rev.*, 58:313–323, 1951.

R. Bush and F. Mosteller. A mathematical model for simple learning. In S. Fienberg and D. Hoaglin, editors, *Selected Papers of Frederick Mosteller*, Springer Series in Statistics, pages 221–234. Springer New York, 2006.

G. Buzsáki. The "where is it?" reflex: Autoshaping the orienting response. *Journal of the Experimental Analysis of Behavior*, 37(3):461–484, 1982.

G. Buzsáki. Two-stage model of memory trace formation: A role for "noisy" brain states. *Neuroscience*, 31(3):551–570, 1989.

G. Buzsáki. Theta rhythm of navigation: link between path integration and landmark navigation, episodic and semantic memory. *Hippocampus*, 15(7):827–840, 2005.

G. Buzsáki. *Rhythms of the Brain*. Oxford, 2006.

G. Buzsáki, L. W. Leung, and C. H. Vanderwolf. Cellular bases of hippocampal EEG in the behaving rat. *Brain Research*, 287(2):139–171, 1983.

G. Buzsáki, C. A. Anastassiou, and C. Koch. The origin of extracellular fields and currents: Eeg, ecog, lfp and spikes. *Nature Reviews Neuroscience*, 13(6):407–420, 2012.

A. L. Calvert, L. Green, and J. Myerson. Discounting in pigeons when the choice is between two delayed rewards: implications for species comparisons. *Behavioral and neuroscientific analysis of economic decision making in animals*, page 38, 2011.

R. N. Cardinal, N. Daw, T. Robbins, and B. J. Everitt. Local analysis of behaviour in the adjusting-delay task for choice of delayed reinforcement. *Neural Networks*, 15: 617–634, 2002.

M. F. Carr, S. P. Jadhav, and L. M. Frank. Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature Neuroscience*, 14 (2):147–153, 2011.

A. C. Catania. Concurrent performances: a baseline for the study of reinforcement magnitude. *Journal of the Experimental Analysis of Behavior*, 6(2):299–300, 1963.

A. Chadwick, M. C. van Rossum, and M. F. Nolan. Independent theta phase coding accounts for ca1 population sequences and enables flexible remapping. *bioRxiv*, 2014.

E. Charnov. Optimal foraging, the marginal value theorum. *Theoretical Population Biology*, 9:129–136, 1976a.

E. Charnov. Optimal foraging: Attack strategy of a mantid. *The American Naturalist*, 110:141–151, 1976b.

E. L. Charnov and G. H. Orians. Optimal foraging: some theoretical explorations. Unpublished Manuscript, 1973.

S. Cheng and L. M. Frank. New experiences enhance coordinated neural activity in the hippocampus. *Neuron*, 57:303–313, 2008.

S.-H. Chung and R. J. Herrnstein. Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior*, 10(1):67–74, 1967.

N. J. Cohen and H. Eichenbaum. *Memory, Amnesia, and the Hippocampal System*. MIT Press, Cambridge, MA, 1993.

L. L. Colgin. Mechanisms and functions of theta rhythms. *Annual review of neuroscience*, 36:295–312, 2013.

L. L. Colgin, E. I. Moser, and M. B. Moser. Understanding memory through hippocampal remapping. *Trends in Neurosciences*, 31(9):469–477, 2008.

L. L. Colgin, T. Denninger, M. Fyhn, T. Hafting, T. Bonnevie, O. Jensen, M. B. Moser, and E. I. Moser. Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature*, 462(7271):353–357, 2009.

J. Csicsvari, H. Hirase, A. Czurkó, and G. Buzsáki. Oscillatory coupling of hippocampal pyramidal cells and interneurons in the behaving rat. *Journal of Neuroscience*, 19(1): 274–287, 1999.

P. Dasgupta and E. Maskin. Uncertainty and hyperbolic discounting. *American Economic Review*, pages 1290–1299, 2005.

T. J. Davidson, F. Kloosterman, and M. A. Wilson. Hippocampal replay of extended experience. *Neuron*, 63:497–507, 2009.

N. D. Daw. *Reinforcement learning models of the dopamine system and their behavioral implications*. PhD thesis, Carnegie Mellon University, 2003.

N. D. Daw and D. S. Touretzky. Long-term reward prediction in TD models of the dopamine system. *Neural Computation*, 14:2567–2583, 2002.

N. D. Daw, Y. Niv, and P. Dayan. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8:1704–1711, 2005.

N. D. Daw, A. C. Courville, and D. S. Touretzky. Representation and timing in theories of the dopamine system. *Neural Computation*, 18:1637–1677, 2006.

P. Dayan and Y. Niv. Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology*, 18(2):185–196, 2008.

M. Z. Deluty. Self-control and impulsiveness involving aversive events. *Journal of Experimental Psychology: Animal Behavior Processes*, 4(3):250, 1978.

D. Derdikman and M.-B. Moser. A dual role for hippocampal replay. *Neuron*, 65(5): 582–584, 2010.

K. Diba and G. Buzsáki. Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience*, 10:1241–1242, 2007.

A. Dickinson and B. Balleine. Motivational control of goal-directed action. *Animal Learning & Behavior*, 22(1):1–18, 1994.

G. Dragoi and G. Buzsaki. Temporal encoding of place sequences by hippocampal cell assemblies. *Neuron*, 50(1):145–157, 2006.

G. Dragoi and S. Tonegawa. Preplay of future place cell sequences by hippocampal cellular assemblies. *Nature*, 469:397–401, 2011.

G. Dragoi and S. Tonegawa. Distinct preplay of multiple novel spatial experiences in the rat. *Proceedings of the National Academy of Sciences, USA*, 110(22):9100–9105, 2013.

Y. Dudai and M. Eisenberg. Rites of passage of the engram: Reconsolidation and the lingering consolidation hypothesis. *Neuron*, 44(1):93–100, 2004.

