Behavioral/Systems/Cognitive

Evidence for Hyperbolic Temporal Discounting of Reward in Control of Movements

Adrian M. Haith,[1,2]* Thomas R. Reppert,[1]* and Reza Shadmehr[1]
Departments of [1]Biomedical Engineering and [2]Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland 21287

Suppose that the purpose of a movement is to place the body in a more rewarding state. In this framework, slower movements may increase accuracy and therefore improve the probability of acquiring reward, but the longer durations of slow movements produce devaluation of reward. Here we hypothesize that the brain decides the vigor of a movement (duration and velocity) based on the expected discounted reward associated with that movement. We begin by showing that durations of saccades of varying amplitude can be accurately predicted by a model in which motor commands maximize expected discounted reward. This result suggests that reward is temporally discounted even in timescales of tens of milliseconds. One interpretation of temporal discounting is that the true objective of the brain is to maximize the rate of reward—which is equivalent to a specific form of hyperbolic discounting. A consequence of this idea is that the vigor of saccades should change as one alters the intertrial intervals between movements. We find experimentally that in healthy humans, as intertrial intervals are varied, saccade peak velocities and durations change on a trial-by-trial basis precisely as predicted by a model in which the objective is to maximize the rate of reward. Our results are inconsistent with theories in which reward is discounted exponentially. We suggest that there exists a single cost, rate of reward, which provides a unifying principle that may govern control of movements in timescales of milliseconds, as well as decision making in timescales of seconds to years.

Introduction

Temporal discounting of reward is a ubiquitous phenomenon in decision making. Across many types and magnitudes of reward, multiple timescales, and various species, small, immediate rewards are often preferred over larger, delayed rewards. Mathematically, temporal discounting of reward may be described in terms of a multiplicative discount function:

$$V(t_0 + t) = V(t_0)\,F(t). \quad (1)$$

In Equation 1, reward value at current time t_0 is discounted by a function $F(t)$ to produce value at time $t_0 + t$, with $F(0) = 1$. The two most common forms of $F(t)$ that have been used to describe discounting are exponential

$$F(t) = \exp(-kt) \quad (2)$$

and hyperbolic

$$F(t) = 1/(1 + \beta t). \quad (3)$$

For example, exponential temporal discounting is routinely used in a form of reinforcement learning known as temporal difference learning (Sutton and Barto, 1981), which provides a prominent theory of learning in the basal ganglia (Schultz et al., 1997). Exponential discounting has also been suggested in models of human decision making (Schweighofer et al., 2006). Hyperbolic discounting, however, is more consistent with behavioral data in humans (Myerson and Green, 1995) and monkeys (Kobayashi and Schultz, 2008). While it is clear that the brain temporally discounts reward, the exact shape of this function is not entirely clear. Perhaps more significantly, the reason why temporal discounting occurs at all is poorly understood.

Recently we proposed that the way the brain discounts reward may have implications for control of movements (Shadmehr et al., 2010). Suppose that a movement is made with the purpose of acquiring some rewarding state that has value $V(t_0) = a$. In this framework, the duration of the movement acts as a delay in acquiring reward. Performing a movement slowly diminishes the value of reward upon its acquisition (movement end), making it preferable to move quickly. Fast movements, however, are more variable (Fitts, 1954; Schmidt et al., 1979), reducing the probability of success for the movement. Therefore, the expected value of a stimulus that is acquired after some movement duration τ is affected by two factors: probability of successfully acquiring the stimulus $P[\text{success}|\tau]$, which increases with duration τ, and temporal discounting of reward value $F(\tau)$, which decreases with duration τ:

$$E[\text{reward}|\tau] = aP[\text{success}|\tau]\,F(\tau). \quad (4)$$

Thus, if the objective for the brain is to produce movements that maximize the expected value of reward, then movement speed and duration or, collectively, vigor should be a balance between the competing concerns of time and variability.

Received Jan. 30, 2012; revised May 25, 2012; accepted July 5, 2012.

Author contributions: A.M.H. and R.S. designed research; A.M.H. and T.R.R. performed research; T.R.R. analyzed data; A.M.H. and R.S. wrote the paper.

This work was supported by grants from the NIH (NS37422) and the Human Frontiers Science Program. T.R.R. is supported by a National Science Foundation Graduate Research Fellowship.

*A.M.H. and T.R.R. contributed equally to this work.

Correspondence should be addressed to Adrian M. Haith, 210 Carnegie, 600 North Wolfe Street, Johns Hopkins School of Medicine, Baltimore, MD 21287. E-mail: adrian.haith@jhu.edu.

DOI:10.1523/JNEUROSCI.0424-12.2012

 0270-6474/12/3211727-10$15.00/0

However, our proposed link between temporal discounting in motor control and decision making is tenuous: the movements that we are considering (saccades) are tens of milliseconds in duration. Why should a few milliseconds make a meaningful difference in the value of reward? Here, we show that one interpretation of hyperbolic temporal discounting is that the brain selects actions so as to maximize the rate of reward. This idea leads to a novel prediction about how the brain should select vigor in response to changes in the intertrial interval between movements. We propose that rate of reward provides a unifying principle that governs control of movements in timescale of milliseconds, as well as decision making in timescales of seconds to years.

