EJN fens European Journal of Neuroscience

*European Journal of Neuroscience, Vol. 35, pp. 1052–1064, 2012* doi:10.1111/j.1460-9568.2012.08058.x

# A theoretical account of cognitive effects in delay discounting

Zeb Kurth-Nelson,[1] Warren Bickel[2] and A. David Redish[3]
[1]Wellcome Trust Centre for Neuroimaging, University College London, London, UK
[2]Virginia Tech Carilion School of Medicine and Research Institute, Virginia Tech, Roanoke, VA, USA
[3]Department of Neuroscience, University of Minnesota, 6-145 Jackson Hall, 321 Church Street SE, Minneapolis, MN 55455, USA

*Keywords*: cognitive resources, delay discounting, episodic future thinking, working memory

## Abstract

Although delay discounting, the attenuation of the value of future rewards, is a robust finding, the mechanism of discounting is not known. We propose a potential mechanism for delay discounting such that discounting emerges from a search process that is trying to determine what rewards will be available in the future. In this theory, the delay dependence of the discounting of future expected rewards arises from three assumptions. First, that the evaluation of outcomes involves a search process. Second, that the value is assigned to an outcome proportionally to how easy it is to find. Third, that outcomes that are less delayed are typically easier for the search process to find. By relaxing this third assumption (e.g. by assuming that episodically-cued outcomes are easier to find), our model suggests that it is possible to dissociate discounting from delay. Our theory thereby explains the empirical result that discounting is slower to episodically-imagined outcomes, because these outcomes are easier for the search process to find. Additionally, the theory explains why improving cognitive resources such as working memory slows discounting, by improving searches and thereby making rewards easier to find. The three assumptions outlined here are likely to be instantiated during deliberative decision-making, but are unlikely in habitual decision-making. We model two simple implementations of this theory and show that they unify empirical results about the role of cognitive function in delay discounting, and make new neural, behavioral, and pharmacological predictions.

## Introduction

Both human and non-human animals discount future rewards, preferring smaller rewards delivered sooner over larger rewards that will only be available after a delay (Madden & Bickel, 2010). In both human and non-human animals, the ability to wait for a larger reward is positively related to self-control abilities (Mischel & Underwood, 1974; Baumeister *et al.*, 1994; Peters & Büchel, 2011) and inversely related to addiction liabilities (Ainslie, 2001; Bickel & Marsch, 2001; Odum *et al.*, 2002; Perry *et al.*, 2005; Anker *et al.*, 2009; Heyman, 2009; Stanger *et al.*, 2011). Most theories suggest that the temporal discounting of future rewards is related to issues of uncertainty, risk, and investment (Samuelson, 1937; Sozou, 1998; Redish & Kurth-Nelson, 2010); however, the mechanisms underlying the temporal discounting of future rewards remain unresolved.

In non-human animals, discounting functions are usually measured through direct choices of actual rewards, often through the use of adjusting delay tasks (Mazur, 1997), in which selecting smaller-sooner or larger-later rewards changes the delay to the larger-later reward. Modeling adjusting delay tasks is possible using reinforcement-learning models that adjust decisions based on feedback (Kurth-Nelson & Redish, 2009). Most human experiments, however, have measured discounting rates through questionnaires and single-trial events, in which discounting decisions are made in novel situations, often without any immediate feedback (Madden & Bickel, 2010). Modeling decision-making in novel conditions requires discounting functions that can arise from imagined futures.

Discounting experiments generally find reasonable fits to hyperbolic discounting functions, in which the value of the future reward is discounted by $1/(1 + delay)$ (Madden & Bickel, 2010). Although this fit is often very good, other functions have been proposed, including exponential (Schweighofer *et al.*, 2006), exponential with a bonus for immediate rewards (Laibson, 1997; McClure *et al.*, 2004), and the sum of multiple exponentials (Sozou, 1998; Kurth-Nelson & Redish, 2009). These models all assume that the value of a delayed reward is a simple function of the delay to that reward, following simple delay-dependent assumptions of risk, uncertainty, and investment in economic (Samuelson, 1937; Sozou, 1998) and reinforcement-learning (Sutton & Barto, 1998; Redish & Kurth-Nelson, 2010) models.

Discounting functions, however, are also modulated by cognitive and representational factors. For example, subjects with increased cognitive resources tend to show slower discounting rates. As a trait, higher cognitive skills are correlated with better self-control and slower discounting rates (Mischel & Grusec, 1967; Mischel & Underwood, 1974; Burks *et al.*, 2009). As a state, cognitive resources

*Correspondence*: A. D. Redish, as above.
E-mail: redish@umn.edu

*Received 18 October 2011, accepted 2 February 2012*

can be modulated by training working memory (presumably increasing cognitive resources), which slows discounting (Bickel *et al.*, 2011), or by imposing a cognitive load (presumably decreasing cognitive resources), which speeds discounting (Vohs & Faber, 2007). Representationally, discounting rates depend on the episodic nature of the potential options in a questionnaire – if the delayed option is primed with an episodic cue, then subjects are more likely to select it, leading to a decrease in the rate of decay of the discounting function (Peters & Büchel, 2010; Benoit *et al.*, 2011). Similarly, subjects prefer options presented with round numbers ($7 instead of $7.03), showing slower discounting functions when the delayed option is presented in round numbers ($7.03 now vs. $20 next week) and faster discounting functions when the immediate option is presented in round numbers ($7 now vs. $20.03 next week) (Ballard *et al.*, 2009). Ebert & Prelec (2007) report that the time sensitivity of discounting functions is susceptible to manipulations of attention, and that these manipulations differ between near-future and far-future options. These results suggest that there are other factors besides time involved in the delay discounting phenomenon; standard economic and reinforcement-learning models do not capture these modulatory effects (Daw, 2003; Glimcher, 2008).

Current theories of decision-making systems suggest that there are multiple systems that can drive decision-making (Daw *et al.*, 2005; Redish *et al.*, 2008; van der Meer *et al.*, 2012; Montague *et al.*, 2012), including both deliberative and non-deliberative (habit) systems. Deliberative decision-making in particular has been proposed to depend on search processes that proceed through potential futures (Johnson & Redish, 2007; van der Meer *et al.*, 2012) and/or the creation of imagined expectations (episodic future thinking) (Atance & O'Neill, 2001; Schacter *et al.*, 2007).

Here, we propose that discounting arises from a cognitive search process that is trying to identify rewarding situations in the future. We suggest that temporal discounting emerges from the correlation between delay and the ease of identifying future rewarding situations. As cognitive processes underlie the computation of value, the modulation of discounting by representation and cognitive resources emerges directly from the search process.

TABLE 1. Parameters used in model 1 (abstract search)

| | |
|---|---|
| Search time | 5000 |
| Number of searches | 1000 |
| Reward radius | 0.5 |
| Default basin depth | 1 |
| Inverse friction $F$ | 0.9 |
| Thermal noise $\vec{\mathbf{n}}$ | Uniform distribution over the range (−0.5, 0.5) in each dimension |
| Inertia $m$ | 0.1 |

TABLE 2. Parameters used in model 2 (Hopfield)

| | |
|---|---|
| Size of the set of patterns | 15 |
| Search time limit | 250 |
| Number of units | 80 |
| $N_T$, number of trials | 1000 |
| $N$, noise in weights | Uniform distribution over (−0.05, 0.05) |
| $e_p$, strength of pattern when not directly modeled $p$ | 1 |

## Theory

We present a theory of discounting in deliberative decision-making. The key to deliberative decision-making (sometimes referred to as 'model-based reinforcement learning') (Kaelbling *et al.*, 1996; Sutton & Barto, 1998) is the ability to evaluate potential future outcomes on the fly, permitting flexible behavior in the face of changing contingencies or changing motivational states (Doya, 1999; Suri, 2002; Daw *et al.*, 2005; Niv *et al.*, 2006; Balleine & Ostlund, 2007; Johnson *et al.*, 2007; Hill, 2008; van der Meer *et al.*, 2012). Several authors have proposed that this evaluation process involves the episodic projection of oneself into the future to vicariously sample potential outcomes and thereby establish a subjective value (Buckner & Carroll, 2007; Gilbert & Wilson, 2007; Johnson & Redish, 2007; Schacter *et al.*, 2007; van der Meer & Redish, 2009; van der Meer *et al.*, 2012).

Even in situations in which one is told the reward that one will receive (e.g. if you are offered $100 next week), one needs to determine the world situation next week in order to evaluate the usefulness (the subjective value) of that $100 next week. The subjective value of that $100 next week is very different if one expects to win the lottery tomorrow. Thus, from a neural decision-making systems point of view, this situation still has to be evaluated in order to establish a subjective value for it. We suggest that while holding on to that semantic association of $100 next week (which modulates both the search process and the situation/reward association mapping), the agent projects an episodic self-representation forward in time. The ability to construct a vivid, associated forward projection will determine how easily one can evaluate the subjective value of that $100.

The exact mechanisms of this episodic projection in the brain are likely to be very complex. There is likely to be a combination of explicit search through a graph of states using causal models at mixed levels of abstraction (Newell & Simon, 1972; Rich & Knight, 1991; Nilsson, 1998; Smith *et al.*, 2006; Botvinick & An, 2009); attractor networks settling into representations (Hertz *et al.*, 1991) or following energetically-favorable pathways encoded by experienced sequences (Johnson & Redish, 2007); and application of correlational and associational knowledge. A complete theory of these processes is beyond the scope of this article. However, we propose that episodic construction has some essential properties that give rise to temporal discounting and allow us to make qualitative neural, behavioral, and pharmacological predictions about discounting. We therefore first present the essential properties and then present two instantiations of this theory (one based on abstract search processes and one based on settling processes of attractor networks). We show the same qualitative results in each instantiation: temporal discounting of delayed rewards, and modulation of discounting by the representation of outcomes and by cognitive resources.