V. Ego-Stengel and M. A. Wilson. Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus*, 20(1):1–10, 2010.

H. Eichenbaum. Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, 44:109–120, 2004.

H. Eichenbaum. Memory on time. *Trends in Cognitive Sciences*, 2013.

H. Eichenbaum, P. Dudchenko, E. Wood, M. Shapiro, and H. Tanila. The hippocampus, memory, and place cells: Is it spatial memory or a memory space? *Neuron*, 23: 209–226, 1999.

A. D. Ekstrom, J. Meltzer, B. L. McNaughton, and C. A. Barnes. NMDA receptor antagonism blocks experience-dependent expansion of hippocampal "place fields". *Neuron*, 31:631–638, 2001.

J. Epsztein, M. Brecht, and A. K. Lee. Intracellular determinants of hippocampal ca1 place and silent cell activity in a novel environment. *Neuron*, 70(1):109–120, 2011.

A. A. Fenton and R. U. Muller. Place cell discharge is extremely variable during individual passes of the rat through the firing field. *Proceedings of the National Academy of Sciences, USA*, 95:3182–3187, 1998.

A. A. Fenton, W. W. Lytton, J. M. Barry, P.-P. Lenck-Santini, L. E. Zinyuk, S. Kubik, J. Bures, B. Poucet, R. U. Muller, and A. V. Olypher. Attention-like modulation of hippocampus place cell discharge. *J. Neurosci.*, 30(13):4613–4625, 2010. doi: 10.1523/JNEUROSCI.5576-09.2010.

D. Foster and J. Knierim. Sequence learning and the role of the hippocampus in rodent navigation. *Current Opinion in Neurobiology*, 22:294–300, 2012.

D. J. Foster and M. A. Wilson. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, 440(7084):680–683, 2006.

D. J. Foster and M. A. Wilson. Hippocampal theta sequences. *Hippocampus*, 17:1093–1099, 2007.

S. Frederick, G. Loewenstein, and T. O'donoghue. Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401, 2002.

E. Freidin and A. Kacelnik. Rational choice, context dependence, and the value of information in european starlings (*Sturnus vulgaris*). *Science*, 334:1000–1002, 2011.

T. F. Freund and M. Antal. Gaba-containing neurons in the septum control inhibitory interneurons in the hippocampus. *Nature*, 336(6195):170–173, 1988.

T. F. Freund and G. Buzsáki. Interneurons of the hippocampus. *Hippocampus*, 6(4): 345–370, 1996.

M. Fyhn, T. Hafting, A. Treves, M.-B. Moser, and E. I. Moser. Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446:190–194, 2007.

C. Gallistel, A. King, D. Gottlieb, F. Balci, E. Papachristos, M. Szalecki, and K. Carbone. Is matching inate? *Journal of the Experimental Analysis of Behavior*, 87: 161–199, 2007.

T. Gener, L. Perez-Mendez, and M. V. Sanchez-Vives. Tactile modulation of hippocampal place fields. *Hippocampus*, 23(12):1453–1462, 2013.

P. R. Gill, S. J. Mizumori, and D. M. Smith. Hippocampal episode fields develop with learning. *Hippocampus*, 21(11):1240–1249, 2011.

G. Girardeau and M. Zugaro. Hippocampal ripples and memory consolidation. *Current Opinion in Neurobiology*, 21:452–459, 2011.

G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzs'aki, and M. B. Zugaro. Selective suppression of hippocampal ripples impairs spatial memory. *Nature Neuroscience*, 12: 1222–1223, 2009.

P. W. Glimcher, C. Camerer, and R. A. Poldrack, editors. *Neuroeconomics: Decision Making and the Brain.* Academic Press, 2008.

C. M. Gray, P. E. Maldonado, M. Wilson, and B. L. McNaughton. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings on cat striate cortex. *Journal of Neuroscience Methods*, 63:43–54, 1995.

J. Green and A. Arduini. Hippocampal electrical activity in arousal. *Journal of Neurophysiology*, 17:531–557, 1954.

L. Green, E. Fisher, S. Perlow, and L. Sherman. Preference reversal and self control: Choice as a function of reward amount and delay. *Behaviour Analysis Letters*, 1981.

H. J. Groenewegen, E. Vermeulen-Van der Zee, A. te Kortschot, and M. P. Witter. Organization of the projections from the subiculum to the ventral striatum in the rat. A

study using anterograde transport of *phaseolus vulgaris* leucoagglutinin. *Neuroscience*, 23(1):103–120, 1987.

A. Gupta, M. van der Meer, D. Touretzky, and A. Redish. Segmentation of spatial experience by hippocampal θ sequences. *Nature Neuroscience*, 15:1032–1039, 2012.

A. S. Gupta, M. A. A. van der Meer, D. S. Touretzky, and A. D. Redish. Hippocampal replay is not a simple function of experience. *Neuron*, 65(5):695–705, 2010.

E. R. Guthrie. *The Psychology of Learning (revised edition)*. Harpers, New York, 1952.

J. F. Guzowski, B. L. McNaughton, C. A. Barnes, and P. F. Worley. Environment-specific expression of the immediate-early gene *ARC* in hippocampal neuronal ensembles. *Nature Neuroscience*, 2(12):1120–1124, 1999.

K. D. Harris, D. A. Henze, H. Hirase, X. Leinekugel, G. Dragol, A. Czurkó, and G. Buzsáki. Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature*, 417:738–741, 2002.

D. Hassabis, D. Kumaran, and E. Maguire. Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience*, 27:14365–74, 2007a.

D. Hassabis, D. Kumaran, S. D. Vann, and E. A. Maguire. Patients with hippocampal amnesia cannot imagine new experiences. *PNAS*, 104:1726–1731, 2007b.