Materials and Methods

Our concern is the general question of why movements have their specific kinematic properties, i.e., why movements of a given amplitude have a particular duration and velocity. Here, we present a novel framework for considering the influence of temporal discounting of reward on movement vigor. Our focus is on saccades, as numerous theories have been proposed to explain the kinematic patterns of these simple movements, enabling us to focus on the question of how temporal discounting influences choice of movement vigor. Our principal new theoretical result, presented in the Results section, is that the shape of the discount function should leave its signature in how the brain alters saccadic vigor in response to changes in intertrial intervals between saccades. We will first present the computational methods that we used to study the theoretical relationship between movement vigor and temporal discount functions, and then the experiments that we performed to test some of the predictions.

Model of eye plant. We modeled the oculomotor plant as a second order dynamical system:

$$m\ddot{x} = -kx - b\dot{x} + f. \tag{5}$$

In this equation, x is the lateral deviation from the equilibrium point of the eye, m is the inertia of the eye, k is stiffness, b is viscosity, and f is the instantaneous force generated by the extra-ocular muscles, which act as a first-order linear filter of the motor command u:

$$\gamma\dot{f} = -f + u. \tag{6}$$

Here, γ is a time-constant that determines how quickly motor commands are transmitted into forces. If we represent the full state of the plant by the vector $\mathbf{x} = [x, \dot{x}, f]^T$, the dynamics can be more compactly expressed in continuous time as:

$$\dot{\mathbf{x}} = A_c\mathbf{x} + \mathbf{b}_c u, \tag{7}$$

with $A_c = \begin{pmatrix} 0 & 1 & 0 \\ -k/m & -b/m & 1/m \\ 0 & 0 & -1/\gamma \end{pmatrix}$ and $\mathbf{b}_c = \begin{pmatrix} 0 \\ 0 \\ 1/\gamma \end{pmatrix}$. As with our previous work (Shadmehr et al., 2010), we set the parameters of the eye plant to match the three timescales described by Robinson (1986): $\tau_1 = 0.224$, $\tau_2 = 0.013$, and $\tau_3 = 0.004$ s. This can be achieved by setting $k = 1$, $b = \tau_1 + \tau_2$, $m = \tau_1\tau_2$, and $\gamma = \tau_3$. These equations were converted into discrete-time using matrix exponentials for a time step Δ of 0.1 ms:

$$\mathbf{x}_{t+\Delta} = A\mathbf{x}_t + \mathbf{b}u_t.$$

Next, we added signal-dependent $\varepsilon_t \sim N(0, \kappa^2 u_t^2 \Delta)$ and non-signal-dependent $\chi_t \sim N(0, \lambda^2 \Delta)$ noise to the model:

$$\mathbf{x}_{t+\Delta} = A\mathbf{x}_t + \mathbf{b}(u_t + \varepsilon_t + \chi_t) \tag{8}$$

as described by van Beers (2007). It is difficult to reliably estimate the magnitude of signal-dependent noise from empirical data due to the many potential sources of variability in eye movements, although constant noise is more reliably inferred (van Beers, 2007). We therefore set $\lambda = 0.0075$ kg m s^{-2} to match the horizontal endpoint variability reported in that work, and left the magnitude of signal-dependent noise κ as an open parameter.

To extend this one-dimensional model to a more realistic two-dimensional one, we assumed independent vertical and horizontal components of the eye, with independent sources of noise. Following the method of van Beers (2007), we scaled variability in the vertical direction by a factor of 1.14 to reflect the sparser innervation of muscles in that direction relative to horizontal. Our experiments presented targets approximately along the horizontal axis. As a result, the magnitude of signal-dependent noise introduced for the vertical component of movement was found to be negligible for all movements we considered and we therefore included only signal-independent variability along this axis. We used this two-dimensional plant model in all subsequent simulations.

Making saccades to maximize probability of success. Suppose that success or failure of a point-to-point movement such as a saccade is determined by whether or not the location of the effector at the end of the movement falls within a specified goal region. The probability of success depends on the distribution of the effector endpoints. For our linear system with Gaussian additive and multiplicative noise (Eq. 8), for a movement of duration τ, the endpoint distribution for any sequence of motor commands $u_0, \ldots, u_{\tau-1}$ is Gaussian. Suppose that, in one dimension, for a particular sequence of motor commands the end position x_τ of the saccade has a distribution with mean μ and variance σ^2. If the target of the movement is at location a with respect to the fovea, and the fovea has width w, then the probability of success (i.e., acquiring reward) is defined as:

$$P[\text{success}] = \int_{a-\frac{1}{2}w}^{a+\frac{1}{2}w} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_\tau - \mu)^2}{2\sigma^2}\right) dx_\tau. \tag{9}$$

We assumed that $w = 1°$ (approximate width of the fovea). If we assume that the mean of the endpoint distribution is aligned with the target center, the probability of success becomes:

$$P[\text{success}] = \text{erf}\left(\frac{w}{2\sqrt{2\sigma^2}}\right). \tag{10}$$

In two dimensions, we made the simplifying assumption that the overall probability of success was given by a product of the corresponding probabilities of success along the horizontal and vertical axes (effectively assuming that the target was a $1° \times 1°$ square).