### *Assumption 1*

Construction can be viewed as a search process (Winstanley *et al.*, 2012). Neurophysiologically, we suggest that search occurs through the activation space of the neural network that represents episodes. It follows a trajectory from the present toward a representation of a future outcome (Tolman, 1939; Buckner & Carroll, 2007; Johnson & Redish, 2007). As noted above, this is likely to be very complex, but ultimately, the representation starts from the present, and must reach a future outcome in order to value that outcome. For example, the search may involve multiple hierarchical levels, perhaps through cognitive chunks (Newell & Simon, 1972; Botvinick & An, 2009), but each step probably occurs through the settling of dynamic attractor models, such

as content-addressable memories, which in itself is mathematically akin to a search process (Hertz *et al.*, 1991). Of course, the topology of the neural activation space may look quite different from the topology of the representation space, but we assume that there is some degree of homology. In general, if something is farther away, there are more steps to go through, even though these steps may be in a variety of different spaces.

### *Assumption 2*

The search process values rewards proportionally to how easy they are to find. The search process attempts to find rewards in the future, but in most real-world situations, and even in many experimental tasks, it is not feasible to fully search out every possible outcome. Naturally, the search process can only value rewards if it can find them. An all-or-nothing effect may be partially smoothed out by multiple searches (whether discrete or continuous, i.e. evolving a distribution through the search space). These multiple searches may occur serially (Johnson & Redish, 2007) or in parallel (Botvinick & An, 2009). From the point of view of making adaptive decisions, valuing rewards proportionally to how easy they are to find may reflect that easy-to-search-to rewards are also statistically easier to obtain in the real world. This is expected to be a consequence of the fact that the learning mechanisms that create the representation space through which the search operates tend to associate things that are linked in the real world.

Neurophysiologically, it is likely that evaluation occurs through a downstream brain network that has learned to associate which kinds of situations are valuable and which are not, probably involving the orbitofrontal cortex (Padoa-Schioppa & Assad, 2006; Zald & Rauch, 2006; Schoenbaum *et al.*, 2009) and ventral striatum (Pennartz *et al.*, 2009; Roesch *et al.*, 2009; McDannald *et al.*, 2011; van der Meer & Redish, 2011). If we assume that the classification process is locally smooth, then rewards will be more valued if they are easier for the episodic projection to find, because having a representation closer to something that has been learned to be rewarded will be more subjectively valuable.

### *Assumption 3*

Multiple factors influence how easy a reward is to find, including the temporal distance to the reward and the ease of constructing the future representation. As events are generally continuous through time, there is an inherent overlap between 'now' and any given future, which decays with a regular time-course (Rachlin, 2004; Howard *et al.*, 2005). Previous researchers have suggested that this creates a separation between current and future 'selves' such that future selves are farther away than one's current 'self' (Ainslie, 1992, 2001; Trope & Liberman, 2003; Rachlin, 2004). Thus, we assume that the ease of construction of a future outcome is correlated with the delay to the outcome. However, because the fundamental cause of temporal discounting in our theory is that more distant rewards are more difficult to find, it is possible to modulate discounting by creating situations where these are dissociated. One interesting example is when outcomes are made easier to find by altering the energy landscape around them. We will show, for example, that in an attractor-network instantiation of this theory (model 2, below), increased interconnection strength of a memory (producing a deeper basin of attraction) allowed the network to find that memory faster and more reliably, leading to slower discounting. We argue that increased interconnection strength is a feature of vivid representations. Therefore, boosting the vividness of the outcome leads to slower discounting, as observed experimentally (Peters & Büchel, 2010; Benoit *et al.*, 2011).

Neurophysiologically, assumption 3 is met because it is easier for the projection to reach temporally-nearer outcomes because they also tend to be nearer in feature space. The distance in 'search space' is the ease of constructing that future (Schacter & Addis, 2007; Schacter *et al.*, 2008). Situations that are more separated in time tend to be more separated in feature space because everything in the world is changing, and by the nature of learning in associative networks, situations that are more separated in feature space are likely to be more separated in activation space. Of course, the topology of the activation space is determined by the dynamics of the network, and things may be topologically closer or further than their distance in raw activation space. We assume that there is a finite time available for searching, so the search cannot simply run forever until it finds all outcomes. Experimentally, discounting is often measured with questionnaires, where the experimenter asks the subject's preference between, e.g. \$100 today or \$1000 in 1 year. In this case, even though the subject 'knows where the reward is' (e.g. \$1000 in 1 year), his or her episodic representation must still follow a trajectory from the present to that outcome in order to perceive it as valuable. We suggest that the broader space of possibilities ('where will I be in a year?') leads to a decrease in the expected value of that delayed outcome. In summary, this theory constitutes a novel explanation for the mechanism of delay discounting – rewards that are temporally near will be more likely to be found and will be valued more highly. We now present two models that instantiate, respectively, the generic assumptions and their mapping on to episodic construction.

## Materials and methods

### *Model 1: Abstract search*

This model simulates the process of estimating the value of a situation by identifying the future rewards that will be available.

Searches in model 1 were performed in the real number plane, $R^2$. Searches always started from the origin. Certain locations within the representation space were defined as rewarded. For simplicity, we assumed that rewards are either present or not (i.e. they were not probabilistically delivered). Multiple searches were performed, and the value produced by the model was defined as the number of searches that found a reward, divided by the total number of searches. Delay to a reward was modeled as the Euclidean distance (following assumption 3); thus, a reward available after five time-steps would be located five units from the origin.

Search dynamics proceeded by being updated through standard particle motion equations, stochastically following an imposed energy gradient

$$\vec{\mathbf{x}}(t+1) = \vec{\mathbf{x}}(t) + \Delta t \cdot \vec{\mathbf{V}}(t) \quad (1)$$

$$\vec{\mathbf{V}}(t+1) = F \cdot \left(\vec{\mathbf{V}}(t) + m \cdot (\nabla S + \vec{\mathbf{n}})\right) \quad (2)$$

where $\vec{\mathbf{x}}(t)$ was the location of the particle at time $t$, and $\vec{\mathbf{V}}(t)$ was the velocity at time $t$. The inverse friction (slipperiness) $F$ was set to 0.9 for all simulations. The thermal noise $\vec{\mathbf{n}}$ was isotropic; it was independently drawn at each time-step from a uniform distribution over the range (−0.5, 0.5) in each dimension. The inertia $m$ was held at 0.1 for all simulations.

The energy gradient was defined by the surface

$$S = \sum_i B_i \tag{3}$$

where each $B_i$ is one basin. Basins were defined by the following equation

$$B_i = \frac{D_i}{\sqrt{(x - x_i)^2 + (y - y_i)^2 + 4)}} \tag{4}$$

where $D_i$ defined the depth of basin $i$, and $(x_i, y_i)$ defined the center of the basin. Qualitatively similar results were found with a number of different basin shapes. These parameters were chosen as being robust in order to allow examination of the effects of other parameters on the system. $D_i$ was set to 1 unless it was being explicitly manipulated. In Fig. 1, the number of basins was zero, so the energy landscape was flat. In Figs 2 and 3, the number of basins was one and this basin was centered on the reward. In Fig. 3, there was a basin centered on the reward and also other basins centered at uniformly random locations within (−50, 50) in each dimension.

If, at any time-step, the search came within 0.5 Euclidean distance of a reward, the search terminated and returned with a positive value saying that the reward was found. If the search reached its maximum duration (held to 5000 time-steps for all simulations in which it was not explicitly manipulated) without finding reward, the search terminated and returned with a value of 0. On each trial, 1000 independent searches were performed. The subjective value was taken to be the proportion of searches that found a given reward. Parameters used for model 1 are listed in Table 1.

### *Model 2: Hopfield network*

This simulation used a standard Hopfield network (Hopfield, 1982; Hertz *et al.*, 1991). Each unit $i$ had a state $s_i$ of 1 or −1. A set $P$ of randomly defined patterns was encoded in the weights of the network. They were defined as

$$w_{ij} = N + \sum_{p \in P} p_i p_j \tag{5}$$

where $p_i$ is the $i$th bit of pattern $p$, and $N$ is a noise contribution, drawn from a uniform random distribution centered about zero. The weights were symmetric, so $w_{ij} = w_{ji}$. New random patterns were used on each trial to reduce the systematic bias that would arise from any particular configuration of patterns (in other words, the weights were regenerated from scratch on each trial).

One of the encoded patterns was defined as the target, to simulate where reward was located. The other patterns were non-targets.

On each trial, the activation state of the network was initialized to a pattern that was generated by starting from the target pattern and flipping $H$ bits. The delay on that trial was defined as $H$, the Hamming distance between the starting position and the target. The network was then updated asynchronously until either (i) the search time limit was reached, or (ii) the target pattern was reached. Exactly one unit was randomly selected for update on each time-step. The new state of this unit $i$ was

$$s_i \leftarrow \begin{cases} -1 & \text{if } \sum_j s_j w_{ji} < 0 \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

where $j$ indexes the set of units. Thus, on each time-step one unit was selected for update, but that unit may or may not have changed its state on that time-step, depending on whether it already had the same sign as the weighted sum of its inputs.