B. Hayden, J. Pearson, and M. Platt. Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, 14:4178–4187, 2011.

D. O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949. Reissued 2002 LEA.

S. E. Henly, A. Ostdiek, E. Blackwell, S. Knutie, A. S. Dunlap, and D. W. Stephens. The discounting-by-interruptions hypothesis: model and experiment. *Behavioral Ecology*, 19(1):154–162, 2008.

R. Herrnstein. Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4(3):267–272, 1961.

R. J. Herrnstein. On the law of effect. *Journal of the Experimental Analysis of Behavior*, 13(2):243–266, 1970.

R. J. Herrnstein. *The Matching Law*. Harvard Univ Press, 1997.

P. A. Hetherington and M. L. Shapiro. Hippocampal place fields are altered by the removal of single visual cues in a distance-dependent manner. *Behavioral neuroscience*, 111(1):20, 1997.

A. J. Hill and P. J. Best. Effects of deafness and blindness on the spatial correlates of hippocampal unit activity in the rat. *Experimental neurology*, 74(1):204–217, 1981.

C. S. Holling. Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist*, 91(07):385–398, 1959.

S. A. Hollup, S. Molden, J. G. Donnett, M.-B. Moser, and E. I. Moser. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, 21(5):1635–1644, 2001.

J. Holmes and W. R. Adey. Electrical activity of the entorhinal cortex during conditioned behavior. *American Journal of Physiology–Legacy Content*, 199(5):741–744, 1960.

C. L. Hull. *Principles of behavior*. Appleton-Century-Crofts, New York, 1943.

J. R. Huxter, T. J. Senior, K. Allen, and J. Csicsvari. Theta phase-specific codes for two-dimensional position, trajectory and heading in the hippocampus. *Nature Neuroscience*, 11:587–594, 2008.

J. Jackson and A. D. Redish. Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. *Hippocampus*, 17:1209–1229, 2007.

J. C. Jackson, A. Johnson, and A. D. Redish. Hippocampal sharp waves and reactivation during awake states depend on repeated sequential experience. *Journal of Neuroscience*, 26:12415–12426, 2006.

S. Jadhav, C. Kemere, P. German, and L. Frank. Awake hippocampal sharp-wave ripples support spatial memory. *Science*, 336:1454–1458, 2013.

T. M. Jay and M. P. Witter. Distribution of hippocampal CA1 and subicular efferents in the prefrontal cortex of the rat studied by means of anterograde transport of *Phaseolus vulgaris*-leucoagglutinin. *The Journal of Comparative Neurology*, 313(4):574–586, 2004.

K. Jeffery, A. Gilbert, S. Burton, and A. Strudwick. Preserved performance in a hippocampal-dependent spatial task despite complete place cell remapping. *Hippocampus*, 13:133–147, 2003.

K. J. Jeffery, R. L. Anand, and M. I. Anderson. A role for terrain slope in orienting hippocampal place fields. *Experimental brain research*, 169(2):218–225, 2006.

O. Jensen and J. E. Lisman. Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells. *Learning and Memory*, 3(2-3): 279–287, 1996.

O. Jensen and J. E. Lisman. Position reconstruction from an ensemble of hippocampal place cells: contribution of theta phase encoding. *Journal of Neurophysiology*, 83(5): 2602–2609, 2000.

W. S. Jevons. *Theory of Political Economy*. Macmillan, 1871.

A. Johnson and D. Crowe. Revisiting tolman: Theories and cognitive maps. *Cognitive Critique*, 1(1):43–72, 2009.

A. Johnson and A. D. Redish. Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. *Neural Networks*, 18(9):1163–1171, 2005.

A. Johnson and A. D. Redish. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience*, 27(45):12176–12189, 2007.

A. Johnson, M. A. A. van der Meer, and A. D. Redish. Integrating hippocampus and striatum in decision-making. *Current Opinion in Neurobiology*, 17(6):692–697, 2007.

A. Johnson, A. A. Fenton, C. Kentros, and A. D. Redish. Looking for cognition in the structure in the noise. *Trends in Cognitive Sciences*, 13(2):55–64, 2009.

R. Jung and A. Kornmüller. Eine methodik der ableitung iokalisierter potentialschwankungen aus subcorticalen hirngebieten. *European Archives of Psychiatry and Clinical Neuroscience*, 109(1):1–30, 1938.

A. Kacelnik. Normative and descriptive models of decision making: time discounting and risk sensitivity. In G. R. Bock and G. Cardew, editors, *Characterizing Human Psychological Adaptations*, volume 208 of *Ciba Foundation Symposia*, pages 51–66. Wiley, Chichester UK, 1997. Discussion 67-70.

D. Kahneman and A. Tversky. Choices, values, and frames. *American psychologist*, 39 (4):341, 1984.

T. Kalenscher and C. M. A. Pennartz. Is a bird in the hand worth two in the future? the neuroeconomics of intertemporal decision-making. *Progress in Neurobiology*, 84: 284–315, 2008.

L. J. Kamin. Predictability, surprise, attention, and conditioning. In B. A. Campbell and R. M. Church, editors, *Punishment and AVersive Behavior*, pages 279–296. Appleton-Century-Crofts, New York, 1969.

M. P. Karlsson and L. M. Frank. Awake replay of remote experiences in the hippocampus. *Nature Neuroscience*, 12:913–918, 2009.

C. G. Kentros, N. T. Agnihotri, S. Streater, R. D. Hawkins, and E. R. Kandel. Increased attention to spatial context increases both place field stability and spatial memory. *Neuron*, 42:283–295, 2004.

P. Killeen. The matching law. *Journal of the Experimental Analysis of Behavior*, 17(3): 489–495, 1972.

S. Kim and L. Frank. Hippocampal lesions impair rapid learning of a continuous spatial alternation task. *PLoS ONE*, 4:e5494, 2009.