For a given target position, we needed to compute the probability of success as a function of duration τ. To do so, for a given τ we found the motor commands that maximized the probability of success. This was achieved by finding the motor commands that minimized endpoint variance with the constraint that the mean of the endpoint distribution was at the target position at time τ and remained there for a further 50 ms hold period. We solved this constrained optimization problem analytically using Lagrange multipliers.

Discounting probability of success. How does one decide the vigor with which to perform a movement? A simple hypothesis is that we choose the motor commands that maximize probability of success (Eq. 10; van Beers, 2008). However, as we will see, generating motor commands that maximize probability of success produces movement durations that agree with observed data on small amplitude saccades, but fails for large amplitude saccades. Our theory proposes that this failure is because reward does not have a constant value as a function of time. Rather, reward is discounted as a function of movement duration τ by a temporal discount function $F(\tau)$. We consider different forms for this temporal discount function (as described in Results), compute the discounted reward value, and then numerically find the movement duration that maximizes the expected value of the discounted reward function.

The value of the signal-dependent noise parameter κ is unknown and difficult to estimate. Furthermore, the duration-amplitude relationships of all models we considered were highly sensitive to the exact value of this

parameter. To generate the relationship between saccade amplitude and duration, we fit this single open parameter separately for the discounted reward model and for the maximum probability of reward model by finding the values that yielded a predicted duration of 105 ms for a 30° saccade. This yielded estimates for κ of 0.0066 and 0.0055 for the discounted reward and maximum probability of reward models, respectively. The estimate of κ for the discounted reward model was then used directly to predict the influence of changes in intertrial interval on movement vigor. There, we also assumed that subjects maintained an estimate of the intertrial interval on each trial $\hat{p}_i$, and then updated this estimate based on the observed intertrial interval: $\hat{p}_{i+1} = \hat{p}_i + \eta(p_i - \hat{p}_i)$. We set $\eta = 0.7$. We then set the intersaccade interval in the discount function (δ in Eq. 14) equal to the estimated intertrial interval plus a reaction time of 150 ms and determined the movement duration that theoretically maximized the discounted reward. Peak velocity was computed by simulating the saccade at the optimal movement duration using the motor commands that minimize the endpoint variance.

Experimental methods. A critical prediction of our theoretical work is that, if movement duration is determined by maximizing reward rate, then the brain should alter movement vigor in response to changes in the intertrial interval (ITI) between movements in a particular way. Our theory suggests that alternative forms of discounting may have different characteristic patterns by which changes in ITI lead to changes in movement vigor. We performed experiments to test our predictions. All experimental procedures were approved by the Johns Hopkins Institutional Review Board.

For our main experiment, we recruited $n = 6$ healthy volunteers (mean age 28, range 23–47, five females). Subjects sat in a dark room in front of a CRT monitor (36.5 × 27.5 cm, 1024 × 768 pixel, light gray background, frame rate 120 Hz) with head restrained using either a dental bite bar or chin and forehead rests and their left eye covered. Targets (blue, diameter = 1°) were presented with Matlab 7.4 (MathWorks) using the Psychophysics Toolbox. The screen was placed at a distance of 31 cm from the subject's face, and an EyeLink 1000 (SR Research) infrared camera recording system (sampling rate = 1000 Hz) was used to record movement of the right eye.

Subjects were asked to make saccades between targets having a horizontal separation of 40°, and positioned symmetrically about the center of the screen. Each saccade was cued by the appearance of one of two possible targets, having a vertical separation of 5° between them. Including two potential targets discouraged subjects from generating predictive saccades in advance of the actual target presentation. After an initial training period, subjects completed 12 blocks of 80 trials. Each trial consisted of three parts: intertrial interval, reaction time, and movement time. The intertrial interval began when the subject's gaze was within 3° of the target and ended with extinction of the current target and presentation of the next target. There were three different block types in which the intertrial interval was varied in different ways. In the constant ITI blocks, the ITI was fixed at 1 s throughout the block. In the increasing ITI blocks, the ITI was set to 1 s for the first 10 trials, then was abruptly decreased to 0.4 s, before slowly increasing to 1.6 s over the next 60 trials, then was restored to 1 s for the final 10 trials of the block. In the decreasing ITI blocks, the opposite sequence of ITIs was used, with an initial abrupt increase to 1.6 s preceding a slow decrease to 0.4 s. Blocks were presented in a pseudorandom sequence that was different for each individual subject.