$N_T$ trials were run at each delay of each condition, and the fraction of these trials in which the network reached the target pattern was reported as the subjective value.

The depth of the basin for a particular pattern was manipulated by multiplying the contribution of that pattern to the weights by a constant.

In experiments where the pattern strength of the target pattern was varied, the weights were calculated as

$$w_{ij} = N + \sum_{p \in P} e_p p_i p_j \tag{7}$$

where $e_p$ is the strength of pattern $p$. $e_p$ was always set to 1 for all non-target patterns, but could be > 1 for target patterns. Parameters used for model 2 are listed in Table 2.

## Results

### *Model 1: Abstract search*

This simple model instantiates the three assumptions of our theory in the simplest way: there is an abstract search process, which follows a

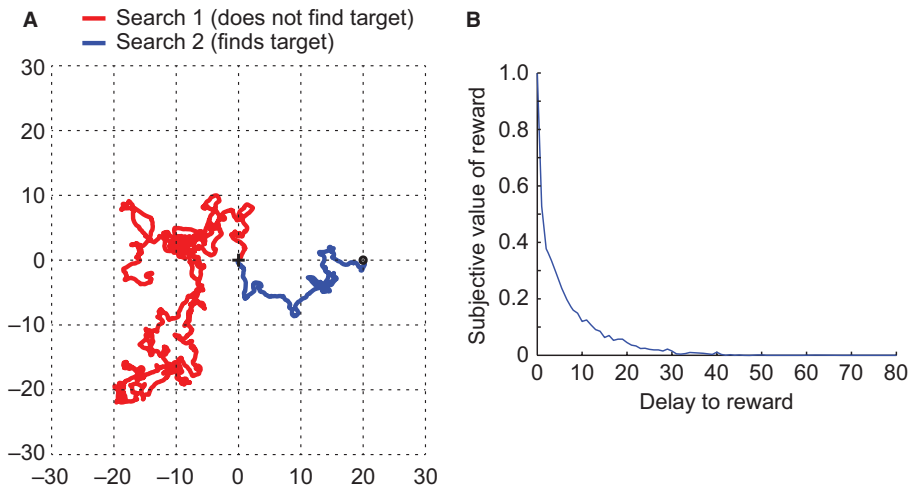


FIG. 1. Discounting arises from a set of simple assumptions about search. (A) Examples of two random searches on the plane. Black plus sign at origin indicates start of searches. Black circle indicates location of reward. (B) Subjective value (defined as the fraction of searches that find the target) is smaller when the delay (defined as distance between start of search and target reward) is larger.

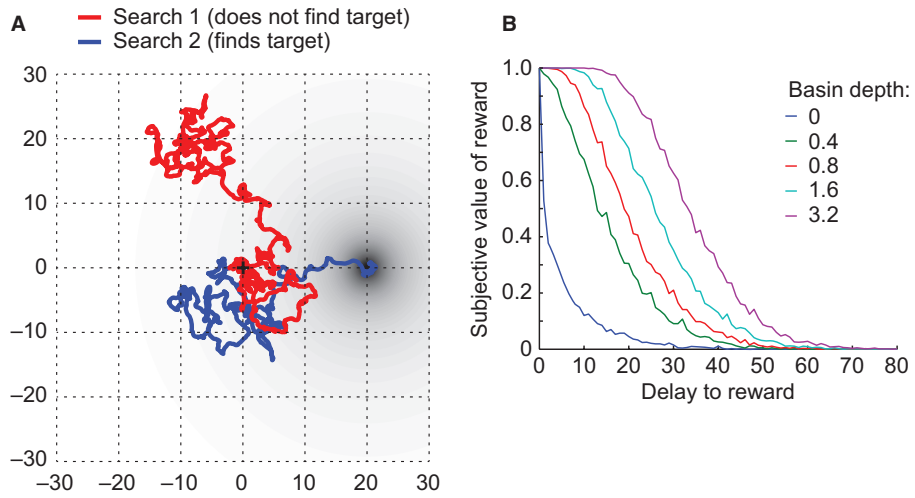


FIG. 2. Discounting is slower when the representation of the outcome is more energetically favorable. (A) Examples of two searches through an energy landscape (indicated by gray shading) with one basin, which is centered at (20, 0). The reward is located at the center of the basin. (B) The attenuation of subjective value with increasing delay (i.e. discounting) is less pronounced when the basin is deeper. Note that the dark blue trace is the same as Fig. 1B.

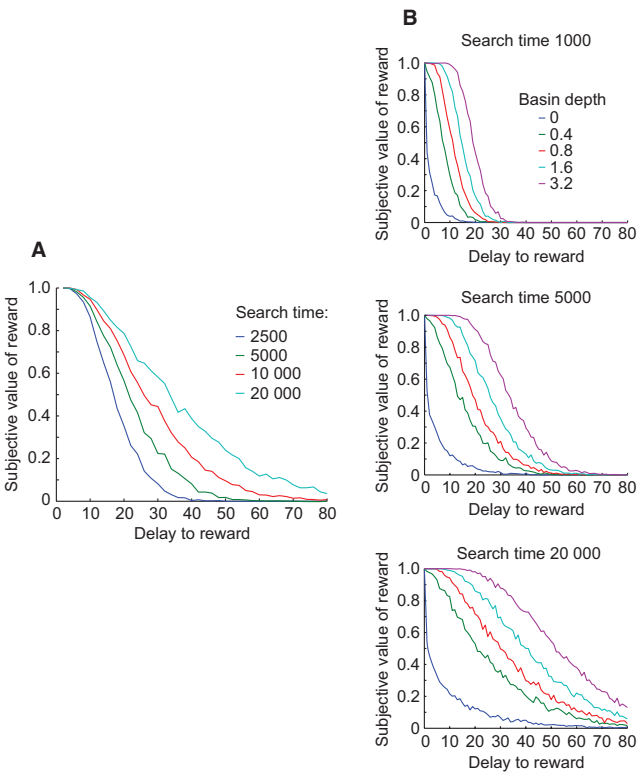


FIG. 3. Discounting is slower when more time is available to search, and this effect is greater when the outcome is more energetically favorable. (A) As maximum search time increases, the discounting function becomes shallower, meaning that the same delay produces less reduction in subjective value. (B) In each panel, discounting curves are shown for a range of different basin depths. As search time is increased from the top panel to the bottom panel, the separation between the curves increases, meaning the effects of search time and basin depth on the curves are supra-additive. The middle panel is the same as Fig. 2B.

trajectory through an abstract space. This space could correspond to anything from neural activation space to sensory feature space to a high-level hierarchical rule space. For this model, the search space is the real plane with a Euclidean distance metric.

In Fig. 1, we used random diffusion as the search dynamics. This is the simplest possible search dynamics, and simply illustrates that the phenomenon of discounting arises from our three theoretical assumptions. Many other forms of search dynamics would produce the same results.

In Figs 2 and 3, we added the simplest possible extra assumption – that the search process has some non-random dynamics that favor particular subsets of the space. Again, we implemented this with the simplest possible instantiation – a gravity-like attraction to radially symmetric energy basins. This gives an example of how the assumption of adding non-random dynamics can produce a change in discounting (slower discounting to outcomes in regions of the representation space that are favored by the dynamics).

Starting from the current location [without loss of generality defined to be (0, 0)], we allowed a large number of serial searches to evolve along random trajectories. Examples of these trajectories are shown in Fig. 1A. Delay was modeled as the Euclidean distance. As a consequence, more delayed outcomes were discounted relative to less delayed outcomes (Fig. 1B).

### *Effect of changing basin depth*

In order to encode memories within our representation space, we defined an energy function over the space, distorting the two-dimensional space with basins. The random diffusion of the search was modulated by a gravity-like effect based on the shape of the energy function (Hertz *et al.*, 1991) (Fig. 2A). Deeper and broader basins were more likely to be found, meaning that changes in the basin shape will have profound effects on the discounting function. Searches probabilistically gravitate towards basins, meaning that searches are more likely to find targets at the bottom of deeper basins before the search's maximum search time expires. This mitigates the effect that distance (which stands in for delay) has in making it more difficult to find the target, thus effectively slowing the apparent discounting rate (Fig. 2B).

It is interesting to note that, under some conditions, the discounting functions produced by this model have an initial 'plateau', e.g. in the purple curve in Fig. 2B; the value remains near 1 until the delay reaches about 20. It is a fundamental prediction of the search + basins theory that, when the search starts within some radius of an

energetically favorable state, it is almost guaranteed to collapse to that state, minimizing the difference between small delays within that radius. Depending on the shape of the energy space, this effect could range from large to unnoticeable.

*Effect of changing maximum search time*

Finding a reward, particularly a distant reward, takes time. Presumably, a search can only proceed for a limited time before the agent will give up on the search. Therefore, to model this limitation, the available time for a given particle to find a reward was limited. If the particle did not find a reward within that limited number of steps, it returned a value of 0 or 'no reward found'. Increasing the search time available to the particle increased the likelihood that the particle will find a distant reward, effectively slowing the apparent discounting rate. Figure 3A shows how changing the search time changes the discounting function.