J. Konorski and S. Miller. On two types of conditioned reflex. *The Journal of General Psychology*, 16(1):264–272, 1937.

R. Kramis, C. Vanderwolf, and B. Bland. Two types of hippocampal rhythmical slow activity in both the rabbit and the rat: Relations to behavior and effects of atropine, diethyl ether, urethane, and pentobarbital. *Experimental Neurology*, 49:58–85, 1975.

B. J. Kraus, R. J. Robinson II, J. A. White, H. Eichenbaum, and M. E. Hasselmo. Hippocampal "time cells": time versus path integration. *Neuron*, 78(6):1090–1101, 2013.

J. Krebs, A. Kacelnik, and P. Taylor. Test of optimal sampling by foraging great tits. *Nature*, 275:27–31, 1978.

H. S. Kudrimoti, C. A. Barnes, and B. L. McNaughton. Reactivation of hippocampal cell assemblies: Effects of behavioral state, experience, and EEG dynamics. *Journal of Neuroscience*, 19(10):4090–4101, 1999.

Z. Kurth-Nelson and A. D. Redish. Temporal-difference reinforcement learning with distributed representations. *PLoS ONE*, 4(10):e7362, 2009.

Z. Kurth-Nelson and A. D. Redish. A reinforcement learning model of pre-commitment in decision making. *Frontiers in Behavioral Neuroscience*, 4:184ff, 2010.

Z. Kurth-Nelson, W. Bickel, and A. D. Redish. A theoretical account of cognitive effects in delay discounting. *European Journal of Neuroscience*, 35(7):1052–1064, 2012.

D. Laibson. Hyperbolic discounting and golden eggs. *Quarterly Journal of Economics*, 112(2):443–477, 1997.

A. K. Lee and M. A. Wilson. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, 36:1183–1194, 2002.

D. Lee, B.-J. Lin, and A. K. Lee. Hippocampal place fields emerge upon single-cell manipulation of excitability during behavior. *Science*, 337(6096):849–853, 2012.

P.-P. Lenck-Santini, A. A. Fenton, and R. U. Muller. Discharge properties of hippocampal neurons during performance of a jump avoidance task. *J. Neurosci.*, 28(27):6773–6786, 2008.

J. K. Leutgeb, S. Leutgeb, M.-B. Moser, and E. I. Moser. Pattern Separation in the Dentate Gyrus and CA3 of the Hippocampus. *Science*, 315(5814):961–966, 2007. doi: 10.1126/science.1135801.

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.

D. Levcik, T. Nekovarova, A. Stuchlik, and D. Klement. Rats use hippocampus to recognize positions of objects located in an inaccessible space. *Hippocampus*, 23:153–161, 2013.

C. Lever, T. Wills, F. Cacucci, N. Burgess, and J. O'Keefe. Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature*, 416:90–94, 2002.

W. B. Levy, A. B. Hocking, and X. B. Wu. Interpreting hippocampal function as recoding and forecasting. *Neural Netowrks*, 18:1242–1264, 2005.

P. R. Lewis and C. Shute. The cholinergic limbic system: projections to hippocampal formation, medial cortex, nuclei of the ascending cholinergic reticular system, and the subfornical organ and supra-optic crest. *Brain*, 90(3):521–540, 1967.

J. Lisman and A. D. Redish. Prediction, sequences and the hippocampus. *Philosophical Transactions of the Royal Society B Biological Sciences*, 364:1193–1201, 2009.

J. E. Lisman and O. Jensen. The theta-gamma neural code. *Neuron*, 77(6):1002–1016, 2013.

G. Loewenstein. The fall and rise of psychological explanations in the economics of intertemporal choice. *Russell Sage Foundation Publications*, 1992.

G. Loewenstein and D. Prelec. Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, 107(2):573–597, 1992.

A. Logue, M. E. Smith, and H. Rachlin. Sensitivity of pigeons to prereinforcer and postreinforcer delay. *Animal Learning & Behavior*, 13(2):181–186, 1985.

R. Lorento do Nó. Studies on the structure of the cerebral cortex II. Continuation of the study of the ammonic system. *J. Psychol. Neurol.*, 46:113–177, 1934.

K. MacCorquodale and P. E. Meehl. On a distinction between hypothetical constructs and intervening variables. *Psychological review*, 55(2):95, 1948.

C. J. MacDonald, K. Q. Lepage, U. T. Eden, and H. Eichenbaum. Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, 71(4):737–749, 2011.

T. V. Maia. Reinforcement learning, conditioning, and the brain: Successes and challenges. *Cognitive, Affective, & Behavioral Neuroscience*, 9(4):343–364, 2009.

E. J. Markus, Y. Qin, B. Leonard, W. E. Skaggs, B. L. McNaughton, and C. A. Barnes. Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *Journal of Neuroscience*, 15:7079–7094, 1995.

D. Marr. Simple memory: A theory of archicortex. *Philosophical Transactions of the Royal Society of London*, 262(841):23–81, 1971.

A. P. Maurer and B. L. McNaughton. Network and intrinsic cellular mechanisms underlying theta phase precession of hippocampal neurons. *Trends in Neurosciences*, 30 (7):325–333, 2007.

A. P. Maurer, S. N. Burke, P. Lipa, W. E. Skaggs, and C. A. Barnes. Greater running speeds result in altered hippocampal phase sequence dynamics. *Hippocampus*, 22(4): 737–747, 2012.

J. Mazur. An adjusting procedure for studying delayed reinforcement. In M. Commons, J. Mazur, J. Nevin, and H. Rachlin, editors, *Quantitative analyses of behavior: The effect of delay and of intervening events*, pages 55–73. Hillsdale, NJ: Erlbaum., 1987.

J. Mazur. Theories of probabilistic reinforcement. *Journal of Experimental Analysis of Behavior*, 51(1):87–99, 1989.

B. L. McNaughton, C. A. Barnes, and J. O'Keefe. The contributions of position, direction, and velocity to single unit activity in the hippocampus of freely-moving rats. *Experimental Brain Research*, 52:41–49, 1983.