An additional $n = 5$ subjects (mean age 26, age range 21–47, 4 females) participated in a control experiment in which the previous target did not disappear when the new target was presented, but remained on the screen until the gaze reached the new target. Instead of two potential targets for each saccade, there was only one possible target in this control experiment.

All data analysis was completed using Matlab R2011a (MathWorks). The gaze position data were filtered using a second-order Savitzky-Golay filter with a half-width of 27 ms. Saccade beginning and end were marked using a 20°/s velocity threshold. Five criteria were used to assess saccades: (1) Amplitude between 35 and 45°; (2) Duration between 50 and 350 ms; (3) Reaction time between 100 and 500 ms; (4) No blinking during the saccade; (5) Saccade velocity profile exhibits only one maximum. Any saccade that did not meet all 5 criteria was excluded from the analysis (approximately 10% of all saccades).

Peak velocities in a given block were normalized separately for nasal (leftward) and temporal (rightward) saccades by dividing by the mean peak velocity during the first 10 trials across all blocks. The normalized peak velocity was then averaged across all four repeats per block type for each subject. Saccade duration, amplitude, and reaction time were normalized using the same method. We assessed the effect of changing ITI on the variables of interest using an analysis of covariance on data from all subjects for the middle 60 trials of each block (during which the ITI changed), with trial number serving as a continuous predictor and block type as a categorical predictor. Our hypothesis that changes in ITI should lead to changes in the kinematic properties of saccades then corresponds to a predicted interaction effect between trial number and block type.

Results

An early model of saccades (Harris and Wolpert, 1998) showed that saccade trajectories for a given duration are well predicted by a model in which motor commands are selected to minimize endpoint variance in the presence of signal-dependent noise (term ε_t in Eq. 8). That idea is equivalent to finding motor commands that maximize the probability of success for a given duration. However, the question of how movement durations are selected was not addressed. A more recent work (van Beers, 2007) empirically demonstrated that in addition to signal-dependent noise, the oculomotor plant suffers from non-signal-dependent noise (term χ_t in Eq. 8). The non-signal-dependent noise acts as a natural cost that penalizes movement durations: the longer the duration of the movement, the greater the endpoint variance due to accumulation of this kind of noise. The impact of non-signal-dependent noise grows monotonically with saccade duration, producing a greater penalty for longer durations of movement. Therefore, if one assumes that the objective is to maximize probability of success (or, equivalently minimize endpoint variance), then one can compute the optimal movement duration for any given saccade amplitude (van Beers, 2008).

We adopted a model of the oculomotor plant based on previous publications (Robinson, 1986; van Beers, 2007). For reasons which we explain below, we used a value of $\kappa = 0.0066$ for the signal-dependent noise magnitude in our simulations. Using this model, we computed the probability that at the end of a 10° saccade the target would be placed on the fovea (Fig. 1*A*, top subplot). Increasing movement duration initially increases the probability of success (this is because of the diminishing impact of signal-dependent noise). However, as movement durations become long, the probability of success tends to decline (this is because of the increasing influence of non-signal-dependent noise). Therefore, if the objective is to maximize probability of success, a 10° saccade should last ~56 ms, a prediction that falls within the range of durations measured for actual saccades of that amplitude (Collewijn et al., 1988).

Figure 1*B* (top subplot) shows the probability of success for a saccade of 40° amplitude. Under the maximum probability of success hypothesis, a 40° saccade has duration of approximately 200 ms, a value that far exceeds that of observed data (around 135 ms). As the saccade amplitude increases further, the optimal duration begins to increase at an increasing rate and rapidly becomes unrealistic. Therefore, whereas maximizing probability of success under these noise properties produces durations that match observed data for small amplitudes, this policy grossly overestimates saccade durations for large amplitudes (Fig. 1*C*). The reason for this failure is that the variance due to signal-independent noise saturates at around 200–300 ms. Increasing saccade duration beyond this point carries little additional cost from signal-independent noise, but continues to

reduce the impact of signal-dependent noise—particularly for large amplitude saccades. This characteristic is independent of the particular value of κ used in our simulations. The dashed line in Figure 1*C* shows the predicted durations given a value of $\kappa = 0.0055$, which was chosen so that maximizing the probability of reward of a 40° saccade yielded a correct prediction of 135 ms. Even for this optimized parameter, large amplitude saccades are predicted to have unrealistically long durations. If instead we were to fix κ based on the duration of larger amplitude saccades, the predictions for lower amplitude saccades become unrealistically short. As a result, we find that saccade durations are inconsistent with a policy that maximizes probability of success.