When we systematically varied both the maximum search time and the basin depth, we found a non-linear interaction between these two parameters (Fig. 3B), such that the extent to which a deeper basin slowed the discounting function became much larger when there was a greater maximum search time. This is a novel prediction – that the effect of episodicity, as in Peters & Büchel (2010), should become more pronounced in subjects with more cognitive resources (e.g. following working memory training). It arose in the model because, if the basin gradient was much less than the thermal noise, extra search time had relatively little effect (the expected time to reach a given point was exponentially increasing with its distance). However, if the basin gradient was non-negligible relative to the thermal noise, then extra search time allowed the search to follow the gradient into the basin.

*Effect of increasing the number of distractor basins*

In Fig. 3, there was only one basin, and it was centered on the rewarding target. We also tried adding other, non-target basins in the energy landscape, such that they contributed to the overall gradient, but reaching their center did not impart a reward. This simulates the effect of other coherent situations encoded by the representational system. We found that increasing the number of basins caused the model to discount faster (Fig. 4A), because these non-target basins could draw the search away from the rewarding target. In these simulations, the depth of each basin was 1.

We also found that, as search time increased, the effect of changing the number of basins became even more pronounced (compare Fig. 4A and B). This is because, with many non-target basins, it becomes very likely that the search process will fall into one of these incidental basins between the starting point and the target, so that additional search time would be wasted on sitting at the bottom of that incidental basin.

This makes the novel prediction that increasing cognitive resources should have little effect on discount rates when there are many other targets encoded in the search space.

## Model 2: Hopfield network

One interesting instantiation of our search assumptions is a settling attractor network representing an episodic future imagination process. In this section we show that a settling attractor network model exhibits the same discounting behavior as we showed for an abstract search process in the previous section. The model here consisted of a standard Hopfield network, with $N_P$ patterns encoded in the weights (Fig. 5A). One of the patterns was defined as the rewarding target.

*Hopfield network discounts delayed rewards*

We modeled the process of evaluating the subjective value of a future outcome. The system starts from a present state and projects forward to imagine a future state. Delay to the future outcome was modeled as the Hamming distance between the starting state and the outcome state. To generate a starting state for the units that was Hamming distance $H$ from the rewarding target, we flipped a random $H$ bits of the pattern representing the rewarding target.

After the units were initialized to a starting state, the network was allowed to run with asynchronous updates until either the network's state matched the rewarding target, or the maximum number of asynchronous update steps (search time limit) was reached.

As we varied the delay, the subjective value calculated by the model decreased roughly exponentially (Fig. 5B).

*Effect of changing basin depth*

We were also interested in the effect of varying episodicity, or the strength or vividness of a future representation. We assume that the more vivid an outcome is, the stronger are the connections that encode it.

In a standard Hopfield network, the connection weight between two units is the sum (across encoded patterns) of the product of the states of those units in that pattern. In our simulations, we permitted a

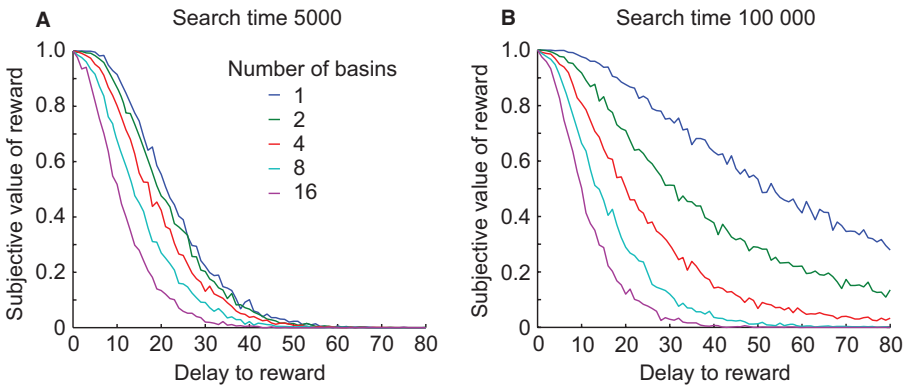


FIG. 4. Discounting is slower when there are fewer incidental minima in the energy space, and this effect is greater with longer search time. (A) In Figs 2 and 3, there was only one basin in the energy landscape, centered at the rewarding target. Now we add additional non-target basins. The depth of each basin, including the target basin, is 1. Increasing the number of basins increases the amount of discounting. (B) Increasing the maximum search time greatly enhances the effect of changing the number of basins, i.e. the separation between curves is greater in B than in A.

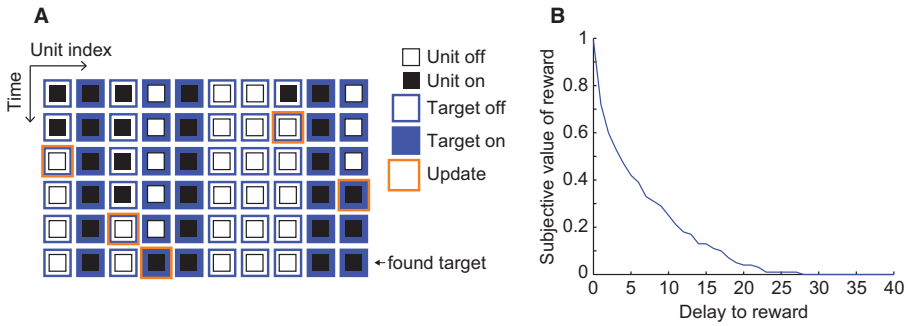


FIG. 5. Discounting arises in the settling of an attractor network. (A) The units in the Hopfield network simulation are updated asynchronously until the target pattern is found or the maximum search time is reached. (B) Subjective value (defined as the fraction of trials where the network settles on the rewarding target) is smaller when the delay (defined as Hamming distance between the starting state of the network and the rewarding target) is larger.

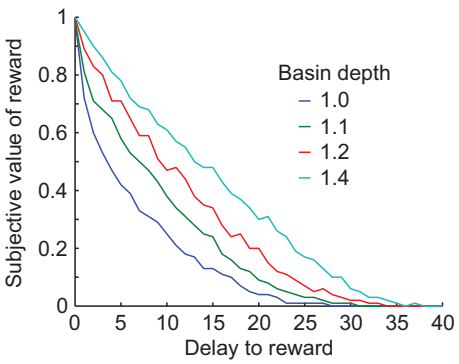


FIG. 6. Discounting is slower when the target pattern is more strongly encoded. As the weight contributions of the rewarding target pattern are increased relative to the other encoded patterns, the discounting function becomes slower.

particular pattern to have stronger connections, by multiplying that pattern's contribution to the weights by a constant > 1.

When we boosted the vividness of the rewarding target (leaving all other encoded patterns the same), this had the effect of increasing the likelihood of the network settling on to the rewarding target, even when the starting state of the network had a large Hamming distance from the rewarding target. Therefore, boosting vividness slowed the discount rate of the model (Fig. 6).

*Effect of changing maximum search time*

To simulate the effect of changing the amount of cognitive resources available for construction, we varied the total number of time-steps that the model was permitted to try to settle on the rewarding target. Increasing the available search time slowed the discounting rate of the model (Fig. 7A), because with more time-steps it was less likely that the search would time out before settling on the rewarding target.

As in the abstract search model, the effect of changing search time interacted with the effect of changing basin depth. The gap between the discounting curves at 1.0 basin depth and 1.4 basin depth grew larger as more search time was available (Fig. 7B). This supra-additive effect occurred because, with a deeper basin, it became more likely that search time was the limiting factor in valuation. In other words, with a deeper basin, it was likely that even searches starting at a large Hamming distance from the rewarded target would eventually find the rewarded target; a larger search time allowed them to do so.

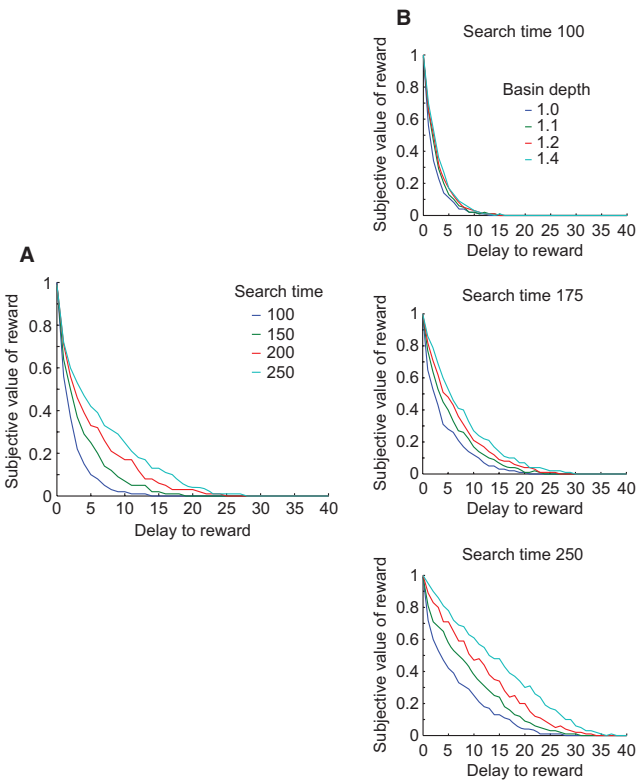


FIG. 7. Discounting is slower when more time is available for the network to settle, and this effect is greater when the target pattern is more strongly encoded. (A) As maximum settling time increases, the discounting function becomes shallower, meaning that the same delay produces less reduction in subjective value. (B) In each panel, discounting curves are shown for a range of different basin depths. As search time is increased from the top panel to the bottom panel, the separation between the curves increases, meaning the effects of search time and basin depth on the curves are supra-additive. The bottom panel is the same as Fig. 2B.