B. L. McNaughton, B. Leonard, and L. Chen. Cortical-hippocampal interactions and cognitive mapping: A hypothesis based on reintegration of the parietal and inferotemporal pathways for visual processing. *Psychobiology*, 17(3):230–235, 1989.

B. L. McNaughton, C. A. Barnes, J. L. Gerrard, K. Gothard, M. W. Jung, J. J. Knierim, H. Kudrimoti, Y. Qin, W. E. Skaggs, M. Suster, and K. L. Weaver. Deciphering the hippocampal polyglot: The hippocampus as a path integration system. *Journal of Experimental Biology*, 199(1):173–186, 1996.

M. R. Mehta and B. L. McNaughton. Rapid changes in hippocampal population code during behavior: A case for Hebbian learning *in vivo*. Presented at *CNS*96*, (the fifth annual *Computation in Neural Systems* meeting) , 1996.

M. R. Mehta, C. A. Barnes, and B. L. McNaughton. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences, USA*, 94:8918–8921, 1997.

M. R. Mehta, M. C. Quirk, and M. A. Wilson. Experience-dependent asymmetric shape of hippocampal receptive fields. *Neuron*, 25:707–715, 2000.

M. R. Mehta, A. K. Lee, and M. A. Wilson. Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, 417:741–746, 2002.

V. M. Miller and P. J. Best. Spatial correlates of hippocampal unit activity are altered by lesions of the fornix and entorhinal cortex. *Brain Research*, 194:311–323, 1980.

M. Minsky. Steps toward artificial intelligence. *Proceedings of the IRE*, 49(1):8–30, 1961.

M. A. Moita, S. Rosis, Y. Zhou, J. E. LeDoux, and H. T. Blair. Hippocampal place cells acquire location-specific responses to the conditioned stimulus during auditory fear conditioning. *Neuron*, 37(3):485–497, 2003.

M. A. Moita, S. Rosis, Y. Zhou, J. E. LeDoux, and H. T. Blair. Putting fear in its place: Remapping of hippocampal place cells during fear conditioning. *Journal of Neuroscience*, 24(31):7015–7023, 2004.

P. R. Montague, P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, 16(5):1936–1947, 1996.

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*, 7:1951–1968, 1987.

R. U. Muller, J. L. Kubie, and J. B. Ranck, Jr. Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *Journal of Neuroscience*, 7:1935–1950, 1987.

Z. Nádasdy, H. Hirase, A. Czurkó, J. Csicsvari, and G. Buzsáki. Replay and time compression of recurring spike sequences in the hippocampus. *Journal of Neuroscience*, 19(2):9497–9507, 1999.

L. Nerad and N. McNaughton. The septal eeg suggests a distributed organization of the pacemaker of hippocampal theta in the rat. *European Journal of Neuroscience*, 24(1): 155–166, 2006.

A. J. Neuringer. Effects of reinforcement magnitude on choice and rate of responding. *Journal of the Experimental Analysis of Behavior*, 10(5):417–424, 1967.

Y. Niv. Cost, benefit, tonic, phasic. what do response rates tell us about dopamine and motivation? *Annals of the New York Academy of Sciences*, 1104(1):357–376, 2007.

Y. Niv, N. D. Daw, D. Joel, and P. Dayan. Tonic dopamine: oppportunity costs and the control of response vigor. *Psychopharmacology*, 191(3):507–520, 2006a.

Y. Niv, D. Joel, and P. Dayan. A normative perspective on motivation. *Trends in Cognitive Sciences*, 10(8):375–381, 2006b.

J. O'Keefe and D. H. Conway. Hippocampal place units in the freely moving rat: Why they fire where they fire. *Experimental Brain Research*, 31:573–590, 1978.

J. O'Keefe and J. Dostrovsky. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely moving rat. *Brain Research*, 34:171–175, 1971.

J. O'Keefe and L. Nadel. *The Hippocampus as a Cognitive Map*. Clarendon Press, Oxford, 1978.

J. O'Keefe and M. Recce. Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, 3:317–330, 1993.

A. V. Olypher, P. Lánský, and A. A. Fenton. Properties of the extra-positional signal in hippocampal place cell discharge derived from the overdispersion in location-specific firing. *Neuroscience*, 111(3):553–566, 2002.

S. M. O'Mara, S. Commins, M. Anderson, and J. Gigg. The subiculum: a review of form, physiology and function. *Progress in neurobiology*, 64(2):129–155, 2001.

J. O'Neill, T. Senior, and J. Csicsvari. Place-selective firing of ca1 pyramidal cells during sharp wave/ripple network patterns in exploratory behavior. *Neuron*, 49:143–155, 2006.

J. O'Neill, B. Pleydell-Bouverie, D. Dupret, and J. Csicsvari. Play it again: reactivation of waking experience and memory. *Trends in Neurosciences*, 33:220–229, 2010.

A. Papale, J. Stott, N. Powell, P. Regier, and A. Redish. Interactions between deliberation and delay-discounting in rats. *Cognitive Affective and Behavioral Neuroscience*, 12:513–526, 2012.

E. Pastalkova, V. Itskov, A. Amarasingham, and G. Buzsaki. Internally generated cell assembly sequences in the rat hippocampus. *Science*, 321(5894):1322–1327, 2008. doi: 10.1126/science.1159775.

C. Pavlides and J. Winson. Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *Journal of Neuroscience*, 9(8):2907–2918, 1989.

I. Pavlov. *Conditioned Reflexes*. Oxford Univ Press, 1927.

J. M. Pearson, B. Y. Hayden, and M. L. Platt. Explicit information reduces discounting behavior in monkeys. *Frontiers in Comparative Psychology*, 1, 2010. doi: 10.3389/fpsyg.2010.00237.

W. D. Penny, P. Zeidman, and N. Burgess. Forward and backward inference in spatial cognition. *PLoS computational biology*, 9(12):e1003383, 2013.