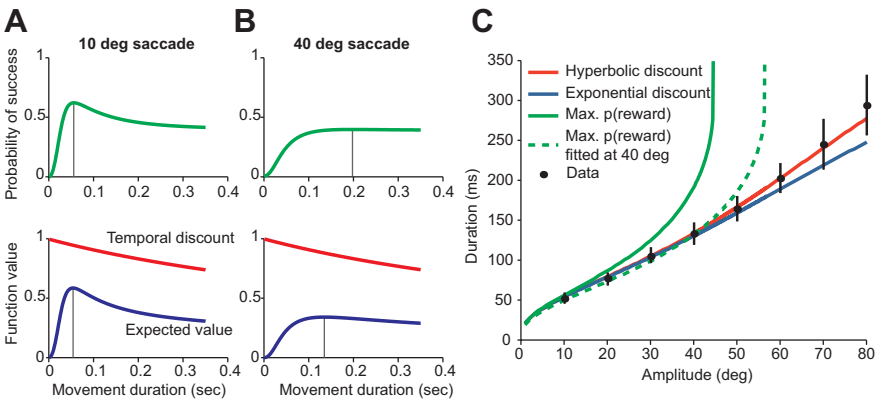

Figure 1. Relationship between saccade duration, endpoint accuracy, and expected value of reward. ***A***, Top, We computed probability of success (probability that the motor commands will place the target on the fovea) for a 10° saccade as a function of movement duration. Signal-dependent noise magnitude was set to $\kappa = 0.0066$. Movement duration that maximizes probability of success is indicated by the vertical line. Bottom, Expected discounted value of the reward attained at the completion of the saccade (blue line) under the assumption that reward is discounted hyperbolically in time (red line). The temporal discount function is $F(\tau) = 1/(1 + \tau)$. For a 10° saccade, the movement duration that maximizes the probability of success is similar to that which maximizes the expected value of reward. ***B***, Same as A but for a 40° saccade. For this saccade amplitude, the movement duration that maximizes the expected value of reward is much shorter than one that maximizes the probability of success. ***C***, Relationship between saccade amplitude and saccade duration predicted by maximum probability of success hypothesis (green), and maximum expected value of reward hypothesis (red and blue). For the expected value of reward hypothesis, durations that maximize hyperbolically discounted expected rewards are shown in red, and durations that maximize exponentially discounted expected rewards are shown in blue. For hyperbolic discounting, $F(\tau) = 1/(1 + \tau)$. For exponential discounting, $F(\tau) = \exp(-\tau)$. Also plotted are experimental data from Collewijn et al. (1988) (filled circles; vertical bars indicate ± 1 SD). The dashed green line indicates predictions of the maximum probability of success model under a noise model fitted to generate accurate predictions for a 40° saccade ($\kappa = 0.0055$).

Temporal discounting of reward

Consider the possibility that the value of the stimulus is not constant as a function of time, but is discounted. As a result, the expected value of the stimulus at movement completion depends on two factors (Eq. 4): probability of success, and a temporal discount function $F(\tau)$ which describes change in stimulus value during the movement. Let us show that a policy that maximizes the expected discounted value of reward reproduces the observed amplitude-duration relationship.

Suppose that stimulus value is discounted hyperbolically $F(\tau) = 1/(\beta\tau + 1)$. We set $\beta = 1\ s^{-1}$ (for reasons we explain below). Under our assumed noise characteristics, the optimum duration for a 10° saccade is around 54 ms and for a 40° saccade, the same parameters produce an optimum saccade duration of 135 ms—both offering a good agreement with observed saccade durations. Indeed, a single hyperbolic discount function can accurately reproduce saccade durations for the entire range of amplitudes of recorded data (Fig. 1*C*). We also considered an exponential discount function, which has general form $F(\tau) = \exp(-\lambda\tau)$. We set $\lambda = 1\ s^{-1}$ so that the hyperbolic and exponential discount functions would share the same gradient at $\tau = 0$ s. This exponential discount model also produces realistic duration predictions for the entire range of saccade amplitudes (Fig. 1*C*). Therefore, saccade durations are consistent with a policy that maximizes the expected discounted value of the stimulus (Eq. 4).

The specific value of $\kappa = 0.0066$ for these simulations was chosen such that a 30° saccade would have an optimal duration of 105 ms under a hyperbolic discounting of reward model. This value of κ is approximately consistent with the magnitude of signal-dependent variability reported by van Beers (2007). However, given the sensitivity of our predicted durations to the details of our underlying model of saccade generation (plant properties and selection of motor commands for a given duration), we cannot be certain about the precise relationship between the discount function and saccade durations. Furthermore, these data do not allow us to dissociate between hyperbolic and exponential forms of discounting. We can, however, reject the possibility that saccade durations are selected to maximize expected undiscounted reward, since no single value of κ could account for saccade durations across all amplitudes. Therefore, at this point we can only conclude that temporal discounting of reward plays a role in determining movement durations. The exact shape of the discount function remains unclear.