*Effect of increasing the number of distractor basins*

Auto-associative memories may encode many or few stored representations. Encoding more representations in the same network generally produces a more complex energy landscape with more local minima. We were interested in how the other patterns encoded in the network would interact with the search process to the rewarding target, so we varied the number of patterns $N_P$ encoded in the Hopfield network's weights.

Increasing the number of encoded patterns caused discounting to become faster (Fig. 8A). As the number of encoded patterns increased, it became more likely that the network would either settle to one of the non-target patterns, or get trapped in a local minimum that did not correspond to any encoded pattern. It may have also made the paths through the space more tortuous, decreasing the likelihood of finding the rewarding target before the search time limit was reached.

The effect of the number of encoded patterns also interacted with the effect of changing the maximum search time (Fig. 8, compare A and B). With 20 encoded patterns, the discount curves for 250 search time and 500 search time were nearly identical. However, with a smaller number of encoded patterns, the discount curve for 500 search time was much slower than the discount curve for 250 search time. This is because, as more non-target basins crowded the space, it became more likely that movement in any direction away from the starting point would quickly fall into a non-target basin, negating the benefit of any further searching.

## Correspondence to specific experimental results

### *Changes in the episodic cueing of the potential options*

Peters & Büchel (2010) and Benoit *et al.* (2011) examined the effect of manipulating episodic cues to the delayed choice in a questionnaire task, comparing the effect of asking for a preference for larger-later (delayed) over smaller-sooner (immediate) choices when the delayed option was marked by a simple time (e.g. '26€ after 35 days') or by an explicit (and correct) future cue (e.g. '35€ after 45 days during your future vacation in Paris'). Delayed options in the episodic cases were more likely to be preferred, implying a decrease in the discounting rate (Peters & Büchel, 2010; Benoit *et al.*, 2011). We modeled this effect as an increase in the depth of the basin, under the assumption that more easily retrieved episodic memories would be at deeper spots in an attractor network (Kohonen, 1980; Hertz *et al.*, 1991). As can be seen in Figs 4 and 8, increasing the depth of a basin has the effect of increasing the likelihood that a searching particle would find it, and thus decreasing the discounting rate to the episodically-marked future event. In a sense, this increase in depth can be interpreted as a change in the temporal attention being paid to the future event (Ebert & Prelec, 2007). Recently, Radu *et al.* (2011) found that changes in how classical discounting questionnaires are framed produced changes in discounting rates, and, by studying discounting related to both future and past events, concluded that the changes provided additional attention to the temporally-distant event. The changes in basin depth shown in Figs 2 and 6 model this effect.

### *Changes in cognitive abilities and cognitive load*

Several researchers have noted that subjects with higher cognitive abilities show slower discounting functions than subjects with lower or impaired cognitive abilities. For example, Burks *et al.* (2009) correlated experimentally determined discounting rates through questionnaires with intelligence as measured by the Raven's matrices non-verbal IQ test, the Hit-15 ability to plan task, and a quantitative literacy test. They found that increased cognitive intelligence as measured by these parameters (all of which were correlated with each other) correlated with both increased consistency between delays (later delays were more likely to be discounted more than earlier delays) and slowed discounting rates. Franco-Watkins *et al.* (2006) suggest that cognitive load leads to decreased consistency, whereas Baumeister *et al.* (1994), Hinson *et al.* (2003), Gailliot *et al.* (2007), and Vohs & Faber (2007) suggest that cognitive load leads to impulsivity.

We can interpret the state of increased cognitive load and the trait of decreased cognitive skills as reductions in the available resources that can be applied to a search. We simulated this with reduced search times. As can be seen in Fig. 3, decreasing the maximum time before a search gives up increases the discounting rate (speeds discounting), with an especially pronounced effect when the basin around the reward is deeper. Of course, manipulating the available search time is just a very crude proxy for a host of mechanisms that are probably part of the whole picture of cognitive resources. For example, the ability to add more associations might make searches easier by reducing the topological distance between points in the representation space. The general argument is that, by improving the search/construction process, it is more likely that the deliberative decision-making process will be able to find, and therefore value, delayed rewards.

### *Changes in working memory abilities*

Recently, Bickel *et al.* (2011) examined the effects of training working memory on discounting rates and found that working memory training produced dramatic decreases (slowing) in discounting rates. In contrast, a control cohort that received the same training, but with the answers given, thus not requiring working memory (and producing no working memory improvements), showed no significant changes in discounting rates. Presumably, some aspects of working

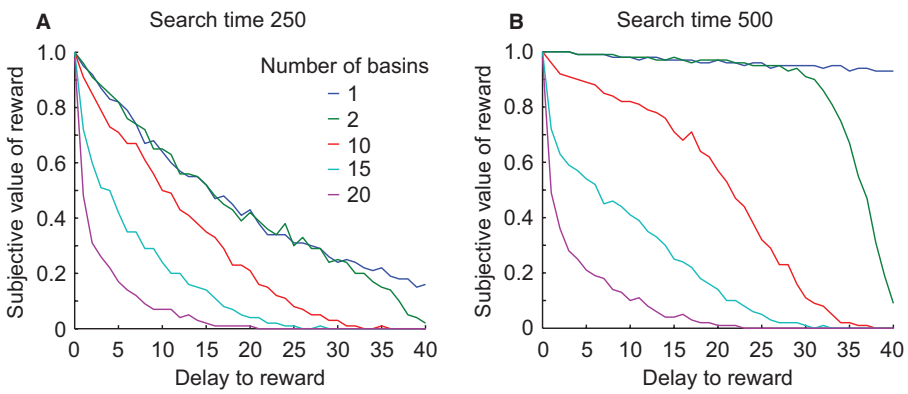


FIG. 8. Discounting is slower when there are fewer patterns trained in the network, and this effect is greater when more time is available for the network to settle. (A) Increasing the number of patterns encoded in the network's weights increases the amount of discounting. (B) Increasing the maximum search time enhances the effect of changing the number of basins, i.e. the separation between curves is greater in B than in A.

memory are required to hold on to a search in progress (Baddeley, 1986; Kliegel *et al.*, 2008), and thus we can hypothesize that the maximum search time is related to working memory. As with changes in cognitive abilities and cognitive load, this discounting-as-search-to-reward model suggests an explanation for the decreased discounting rates through modulations in the parameters, most likely to be the length of time that a search can proceed before it is abandoned. As can be seen in Fig. 2, providing increased potential search time leads to a slowing of discounting functions.

## Relationship to neurophysiology

The discounting mechanism proposed here depends on the depth of the basin of attraction of representations during the imagination of future representations (i.e. during episodic future thinking). Episodic future thinking is known to depend on the hippocampus and prefrontal cortex (Addis *et al.*, 2007; Hassabis *et al.*, 2007; Schacter *et al.*, 2007; Buckner, 2010), brain structures that are often modeled as auto-associators with attractor dynamics (Wilson & Cowan, 1972; Zhang, 1996; Redish, 1999; Durstewitz *et al.*, 2000). Our theory suggests that the dynamics of these brain structures involved in episodic future thinking will translate directly into discounting rates during deliberative decision-making, in that more easily imagined futures will be discounted less than futures that are more difficult to imagine. From this correspondence, we can make several predictions.

### *Prediction 1*

Episodic discounting rates will depend on the ability to access prefrontal representations. Extensive research in self-control has suggested that cognitive resources, in particular self-control-related resources, depend on limited resources that are probably related to glucose availability in prefrontal cortical structures (Gailliot *et al.*, 2007; Vohs & Faber, 2007). Our theory that discounting in deliberative decision-making depends on the ability to imagine future outcomes suggests that discounting rates will depend on this self-control-related resource and can be manipulated through the presence or diminishment of these prefrontal glucose resources.

### *Prediction 2*

Subjects who are unable to episodically imagine futures may still show temporal discounting, but that temporal discounting will not be modulated by cognitive resources or the representational nature of the options. Discounting probably arises from a number of processes, likely to be different for different decision-making systems. Although subjects with damage to the prefrontal cortex or hippocampus may still show discounting due to other processes (such as simple associations), the evidence that subjects with hippocampal or prefrontal damage cannot accomplish episodic future thinking (Addis *et al.*, 2007; Hassabis *et al.*, 2007) implies that any discounting functions that they do show will be immune to cognitive resource and representational manipulations. A recent study by Kwan *et al.* (2011) found that an amnesic patient showed linear discounting rates, whereas controls showed hyperbolic discounting rates.

### *Prediction 3*

The depth of the basin affects discounting rates. Several models (Seamans & Yang, 2004; Yamashita & Tanaka, 2005; Redish *et al.*, 2007; Durstewitz *et al.*, 2010) have suggested that the basin depth in prefrontal and hippocampal representations depends on dopaminergic tone, with low tonic levels of dopamine implying shallow basins and fragile representations, and high tonic levels of dopamine implying deep basins and rigid representations. This suggests an inverted U in discounting rates as a function of dopaminergic tone, with overly low tonic dopamine leading to difficulty in finding rewards, speeding up discounting rates, and overly high tonic dopamine leading to difficulty in moving through space, also speeding up discounting rates. This may also produce discounting effects as a function of genetic variation in dopamine single-nucleotide polymorphisms (SNP) (Frank *et al.*, 2007, 2009).