C. T. Perin. The effect of delayed reinforcement upon the differentiation of bar responses in white rats. *Journal of Experimental Psychology*, 32(2):95, 1943.

J. Peters and C. Büchel. Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, 66(1): 138–148, 2010.

B. Pfeiffer and D. Foster. Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447):74–79, 2013.

E. S. Phelps and R. A. Pollak. On second-best national saving game-equilibrium growth. *Review of Economic Studies*, 35(2), 1968.

L. Pompilio and A. Kacelnik. Context-dependent utility overrides absolute memory as a determinant of choice. *Proceedings of the National Academy of Sciences, USA*, 107: 508–512, 2010.

G. H. Pyke, H. R. Pulliam, and E. Charnov. Optimal foraging: a selective review of theory and tests. *Quarterly Review of Biology*, 52:137–154, 1977.

G. J. Quirk, R. U. Muller, and J. L. Kubie. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *Journal of Neuroscience*, 10(6): 2008–2017, 1990.

H. Rachlin. On the tautology of the matching law. *Journal of the Experimental Analysis of Behavior*, 15(2):249–251, 1971.

H. Rachlin and L. Green. Commitment, choice, and self-control. *JEAB*, 17(1):15–22, 1972.

J. Rae. *Statement of some new principles on the subject of political economy: exposing the fallacies of the system of free trade, and of some other doctrines maintained in the "Wealth of nations."*. Hillard, Gray, 1834.

J. Rae and C. W. Mixter. *The sociological theory of capital*. Macmillan, 1905.

S. Ramón y Cajal. Histology of the nervous system of man and vertebrates. *Oxford Univ. Press, New York*, 1911. Republished in 1995; translation by N. Swanson and L.W. Swanson.

A. D. Redish. *Beyond the cognitive map: Contributions to a computational neuroscience theory of rodent navigation.* PhD thesis, Carnegie Mellon University, 1997.

A. D. Redish. *Beyond the Cognitive Map: From Place Cells to Episodic Memory.* MIT Press, Cambridge MA, 1999.

A. D. Redish. Addiction as a computational process gone awry. *Science*, 306(5703): 1944–1947, 2004.

A. D. Redish and Z. Kurth-Nelson. Neural models of temporal discounting. In G. Madden and W. Bickel, editors, *Impulsivity: The Behavioral and Neurological Science of Discounting*, pages 123–158. APA books, 2010.

A. D. Redish and N. C. Schmitzer-Torbert. MCLUST spike sorting toolbox, version 3.0. *http://redishlab.neuroscience.umn.edu/* , 2008.

A. D. Redish, F. O. Battaglia, M. K. Chawla, A. D. Ekstrom, J. L. Gerard, P. Lipa, E. S. Rosenzweig, P. F. Worley, J. F. Guzowski, B. L. McNaughton, and C. A. Barnes. Independence of firing correlates of anatomically proximate hippocampal pyramidal cells. *Journal of Neuroscience*, 21(RC134):1–6, 2001.

A. D. Redish, S. Jensen, A. Johnson, and Z. Kurth-Nelson. Reconciling reinforcement learning models with behavioral extinction and renewal: Implications for addiction, relapse, and problem gambling. *Psychological Review*, 114(3):784–805, 2007.

R. A. Rescorla. Pavlovian conditioned inhibition. *Psychological Bulletin*, 72(2):77, 1969.

R. A. Rescorla. Second-order conditioning: implications for theories of learning. 1973.

R. A. Rescorla and A. R. Wagner. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokesy, editors, *Classical Conditioning II: Current Research and Theory*, pages 64–99. Appleton Century Crofts, New York, 1972.

D. Robbe and G. Buzsaki. Alteration of Theta Timescale Dynamics of Hippocampal Place Cells by a Cannabinoid Is Associated with Memory Impairment. *J. Neurosci.*, 29(40):12597–12605, 2009. doi: 10.1523/JNEUROSCI.2407-09.2009.

D. Robbe, S. M. Montgomery, A. Thome, P. E. Rueda-Orozco, B. L. McNaughton, and G. Buzsaki. Cannabinoids reveal importance of spike timing coordination in hippocampal function. *Nature Neuroscience*, 9:1526–1533, 2006.

M. R. Roesch, D. J. Calu, , and G. Schoenbaum. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, 10:1615–1624, 2007.

A. G. Rosati, J. R. Stevens, B. Hare, and M. D. Hauser. The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. *Current Biology*, 17(19):1663–1668, 2007.

D. Rowland, Y. Yanovich, and C. Kentros. A stable hippocampal representation of a space requires its direct experience. *Proceedings of the National Academy of Sciences, USA*, 108:14654–14658, 2011.

A. Rubenstein. "economics and psychology"? the case of hyperbolic discounting. *International Economic Review*, 44(4):1207–1216, 2003.

G. A. Rummery and M. Niranjan. *On-line Q-learning using connectionist systems*. University of Cambridge, Department of Engineering, 1994.

A. V. Samsonovich and G. A. Ascoli. A simple neural network model of the hippocampus suggesting its pathfinding role in episodic memory retrieval. *Learning and Memory*, 12(2):193–208, 2005.

A. V. Samsonovich and B. L. McNaughton. Path integration and cognitive mapping in a continuous attractor neural network model. *Journal of Neuroscience*, 17(15): 5900–5920, 1997.

P. A. Samuelson. A note on measurement of utility. *The Review of Economic Studies*, 4(2):155–161, 1937.

E. Save, A. Cressant, C. Thinus-Blanc, and B. Poucet. Spatial firing of hippocampal place cells in blind rats. *Journal of Neuroscience*, 18(5):1818–1826, 1998.

D. Schacter and D. Addis. On the nature of medial temporal lobe contributions to the constructive simulation of future events. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521):1245, 2009.