Temporal discounting as reward rate optimization

The results that we have presented thus far are similar to those that we saw in an earlier set of simulations (Shadmehr et al., 2010). In that work we assumed a cost in which movement endpoint errors were penalized with a quadratic function and movement duration was penalized through an added hyperbolic time cost. Here, we instead adopt a more natural framework in which successful movements (endpoint falling within a specified goal region) earn a positive reward and unsuccessful movements earn zero reward, regardless of the magnitude of the error. The value of reward associated with a successful movement is then discounted multiplicatively as a function of time. In addition, whereas our previous work assumed an oculomotor plant with only signal-dependent noise, we now consider a more accurate model of the oculomotor plant with both signal-dependent and signal-independent noise sources. One may argue over the relative merits of each model, but the fact is that both the current and the previous work suffer from two fundamental concerns: (1) We have merely shown that observed data on saccade durations are consistent with our temporal discounting framework. However, there may be many kinds of costs that are also consistent with these data. (2) In decision making, reward is temporally devalued over timescales of minutes, days, or years. In our model of saccades, reward is discounted over a timescale of milliseconds. It seems improbable that a few milliseconds should produce any meaningful change in the perceived value of reward. To address these concerns, we must first understand the deeper question of why the brain should discount reward at all.

A common interpretation of temporal discounting is that the risk of not getting a predicted reward increases with delay. If reward remains available for a duration that follows an exponential distribution (and reward disappearance behaves like a Poisson process), then exponential discounting maximizes the total expected reward. Such a framework can also account for hyperbolic discounting if the exponential distribution is replaced with a mixture of exponential distributions with different time constants (Daw, 2003). An alternative interpretation is that temporal discounting reflects a desire to maximize the rate of reward acquisition, rather than the absolute value of each acquired reward (Kacelnik, 1997; Daw, 2003; Niv et al., 2007). To explain this, consider a choice between an immediate reward with magnitude α_1, and a larger reward α_2 at some time in the future τ. If we assume that the next such decision will not occur immediately after we receive the reward, but after some average period of time δ (due to reaction time, intertrial interval, etc.), then the reward rates associated with each choice are

$$R_1 = \frac{\alpha_1}{\delta} \quad R_2 = \frac{\alpha_2}{\tau + \delta}. \tag{11}$$

Suppose we vary α_2 and find the value for which we select the immediate but smaller reward α_1 at 50% probability. According to the rate of reward theory, this indifference between the immediate but smaller reward and the delayed but larger reward is occurring because $R_1 = R_2$. This condition occurs when

$$\alpha_1 = \frac{\alpha_2}{1 + \delta^{-1}\tau}. \tag{12}$$

We see that if we make choices in such a way as to maximize the rate of reward, then effectively we discount the value of the delayed reward α_2 hyperbolically with a rate that depends on the average duration δ between opportunities to earn reward. The key new idea that emerges is that hyperbolic temporal discounting arises because the underlying objective of the brain is to optimize the rate of reward R (i.e., reward per unit of time):

$$E[R] = aP[\text{success}|\tau]\frac{1}{\tau + \delta}, \tag{13}$$

where δ is the inter-movement interval, and τ is movement duration. This is proportional to the expected hyperbolically discounted reward (Eqs. 3 and 4) when $\beta = \delta^{-1}$. In the simulations presented in Figure 1, we deliberately set β equal to 1 s^{-1}, corresponding to an inter-movement interval of 1 s, roughly consistent with the experimental paradigm for the study (Collewijn et al., 1988) that collected the data in Figure 1*C* (although exact ITI data were not reported in that paper). That is, the specific hyperbolic temporal discount used to produce the simulations in Figure 1*C* is equivalent to Equation 13 in which ITI is around 1 s. This establishes the plausibility that the rate of reward hypothesis could in principle account for vigor of saccades.

Predictions of the rate of reward theory

While the above findings establish the plausibility of the rate of reward hypothesis, a far stronger prediction of this theory is that if we change the intersaccade interval δ, the brain will change the vigor of saccades. In a typical experiment one gives a sequence of targets, and the subject makes a sequence of movements to these targets. Equation 13 predicts that the expected reward rate will depend on the average duration of each movement τ plus the average inter-movement interval δ. If we change δ, for example by increasing the time between the end of one movement and presentation of the target for the next movement, then the vigor with which that movement is performed should change. Here is the critical prediction of Equation 13: an increase in inter-movement intervals should reduce movement vigor (produce slower movements), whereas a decrease in inter-movement intervals should increase movement vigor (see also Results, Sensitivity to characteristics of the discount function).

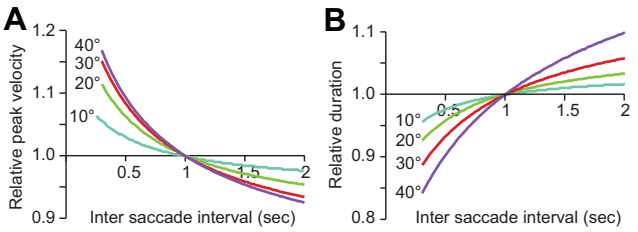

Figure 2. Changes in saccade vigor associated with changes in inter-movement interval, as predicted by a rate of reward cost function (Eq. 13). ***A***, Changes in peak velocity. For each ITI, and each amplitude, we computed the saccade duration that maximized rate of reward. We then computed the peak velocity of that saccade and normalized it with respect to the peak velocity at ITI of 1 s. The simulations show how much the peak velocity should increase (or decrease) as a function of ITI between saccades. The effect is greatest for saccade size of around 40°. ***B***, Changes in saccade duration as a function of changes in ITI.