## Discussion

We propose that the discounting of expected rewards with time is due to a combination of three theoretical ideas: (i) that model-based reinforcement learning entails a search through the future [planning (Johnson & Redish, 2007), episodic future thinking (Buckner & Carroll, 2007)], (ii) that temporally-delayed rewards are more difficult to find [because they are more contextually separated (Rachlin, 2004; Howard *et al.*, 2005) and thus farther away in the representation space], and (iii) that one calculates expected value as a function of the ability to find rewards [thus taking into account expected risk and hazard (Sozou, 1998; Redish & Kurth-Nelson, 2010)]. In this article, we show that this simple model can provide explanations for the effects of cognitive and executive functions on discounting rates, including the effects of cueing episodicity of the future reward (Peters & Büchel, 2010; Benoit *et al.*, 2011), inherent cognitive abilities (Franco-Watkins *et al.*, 2006; Burks *et al.*, 2009), compromising cognitive abilities, such as providing a cognitive load (Vohs & Faber, 2007) or depleting cognitive resources (Baumeister *et al.*, 1994; Hinson *et al.*, 2003; Gailliot *et al.*, 2007), and training cognitive abilities (Bickel *et al.*, 2011).

Although the extent to which deliberative (model-based) decision-making depends on vivid episodic construction of future events is not completely clear (Schacter *et al.*, 2007; Daw *et al.*, 2011; Simon & Daw, 2011; van der Meer *et al.*, 2012), we assume here that subjects briefly project themselves into the future to evaluate the subjective value of the outcomes. When a concrete representation of the outcome is not critical (perhaps because things are represented as abstract states) (Trope & Liberman, 2003), our theory reduces to a search through a Markov decision process (Smith *et al.*, 2006), and we would predict that, although cognitive skills are still relevant because they are needed to efficiently search through the graph (Huys *et al.*, 2012), the episodic modulation would have diminishing importance.

### *Search takes time*

One of the key points in any model of deliberative decision-making is that the deliberative process takes time (Gold & Shadlen, 2002; Johnson & Redish, 2007; Ratcliff & McKoon, 2008; Krajbich *et al.*, 2010; van der Meer & Redish, 2010). Forcing decisions to terminate early produces more inconsistent results (Gold & Shadlen, 2002; Ratcliff & McKoon, 2008), but can also lead to falling back on inherent biases and changing discounting functions (Simonson & Tversky, 1992; Lauwereyns, 2010). In the two models, limitations on search time are used as a proxy for available cognitive resources. This makes sense because increased cognitive resources should allow a larger exploration of the representation space. However, the implications of our search-to-find-reward model on forcing decisions to terminate early depend on whether the searches are performed in

parallel or serially. If the searches are performed serially (Johnson *et al.*, 2007), then early stopping would be modeled by decreasing the number of searches performed, which decreases consistency. If the searches are performed in parallel (Botvinick & An, 2008), then early stopping would be modeled by decreasing the maximum time available to each search, which increases impulsivity. However, one has to be careful with interpreting the model in this way because forcing decisions to terminate early may also drive an agent to access different decision-making systems. Decision-making in mammalian systems arises from multiple decision-making processes, which use different information-processing algorithms, including Pavlovian, deliberative (model-based) and habit (model-free) (Daw *et al.*, 2005; Huys, 2007; Redish *et al.*, 2008; van der Meer & Redish, 2010; van der Meer *et al.*, 2012; Montague *et al.*, 2012). Any experiment designed to test this prediction, however, will need to ensure that the manipulation is not pushing the agent from one system to the other (Bechara *et al.*, 1998, 2001). Different training regimens lead to the selection of different strategies (Restle, 1957) that depend on different brain structures (Barnes, 1979; Packard & McGaugh, 1996; Yin & Knowlton, 2004) and different computations (Daw *et al.*, 2005; van der Meer & Redish, 2010). One prediction of the theory presented in this article is that discounting will be more stable and less influenced by cognitive phenomena when an agent is primarily accessing non-deliberative decision-making systems (such as Pavlovian or model-free situation-action systems).

### *How much of delay discounting is really about delay?*

The search-to-find-reward theory proposed here suggests that what appears to be delay discounting is not an inherent consequence of delay itself, but rather of the ability of a search process to find reward (which may often be correlated with delay). This implies that value can be dissociated from delay by manipulating the ease of imagining the future outcome (Peters & Büchel, 2010), which would change the ease with which a search process can find that outcome. The difference in discounting functions seen between an amnesic patient (linear) and controls (hyperbolic) (Kwan *et al.*, 2011) may well be due to the inability of the amnesic patient to constructively imagine future outcomes, being dependent instead on semantic associations (Hassabis *et al.*, 2007; Kwan *et al.*, 2011), analogous to the difference seen between recognition and recall (Eichenbaum *et al.*, 2007).

### *Delay and probability discounting*

It is important to note that the probability of reward delivery in an experiment is not the same thing as ease of finding reward in a search through future outcomes. Species do show probability discounting (Madden & Bickel, 2010; Myerson *et al.*, 2011), as one would expect (Stephens & Krebs, 1986; Glimcher, 2008). The two functions are correlated in pigeons (Green *et al.*, 2010), but one needs to account for the inherent non-linearity of risk-aversion for gains and risk-seeking for losses (Kahneman & Tversky, 1979) in order to relate the two functions (Rachlin, 2006; Myerson *et al.*, 2011). Nevertheless, there remain subtle differences in probability and delay discounting functions that are not easily reconciled (such as different relationships to the amount of reward) (Myerson *et al.*, 2011). As we do not know the animal's representation of the space of situations in which probabilistic rewards are delivered (Redish *et al.*, 2007), it is not necessarily true that a high probability of reward delivery translates directly into ease of finding reward.

### *Symmetric discounting and the symmetry between episodic memory and episodic future thinking*

Yi *et al.* (2006) asked subjects the very strange question 'which would you rather have had, a smaller amount in the recent past or a larger amount in the more distant past?' As the large amount would have been available earlier, economically, all subjects should always have chosen the larger, earlier choice. This is not, however, what subjects did; instead, they showed a similar discounting function, preferring temporally-nearby choices and discounting values into the past. In a similar study looking at past discounting in smokers, smokers discounted the past more than non-smokers, in alignment with the smokers' faster future discounting rates (Bickel *et al.*, 2008). Our search-based model of discounting explains this effect through the distance to that episodic past memory outcome. In our model, discounting does not depend intrinsically on actual delay, but rather the ease of finding the outcome. Theories of episodic memory suggest that episodic memories are 'rebuilt' anew each time (Loftus & Palmer, 1974) and are akin to 'episodic past thinking' using similar mechanisms to 'episodic future thinking' (Buckner & Carroll, 2007). Our search-based model of discounting would thus suggest similar rates between forward and backward discounting. In fact, the discounting rates between past and future gains were strongly linearly correlated for both non-smokers and smokers (Bickel *et al.*, 2008). This suggests that more powerful episodic memories should show slower discounting functions than less powerful episodic memories. Whether the same cognitive effects seen in discounting of future rewards (Vohs & Faber, 2007; Peters & Büchel, 2010; Benoit *et al.*, 2011; Bickel *et al.*, 2011) will also occur in the reflected discounting of past rewards remains unknown, but that would be a prediction of our hypotheses. New evidence examining the difference between attention to future and past events seems to support this hypothesis (Radu *et al.*, 2011).

### *Hierarchical planning*

Agents engaged in episodic future thinking rarely progress through a direct sequence of states leading to that episodic future. For example, a postdoc deciding between faculty jobs is unlikely to include all paths to reach those two jobs ('First, find a moving truck. Second, drive to city X...'). Instead, the postdoc is likely to simply construct that future outcome ('Imagine, I'm now a new professor at university X. I'll interact with person Y. I'll get grad students like those I saw in person Z's lab...'). The specific path that needs to be taken to achieve the outcome is likely to be considered only if it affects the cost of reaching that outcome. There has been a lot of work integrating planning and the identification of hierarchical subgoals in reinforcement learning (Barto & Mahadevan, 2003; Botvinick & An, 2009; Ribas-Fernandes *et al.*, 2011), but an actual model including subgoals is beyond the scope of this article as it would depend on the specific nature of those subgoals. Nevertheless, our search-to-find-reward model suggests that the discounting rate seen will depend not on the total time to reach a reward, but rather on the ability to imagine that reward episodically. This predicts that more complex paths to reward will speed up discounting rates and less complex paths to reward will slow them down. In a sense, discounting rates will depend not on the actual time, but on the number of subgoals and substeps to reach that reward and the difficulty in finding the solution to each of those subgoals.

## Summary and conclusion

In this article, we note that temporal discounting of future rewards occurs even in single-trial experiments (such as with questionnaires)

and in novel situations (as in real-world deliberations). We therefore propose a model in which discounting arises from a correlation between time to a reward and the ability to find that reward in a search or settling process. This theory provides a ready explanation for the influence of cognitive phenomena on discounting functions and makes a number of predictions. Specifically, it predicts that the influence of these cognitive phenomena will be primarily seen when accessing model-based, deliberative decision-making systems and that the discounting function will depend not on the absolute delay to a reward, but on the ease of finding that reward in an episodically-imagined future.

## Acknowledgements

This work was supported by NIH grant R01 DA024080.

## References

Addis, D.R., Wong, A.T. & Schacter, D.L. (2007) Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration. *Neuropsychologia*, **45**, 1363–1377.

Ainslie, G. (1992) *Picoeconomics: The Strategic Interaction of Successive Motivational States Within the Person*. Cambridge University Press, Cambridge, UK.

Ainslie, G. (2001) *Breakdown of Will*. Cambridge University Press, Cambridge, UK.