D. L. Schacter, D. R. Addis, and R. L. Buckner. Remembering the past to imagine the future: the prospective brain. *Nature Reviews Neuroscience*, 8:657–661, 2007.

D. L. Schacter, D. R. Addis, and R. L. Buckner. Episodic simulation of future events. concepts, data, and applications. *Annals of the New York Academy of Sciences*, 1124: 39–60, 2008.

K. Schaffer. Beitrag zur histologie der Ammons Hornformation. *Archiv für Mikroscopische Anatomie*, 39(1), 1892.

B. Schmidt and A. D. Redish. Neuroscience: Navigation with a cognitive map. *Nature*, 2013.

N. Schmitzer-Torbert, J. Jackson, D. Henze, K. Harris, and A. Redish. Quantitative measures of cluster quality for use in extracellular recordings. *Neuroscience*, 131(1): 1–11, 2005.

W. Schultz, P. Dayan, and R. Montague. A neural substrate of prediction and reward. *Science*, 275:1593–1599, 1997.

N. Schweighofer, K. Shishida, C. E. Han, Y. O. S. C. T. S. Yamawaki, and K. Doya. Humans can adopt optimal discounting strategy under real-time constraints. *PLoS Computational Biology*, 2(11):e152, 2006.

T. Seidenbecher, T. R. Laxmi, O. Stork, and H.-C. Pape. Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. *Science*, 301(5634):846–850, 2003.

N. W. Senior. *An outline of the science of political economy.* W. Clowes and Sons, Stamford Street, 1836.

M. L. Shapiro, H. Tanila, and H. Eichenbaum. Cues that hippocampal place cells encode: dynamic and hierarchical representation of local and distal stimuli. *Hippocampus*, 7 (6):624–642, 1997.

G. M. Shepherd and C. Koch. Introduction to synaptic circuits. In G. M. Shepherd, editor, *The Synaptic Organization of the Brain*, pages 3–31. Oxford University Press, 2nd edition, 1990.

D. Shohamy and N. B. Turk-Browne. Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General*, 142(4):1159, 2013.

A. Singer, M. Carr, M. Karlsson, and L. Frank. Hippocampal swr activity predicts correct decisions during the initial learning of an alternation task. *Neuron*, 77(6): 1163–1173, 2013.

A. C. Singer and L. M. Frank. Rewarded outcomes enhance reactivation of experience in the hippocampus. *Neuron*, 64(6):910–921, 2009. doi: DOI:10.1016/j.neuron.2009.11.016.

W. E. Skaggs and B. L. McNaughton. Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science*, 271:1870–1873, 1996.

W. E. Skaggs, B. L. McNaughton, M. A. Wilson, and C. A. Barnes. Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus*, 6(2):149–173, 1996.

B. F. Skinner. Two types of conditioned reflex and a pseudo type. *The Journal of General Psychology*, 12(1):66–77, 1935.

P. D. Sozou. On hyperbolic discounting and uncertain hazard rates. *The Royal Society London B*, 265:2015–2020, 1998.

L. R. Squire and S. M. Zola-Morgan. The medial temporal lobe memory system. *Science*, 253:1380–1386, 1991.

B. Staw. Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance*, 16:27–44, 1976.

B. Staw and F. Fox. Escalation: the determinants of commitment to a chosen course of action. *Human Relations*, 30:431–450, 1977.

A. P. Steiner and A. D. Redish. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nature Neuroscience*, 2014.

D. Stephens. Decision ecology: Foraging and the ecology of animal decision making. *Cognitive, Affective, and Behavioral Neuroscience*, 8(4):475–484, 2008.

D. Stephens and D. Anderson. The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. *Behavioral Ecology*, 12(3):330–339, 2001.

D. Stephens and J. Krebs. *Foraging Theory*. Princeton, 1986.

D. W. Stephens, J. S. Brown, and R. C. Ydenberg. *Foraging: behavior and ecology*. University of Chicago Press, 2007.

J. Stevens, E. Hallinan, and M. Hauser. The ecology and evolution of patience in two new world monkeys. *Biology Letters*, 1(2):223–226, 2005.

I. H. Stevenson and K. P. Kording. How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139–142, 2011.

M. Suddendorf and M. Corballis. The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, 30:299–312, 2007.

T. Suddendorf and M. C. Corballis. Behavioural evidence for mental time travel in nonhuman animals. *Behavioural Brain Research*, 215:292–298, 2010.

G. R. Sutherland and B. L. McNaughton. Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current Opinion in Neurobiology*, 10(2):180–6, 2000.

R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.

R. S. Sutton and A. G. Barto. *Reinforcement Learning: An introduction*. MIT Press, Cambridge MA, 1998.

L. Swanson, P. Sawchenko, and W. Cowan. Evidence that the commissural, associational and septal projections of the regio inferior of the hippocampus arise from the same neurons. *Brain research*, 197(1):207–212, 1980.

M. Takahashi, J. Lauwereyns, Y. Sakurai, and M. Tsukada. A Code for Spatial Alternation During Fixation in Rat Hippocampal CA1 Neurons. *J Neurophysiol*, 102(1): 556–567, 2009. doi: 10.1152/jn.91159.2008.

M. Takahashi, H. Nishida, A. D. Redish, and J. Lauwereyns. Theta phase shift in spike timing and modulation of gamma oscillation: A dynamic code for spatial alternation during fixation in rat hippocampal area ca1. *Journal of neurophysiology*, 2014.

S. C. Tanaka, K. Doya, G. Okada, K. Ueda, Y. Okamoto, and S. Yamawaki. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nature Neuroscience*, 7:887–893, 2004.

H. Tanila, M. L. Shapiro, and H. Eichenbaum. Discordance of spatial representation in ensembles of hippocampal place cells. *Hippocampus*, 7(6):613–623, 1997.