To illustrate the predictions of Equation 13, we performed a simulation to determine how much saccade peak velocities and durations should change as we alter the ITI. We found that with respect to ITI of 1 s, reducing the ITI predicted increased peak velocities and increasing the ITI predicted decreased peak velocities (Fig. 2*A*). Importantly, the effect was asymmetric: a 0.5 s decrease in ITI predicted a much greater change in peak velocities than a 0.5 s increase in ITI. Furthermore, the effect of ITI on peak velocities grew with saccade amplitude, but tended to saturate at around 40°. Similarly, reducing the ITI predicted decreased saccade durations and increasing ITI predicted increased saccade durations (Fig. 2*B*). Put simply, if the brain is producing motor commands to maximize rate of reward (Eq. 13), then reducing the inter-movement intervals should increase movement vigor.

Change in ITI alters saccade vigor

We performed an experiment to test the prediction that changes in ITI should produce changes in saccade vigor. Subjects made alternate leftward and rightward saccades of 40° amplitude (Fig. 3*A*). We employed three block types, each consisting of 80 trials in which ITI increased, decreased, or remained constant (Fig. 3*B*). Each block began and ended with 10 trials having an ITI of 1 s. Figure 3, *C* and *E*, shows our theoretical predictions regarding saccadic vigor. As ITI is reduced from 1 s to 0.4 s, peak velocities (for a 40° saccade) should increase by approximately 10%. As ITI is increased from 1 s to 1.6 s, peak velocities should decrease by approximately 5%. The experimental results are shown in Figure 3, *D* and *F*. Over the first 10 saccades, the ITI was the same (1 s) in all block types, and the saccade peak velocities did not differ significantly across blocks. For the constant ITI block (black line) there was no clear change in peak velocity other than a trend for the peak velocity to decrease - a 'fatigue-like' effect thought to be associated with stimulus devaluation due to repetition of the stimulus (Chen-Harris et al., 2008). For the increasing ITI block (blue line), the abrupt decrease in ITI from 1 to 0.4 s on trial 10 (bin 5) was accompanied by a sharp increase in peak velocity. Over the next 60 trials the ITI was varied linearly from 0.4 s up to 1.6 s. This was associated with a steady decrease in peak velocity.

In the last 10 trials of the block, after the ITI was decreased abruptly from 1.6 s back to 1 s, the peak velocity began to increase again, becoming similar to the constant ITI block. Saccade peak velocities in the decreasing ITI block (red line) showed the opposite trend.

Analysis of covariance on each block confirmed that the changes in peak velocity followed significantly different trends across block types (ANCOVA, BLOCK × TRIAL interaction, $F_{(2,1074)} = 296.6$, $p < 10^{-10}$). Similarly, we saw a significant effect of ITI on movement duration (BLOCK × TRIAL interaction $F_{(2,1074)} = 111.2, p < 10^{-10}$; Fig. 3*F*). *Post hoc* comparisons of the saccades following the initial abrupt ITI change and control block showed that the effects were significant (paired *t* test, velocity: ITI increase vs control, $p = 0.005$; ITI decrease vs control, $p < 0.01$; duration: ITI increase vs control, $p = 0.007$, ITI decrease vs control, $p = 0.018$). However, manipulation of ITI did not affect saccade amplitude (BLOCK × TRIAL interaction $F_{(2,1074)} = 2.71$, $p > 0.05$), as shown in Figure 3*G*. In addition, we observed significant changes in reaction time (BLOCK × TRIAL interaction $F_{(2,1074)} = 50.9$, $p < 10^{-10}$), with reaction times becoming shorter as ITI decreased (Fig. 3*H*). Therefore, reductions in ITI generally produced saccades with faster velocities, shorter reaction times, and shorter durations.