Anker, J.J., Perry, J.L., Gliddon, L.A. & Carroll, M.E. (2009) Impulsivity predicts the escalation of cocaine self-administration in rats. *Pharmacol. Biochem. Behav.*, **93**, 343–348.

Atance, C.M. & O'Neill, D.K. (2001) Episodic future thinking. *Trends Cogn. Sci.*, **5**, 533–539.

Baddeley, A.D. (1986) *Working Memory*. Clarendon Press; Oxford University Press, Oxford.

Ballard, K., Houde, S., Silver-Babaus, S. & McClure, S.M. (2009) The decimal effect: nucleus accumbens activation predicts within-subject increases in temporal discounting rates. Program No. 93.19. 2009 Neuroscience Meeting Planner. Chicago, IL. Society for Neuroscience. Online.

Balleine, B.W. & Ostlund, S.B. (2007) Still at the choice-point: action selection and initiation in instrumental conditioning. *Ann. N.Y. Acad. Sci.*, **1104**, 147–171.

Barnes, C.A. (1979) Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. *J. Comp. Physiol. Psychol.*, **93**, 74–104.

Barto, A.G. & Mahadevan, S. (2003) Recent advances in hierarchical reinforcement learning. *Discrete Event Dyn. Syst.*, **13**, 341–379.

Baumeister, R.F., Heatherton, T.F. & Tice, D.M. (1994) *Losing Control: How and Why People Fail at Self-Regulation*. Academic Press, San Diego.

Bechara, A., Nader, K. & van der Kooy, D. (1998) A two-separate-motivational-systems hypothesis of opioid addiction. *Pharmacol. Biochem. Behav.*, **59**, 1–17.

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S.W. & Nathan, P.E. (2001) Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia*, **39**, 376–389.

Benoit, R.G., Gilbert, S.J. & Burgess, P.W. (2011) A neural mechanism mediating the impact of episodic prospection on farsighted decisions. *J. Neurosci.*, **31**, 6771–6779.

Bickel, W.K. & Marsch, L.A. (2001) Toward a behavioral economic understanding of drug dependence: delay discounting processes. *Addiction*, **96**, 73–86.

Bickel, W.K., Yi, R., Kowal, B.P. & Gatchalian, K.M. (2008) Cigarette smokers discount past and future rewards symmetrically and more than controls: is discounting a measure of impulsivity? *Drug Alcohol Depend.*, **96**, 256–262.

Bickel, W.K., Yi, R., Landes, R.D., Hill, P.F. & Baxter, C. (2011) Remember the future: working memory training decreases delay discounting among stimulant addicts. *Biol. Psychiatry*, **69**, 260–265.

Botvinick, M.M. & An, J. (2009). Goal-directed decision making in prefrontal cortex: a computational framework. In Koller, D., Bengio, Y.Y., Schuurmans, D., Bouttou, L. & Culotta, A. (Eds), *Advances in Neural Information Processing Systems (NIPS)*. MIT Press, Cambridge, MA, pp. 169–176.

Buckner, R.L. (2010) The role of the hippocampus in prediction and imagination. *Annu. Rev. Psychol.*, **61**, 27.

Buckner, R.L. & Carroll, D.C. (2007) Self-projection and the brain. *Trends Cogn. Sci.*, **11**, 49–57.

Burks, S.V., Carpenter, J.P., Goette, L. & Rustichini, A. (2009) Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proc. Natl. Acad. Sci. USA*, **106**, 7745–7750.

Daw, N.D. (2003) *Reinforcement Learning Models of the Dopamine System and Their Behavioral Implications*. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Daw, N.D., Niv, Y. & Dayan, P. (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat. Neurosci.*, **8**, 1704–1711.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P. & Dolan, R.J. (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron*, **69**, 1204–1215.

Doya, K. (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural Netw.*, **12**, 961.

Durstewitz, D., Seamans, J.K. & Sejnowski, T.J. (2000) Neurocomputational models of working memory. *Nat. Neurosci.*, **3**, 1184–1191.

Durstewitz, D., Vittoz, N.M., Floresco, S.B. & Seamans, J.K. (2010) Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron*, **66**, 438–448.

Ebert, J.E.J. & Prelec, D. (2007) The fragility of time: time-insensitivity and valuation of the near and far future. *Manag. Sci.*, **53**, 1423–1438.

Eichenbaum, H., Yonelinas, A.P. & Ranganath, C. (2007) The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.*, **30**, 123–152.

Franco-Watkins, A.M., Pashler, H. & Rickard, T.C. (2006) Does working memory load lead to greater impulsivity? Commentary on Hinson, Jameson, and Whitney (2003) *J. Exp. Psychol. Learn. Mem. Cogn.*, **32**, 443–447.

Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T. & Hutchison, K.E. (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc. Natl. Acad. Sci. USA*, **104**, 16311–16316.

Frank, M.J., Doll, B.B., Oas-Terpstra, J. & Moreno, F. (2009) Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat. Neurosci.*, **12**, 1062–1068.

Gailliot, M.T., Baumeister, R.F., DeWall, C.N., Maner, J.K., Plant, E.A., Tice, D.M., Brewer, L.E. & Schemeichel, B.J. (2007) Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. *J. Pers. Soc. Psychol.*, **92**, 325–336.

Gilbert, D.T. & Wilson, T.D. (2007) Prospection: experiencing the future. *Science*, **317**, 1351–1354.

Glimcher, P.W. (2008) *Neuroeconomics: Decision Making and the Brain*. Academic Press, London.

Gold, J.I. & Shadlen, M.N. (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, **36**, 299–308.

Green, L., Myerson, J. & Calvert, A.L. (2010) Pigeons' discounting of probabilistic and delayed reinforcers. *J. Exp. Anal. Behav.*, **94**, 113–123.

Hassabis, D., Kumaran, D., Vann, S.D. & Maguire, E.A. (2007) Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Natl. Acad. Sci. USA*, **104**, 1726–1731.

Hertz, J., Krogh, A. & Palmer, R.G. (1991) *Introduction to the Theory of Neural Computation*. Addison-Wesley Pub. Co., Redwood City, CA.

Heyman, G.M. (2009) *Addiction: A Disorder of Choice*. Harvard University Press, Cambridge, MA.

Hill, C. (2008) The rationality of preference construction (and the irrationality of rational choice). *Minn. J. Law, Sci., Tech.*, **9**, 689–742.

Hinson, J.M., Jameson, T.L. & Whitney, P. (2003) Impulsive decision making and working memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, **29**, 298–306.

Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA*, **79**, 2554–2558.

Howard, M.W., Fotedar, M.S., Datey, A.V. & Hasselmo, M.E. (2005) The temporal context model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. *Psychol. Rev.*, **112**, 75–116.

Huys, Q.J. (2007) *Reinforcers and Control: Towards a Computational Aetiology of Depression*. University College London, London.

Huys, Q.J., Eshel, N., O'Lions, E., Sheridan, L., Dayan, P. & Roiser, J.P. (2012) Bonsai trees in your head: how the Pavlovian system sculpts goal-directed choices by pruning decision trees. *PLoS Comput. Biol.*, **8**, e1002410