L. T. Thompson and P. J. Best. Place cells and silent cells in the hippocampus of freely-behaving rats. *Journal of Neuroscience*, 9(7):2382–2390, 1989.

L. T. Thompson and P. J. Best. Long-term stability of the place-field activity of single units recorded from the dorsal hippocampus of freely behaving rats. *Brain Research*, 509(2):299–308, 1990.

E. L. Thorndike. *Animal intelligence: Experimental studies*. Macmillan, 1911.

E. C. Tolman. *Purposive Behavior in Animals and Men*. Appleton-Century-Crofts, New York, 1932.

E. C. Tolman. The determiners of behavior at a choice point. *Psychological Review*, 45 (1):1–41, 1938.

E. C. Tolman. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychological Review*, 46:318–336, 1939.

E. C. Tolman. Cognitive maps in rats and men. *Psychological Review*, 55:189–208, 1948.

E. C. Tolman, B. F. Ritchie, and D. Kalish. Studies in spatial learning. I. Orientation and the short-cut. *Journal of Experimental Psychology*, 36:13–24, 1946.

D. S. Touretzky and A. D. Redish. A theory of rodent navigation based on interacting representations of space. *Hippocampus*, 6(3):247–270, 1996.

M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, and B. L. McNaughton. Population dynamics and theta rhythm phase precession of hippocampal place cell firing: A spiking neuron model. *Hippocampus*, 6(3):271–280, 1996.

E. Tulving and S. Madigan. Memory and verbal learning. *Annual Review of Psychology*, 21:437–484, 1970.

M. A. A. van der Meer and A. D. Redish. Covert expectation-of-reward in rat ventral striatum at decision points. *Frontiers in Integrative Neuroscience*, 3(1):1–15, 2009.

M. A. A. van der Meer and A. D. Redish. Expectancies in decision making, reinforcement learning, and ventral striatum. *Frontiers in Neuroscience*, 2010. doi: 10.3389/neuro.01/006.2010.

T. van Groen and J. M. Wyss. The postsubicular cortex in the rat: characterization of the fourth region of the subicular cortex and its connections. *Brain research*, 529(1): 165–177, 1990.

C. Vanderwolf. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalography and Clinical Neurophysiology*, 26:407–418, 1969.

C. H. Vanderwolf. Limbic-diencephalic mechanisms of voluntary movement. *Psychological Review*, 78(2):83–113, 1971.

A. Viard, C. F. Doeller, T. Hartley, C. M. Bird, and N. Burgess. Anterior hippocampus and goal-directed spatial decision making. *The Journal of Neuroscience*, 31(12):4613–4621, 2011.

E. von Böhm-Bawerk. *Capital and interest: A critical history of economical theory.* Macmillan, 1890.

P. Waelti, A. Dickinson, and W. Schultz. Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, 412:43–48, 2001. doi: 10.1038/35083500.

C. Watkins. *Learning from delayed rewards*. PhD thesis, Cambridge University, 1989.

C. J. Watkins and P. Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

A. Wikenheiser, D. Stephens, and A. Redish. Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. *Proceedings of the National Academy of Sciences, USA*, 110(20):8308–8313, 2013.

A. M. Wikenheiser and A. D. Redish. Changes in reward contingency modulate the trial to trial variability of hippocampal place cells. *Journal of Neurophysiology*, 106 (2):589–598, 2011.

A. M. Wikenheiser and A. D. Redish. Hippocampal sequences link past, present, and future. *Trends in Cognitive Sciences*, 16(7):361–362, 2012.

A. M. Wikenheiser and A. D. Redish. The balance of forward and backward hippocampal sequences shifts across behavioral states. *Hippocampus*, 23(1):22–29, 2013.

M. A. Wilson and B. L. McNaughton. Dynamics of the hippocampal ensemble code for space. *Science*, 261:1055–1058, 1993.

M. A. Wilson and B. L. McNaughton. Reactivation of hippocampal ensemble memories during sleep. *Science*, 265:676–679, 1994.

G. Wimmer and D. Shohamy. The striatum and beyond: hippocampal contributions to decision making. In M. Delgado, E. Phelps, and T. Robbins, editors, *Attention & Performance XXII*, pages 281–310. Oxford.

A. Ylinen, A. Bragin, Z. Nadasdy, G. Jando, I. Szabo, A. Sik, and G. Buzsáki. Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: Network and intracellular mechanisms. *Journal of Neuroscience*, 15(1):30–46, 1995.

B. J. Young, G. D. Fox, and H. Eichenbaum. Correlates of hippocampal complex-spike cell activity in rats performing a nonspatial radial maze task. *Journal of Neuroscience*, 14(11):6553–6563, 1994.

F. D. Zeeb, S. B. Floresco, and C. A. Winstanley. Contributions of the orbitofrontal cortex to impulsive choice: interactions with basal levels of impulsivity, dopamine signalling, and reward-related cues. *Psychopharmacology*, 211(1):87–98, 2010.

K. Zhang, I. Ginzburg, B. L. McNaughton, and T. J. Sejnowski. Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *Journal of Neurophysiology*, 79:1017–1044, 1998.

S. Zhang and D. Manahan-Vaughan. Spatial olfactory learning contributes to place field formation in the hippocampus. *Cerebral Cortex*, page bht239, 2013.

L. Zinyuk, S. Kubik, Y. Kaminsky, A. A. Fenton, and J. Bures. Understanding hippocampal activity by using purposeful behavior: place navigation induces place cell discharge in both task-relevant and task-irrelevant spatial reference frames. *Proceedings of the National Academy of Sciences, USA*, 97(7):3771–3776, 2000.

Y. Ziv, L. D. Burns, E. D. Cocker, E. O. Hamel, K. K. Ghosh, L. J. Kitch, A. El Gamal, and M. J. Schnitzer. Long-term dynamics of ca1 hippocampal place codes. *Nature neuroscience*, 16(3):264–266, 2013.