Johnson, A. & Redish, A.D. (2007) Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *J. Neurosci.*, **27**, 12176–12189.
Johnson, A., van der Meer, M.A. & Redish, A.D. (2007) Integrating hippocampus and striatum in decision-making. *Curr. Opin. Neurobiol.*, **17**, 692–697.
Kaelbling, L.P., Littman, M.L. & Moore, A.W. (1996) Reinforcement learning: a survey. *J. Artif. Intell. Res.*, **4**, 237–285.
Kahneman, D. & Tversky, A. (1979) Prospect theory: an analysis of decision under risk. *Econometrica*, **47**, 263–292.
Kliegel, M., McDaniel, M.A. & Einstein, G.O. (2008) *Prospective memory cognitive, neuroscience, developmental, and applied perspectives*. Lawrence Erlbaun Associates, New York, NY.
Kohonen, T. (1980) *Content-Addressable Memories*. Springer-Verlag, Berlin; New York.
Krajbich, I., Rangel, A. & Armel, C. (2010) Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.*, **13**, 1292–1298.
Kurth-Nelson, Z. & Redish, A.D. (2009) Temporal-difference reinforcement learning with distributed representations. *PLoS ONE*, **4**, e7362.
Kwan, D., Craver, C.F., Green, L., Myerson, J., Boyer, P. & Rosenbaum, R.S. (2011) Future decision-making without episodic mental time travel. *Hippocampus*, in press.
Laibson, D. (1997) Golden eggs and hyperbolic discounting. *Q. J. Econ.*, **112**, 443–477.
Lauwereyns, J. (2010) *The anatomy of bias how neural circuits weigh the options*. MIT Press, Cambridge, MA.
Loftus, E.F. & Palmer, J.C. (1974) Reconstruction of automobile destruction: an example of the interaction between language and memory. *J. Verb. Learn. Verb. Behav.*, **13**, 585–589.
Madden, G.J. & Bickel, W.K. (2010) *Impulsivity: The Behavioral and Neurological Science of Discounting*. American Psychological Association, Washington, DC.
Mazur, J.E. (1997) Choice, delay, probability, and conditioned reinforcement. *Anim. Learn. Behav.*, **25**, 131.
McClure, S.M., Laibson, D.I., Loewenstein, G. & Cohen, J.D. (2004) Separate neural systems value immediate and delayed monetary rewards. *Science (New York, N.Y.)*, **306**, 503–507.
McDannald, M.A., Schoenbaum, G., Lucantonio, F., Burke, K.A. & Niv, Y. (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. *J. Neurosci.*, **31**, 2700–2705.
van der Meer, M.A. & Redish, A.D. (2009) Covert expectation-of-reward in rat ventral striatum at decision points. *Front. Integr. Neurosci.*, **3**, 1.
van der Meer, M.A. & Redish, A.D. (2010) Expectancies in decision making, reinforcement learning, and ventral striatum. *Front. Neurosci.*, **4**, 6.
van der Meer, M.A. & Redish, A.D. (2011) Ventral striatum: a critical look at models of learning and evaluation. *Curr. Opin. Neurobiol.*, **21**, 387–392.
van der Meer, M.A., Kurth-Nelson, Z. & Redish, A.D. (2012) Information processing in decision-making systems. *Neuroscientist*, in press.
Mischel, W. & Grusec, J. (1967) Waiting for rewards and punishments: effects of time and probability on choice. *J. Pers. Soc. Psychol.*, **5**, 24–31.
Mischel, W. & Underwood, B. (1974) Instrumental ideation in delay of gratification. *Child Dev.*, **45**, 1083–1088.
Montague, P.R., Dolan, R.J., Friston, K.J. & Dayan, P. (2012) Computational psychiatry. *Trends Cogn. Sci.*, **16**, 72–80.
Myerson, J., Green, L. & Morris, J. (2011) Modeling the effect of reward amount on probability discounting. *J. Exp. Anal. Behav.*, **95**, 175–187.
Newell, A. & Simon, H.A. (1972) *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, NJ.
Nilsson, N.J. (1998) *Artificial Intelligence: A New Synthesis*. Morgan Kaufmann Publishers, San Francisco, CA.
Niv, Y., Joel, D. & Dayan, P. (2006) A normative perspective on motivation. *Trends Cogn. Sci.*, **10**, 375–381.
Odum, A., Madden, G. & Bickel, W. (2002) Discounting of delayed health gains and losses by current, never- and ex-smokers of cigarettes. *Nicotine Tob. Res.*, **4**, 295–303.
Packard, M.G. & McGaugh, J.L. (1996) Inactivation of hippocampus or caudate nucleus with lidocaine differentially affects expression of place and response learning. *Neurobiol. Learn. Mem.*, **65**, 65–72.
Padoa-Schioppa, C. & Assad, J.A. (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature*, **441**, 223–226.
Pennartz, C.M., Berke, J.D., Graybiel, A.M., Ito, R., Lansink, C.S., van der Meer, M.A., Redish, A.D., Smith, K.S. & Voorn, P. (2009) Corticostriatal interactions during learning, memory processing, and decision making. *J. Neurosci.*, **29**, 12831–12838.
Perry, J.L., Larson, E.B., German, J.P., Madden, G.J. & Carroll, M.E. (2005) Impulsivity (delay discounting) as a predictor of acquisition of IV cocaine self-administration in female rats. *Psychopharmacology*, **178**, 2–3.
Peters, J. & Büchel, C. (2010) Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, **66**, 138–148.
Peters, J. & Büchel, C. (2011) The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends Cogn. Sci.*, **15**, 227–239.
Rachlin, H. (2004) *The Science of Self-Control*. Harvard University Press, Cambridge, MA; London.
Rachlin, H. (2006) Notes on discounting. *J. Exp. Anal. Behav.*, **85**, 425–435.
Radu, P.T., Yi, R., Bickel, W.K., Gross, J.J. & McClure, S.M. (2011) A mechanism for reducing delay discounting by altering temporal attention. *J. Exp. Anal. Behav.*, **96**, 363–385.
Ratcliff, R. & McKoon, G. (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comp.*, **20**, 873–922.
Redish, A.D. (1999) *Beyond the cognitive map from place cells to episodic memory*. MIT Press, Cambridge, MA.
Redish, A.D. & Kurth-Nelson, Z. (2010) Neural models of delay discounting. In Madden, G.J. & Bickel, W.K. (Eds), *Impulsivity: The Behavioral and Neurological Science of Discounting*. American Psychological Association, Washington, DC, pp. 123–158.
Redish, A.D., Jensen, S., Johnson, A. & Kurth-Nelson, Z. (2007) Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychol. Rev.*, **114**, 784–805.
Redish, A.D., Jensen, S. & Johnson, A. (2008) A unified framework for addiction: vulnerabilities in the decision process. *Behav. Brain Sci.*, **31**, 415–437.
Restle, F. (1957) Discrimination of cues in mazes: a resolution of the place-vs.-response question. *Psychol. Rev.*, **64**, 217–228.
Ribas-Fernandes, J.J.F., Solway, A., Diuk, C., McGuire, J.T., Barto, A.G., Niv, Y. & Botvinick, M.M. (2011) A neural signature of hierarchical reinforcement learning. *Neuron*, **71**, 370–379.
Rich, E. & Knight, K. (1991) *Artificial Intelligence*. McGraw-Hill, New York.
Roesch, M.R., Schoenbaum, G., Singh, T., Leon, B.P. & Mullins, S.E. (2009) Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. *J. Neurosci.*, **29**, 13365–13376.
Samuelson, P.A. (1937) A note on measurement of utility. *Rev. Econ. Stud.*, **4**, 155–161.
Schacter, D.L. & Addis, D.R. (2007) Constructive memory: the ghosts of past and future. *Nature*, **445**, 27.
Schacter, D.L., Addis, D.R. & Buckner, R.L. (2007) Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci.*, **8**, 657–661.
Schacter, D.L., Addis, D.R. & Buckner, R.L. (2008) Episodic simulation of future events: concepts, data, and applications. *Ann. N. Y. Acad. Sci.*, **1124**, 39–60.
Schoenbaum, G., Stalnaker, T.A., Takahashi, Y.K. & Roesch, M.R. (2009) A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nat. Rev. Neurosci.*, **10**, 885–892.
Schweighofer, N., Shishida, K., Han, C.E., Okamoto, Y., Tanaka, S.C., Yamawaki, S. & Doya, K. (2006) Humans can adopt optimal discounting strategy under real-time constraints. *PLoS Comput. Biol.*, **2**, e152.
Seamans, J.K. & Yang, C.R. (2004) The principal features and mechanisms of dopamine modulation in the prefrontal cortex. *Prog. Neurobiol.*, **74**, 1–58.
Simon, D.A. & Daw, N.D. (2011) Neural correlates of forward planning in a spatial decision task in humans. *J. Neurosci.*, **31**, 5526–5539.
Simonson, I. & Tversky, A. (1992) Choice in context: tradeoff contrast and extremeness aversion. *J. Mark. Res.*, **29**, 281–295.
Smith, A., Li, M., Becker, S. & Kapur, S. (2006) Dopamine, prediction error and associative learning: a model-based account. *Network*, **17**, 61–84.
Sozou, P.D. (1998) On hyperbolic discounting and uncertain hazard rates. *Proc. Roy. Soc. London B Bio. Sci.*, **265**, 2015–2050.
Stanger, C., Ryan, S.R., Fu, H., Landes, R.D., Jones, B.A., Bickel, W.K. & Budney, A.J. (2011) Delay discounting predicts adolescent substance abuse treatment outcome. *Exp. Clin. Psychopharmacol.*, in press.
Stephens, D.W. & Krebs, J.R. (1986) *Foraging Theory*. Princeton University Press, Princeton, NJ.
Suri, R.E. (2002) TD models of reward predictive responses in dopamine neurons. *Neural Netw.*, **15**, 523–533.
Sutton, R.S. & Barto, A.G. (1998) *Reinforcement learning an introduction*. MIT Press, Cambridge, MA.
Tolman, E.C. (1939) Prediction of vicarious trial and error by means of the schematic sowbug. *Psychol. Rev.*, **46**, 318–336.

Trope, Y. & Liberman, N. (2003) Temporal construal. *Psychol. Rev.*, **110**, 403–421.
Vohs, K.D. & Faber, R.J. (2007) Spent resources: self-regulatory resource availability affects impulse buying. *J. Consum. Res.*, **33**, 537–547.
Wilson, H.R. & Cowan, J.D. (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.*, **12**, 1–24.
Winstanley, C., Robbins, T.W., Balleine, B.W., Brown, J.W., Büchel, C., Cools, R., Durstewutz, D., Redish, A.D., Seamans, J.K. & Robbins, T.W. (2012). Cognitive control, cognitive search, and motivational salience: a systems neuroscience approach. In Todd, P.M., Hills, T.T. & Robbins, T.W. (Eds), *Cognitive Search: Evolution, Algorithms, and the Brain (Strüngmann Forum Report)*. MIT Press, Cambridge, MA, in press.
Yamashita, K. & Tanaka, S. (2005) Parametric study of dopaminergic neuromodulatory effects in a reduced model of the prefrontal cortex. *Neurocomputing*, **65**, 579.
Yi, R., Gatchalian, K.M. & Bickel, W.K. (2006) Discounting of past outcomes. *Exp. Clin. Psychopharmacol.*, **14**, 311–317.
Yin, H.H. & Knowlton, B.J. (2004) Contributions of striatal subregions to place and response learning. *Learn. Mem.*, **11**, 459–463.
Zald, D.H. & Rauch, S.L. (2006) *The Orbitofrontal Cortex*. Oxford University Press, Oxford; New York.
Zhang, K. (1996) Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.*, **16**, 2112–2126.