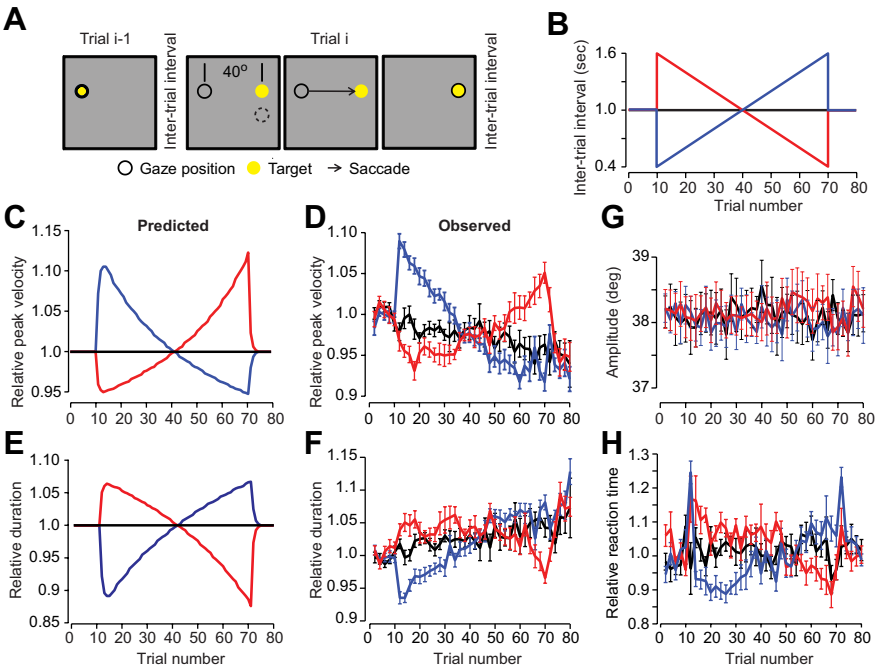

Figure 3. Experimental protocol, model predictions, and results. ***A***, Experimental protocol. Subjects were asked to make alternate leftward and rightward saccades to one of two possible targets. ***B***, The duration between the end of one saccade and displaying the "go" cue for the next saccade (the intertrial interval, ITI) was varied through the course of each 80 trial block. Three possible block types: increasing ITI (blue), decreasing ITI (red), and constant ITI (black). ***C***, Predictions of the rate of reward model regarding changes in peak velocity of saccades. ***D***, Changes in peak velocity with respect to the first 10 saccades for each block type. ***E***, Predictions of the rate of reward model regarding changes in saccade durations. ***F***, Changes in saccade duration. ***G***, Saccade amplitudes. ***H***, Changes in reaction times. Error bars indicate SEM.

Increases in the value of a visual stimulus results in saccades with shorter reaction time in both monkeys (Watanabe and Hikosaka, 2005; Bendiksby and Platt, 2006) and humans (Milstein and Dorris, 2007). We noted that whereas reaction time generally followed the same trends as velocity and duration, in one instance these measures could be dissociated. At the onset of the 11th trial (and the 71st trial) the ITI sharply changed, either increasing or decreasing. We observed an increase in velocity for ITI decrease, and a decrease in velocity for ITI increase (Fig. 3*D*,*F*). In contrast, both the sudden increase and the sudden decrease in ITI produced an increased reaction time. If we view reaction time as a period in which the upcoming movement is planned, this result suggests that the sudden change in ITI resulted in significantly longer time to plan the upcoming movement. Following this increased planning period, there was either a sharp decline (in case of increased ITI) or a sharp increase (in case of reduce ITI) in the vigor of the upcoming saccade. This dissociation allows us to rule out the possibility that changes in movement vigor were directly caused by changes to the reaction time that affected the process of movement planning.

It is noteworthy that for the block type with an initial decrease in ITI, saccadic vigor sharply changed within two trials of this decrease (e.g., maximum saccade velocity was reached within two trials). However, the rate of change in saccade vigor following an increase in ITI was much less, with subjects reaching minimum velocity after 7–8 trials. If we view changes in ITI as change in reward rate, then an unexpected change in ITI is equivalent to a reward rate prediction error. The fact that it takes longer for vigor to decrease than increase may be attributable to differences in learning from positive and negative reward rate prediction errors, suggesting that in this task the brain learns more from positive prediction errors than negative prediction errors.

Control experiment

An alternate interpretation of our experimental results is that when we reduced ITI, we are reducing the time that we are allowing the subject to view the target (in the experimental setup of Fig. 3*A*, the current target disappears when the new target is shown). Perhaps this reduced viewing time is influencing vigor by encouraging the subject to get to the new target earlier so that they can view it for a longer period of time before it disappears. To test for this, we performed a new experiment (Fig. 4*A*). In this version of the task the target that the subject was viewing did not disappear when a new target was presented. Rather, the subject could choose to continue viewing the current target for as long as they wanted. In this way, the viewing time of the current target was chosen by the subject, and not by the experimenter. Remarkably, we still observed that changing ITI produced robust changes in saccade vigor (Fig. 4*B*,*C*). The patterns of change in velocities and durations were essentially identical to that which we had recorded in the main experiment (velocity, ANCOVA, BLOCK × TRIAL interaction, $F_{(2,874)} = 301.1$, $p < 10^{-10}$; duration, BLOCK × TRIAL interaction, $F_{(2,874)} = 106.7$, $p < 10^{-10}$). We did also observe a significant effect of amplitude (BLOCK × TRIAL interaction, $F_{(2,894)} = 19.9$, $p < 10^{-8}$). However the changes in amplitude were of the order of 1% and are not sufficient to account for the changes in peak velocity and duration we observe, which were an order of magnitude larger than would be predicted on the basis of observed changes in amplitude alone.

Interestingly, reaction time in this task was markedly higher than in our main experiment (mean reaction time = 213 ms for the control experiment versus 158 ms for the main experiment). Despite this

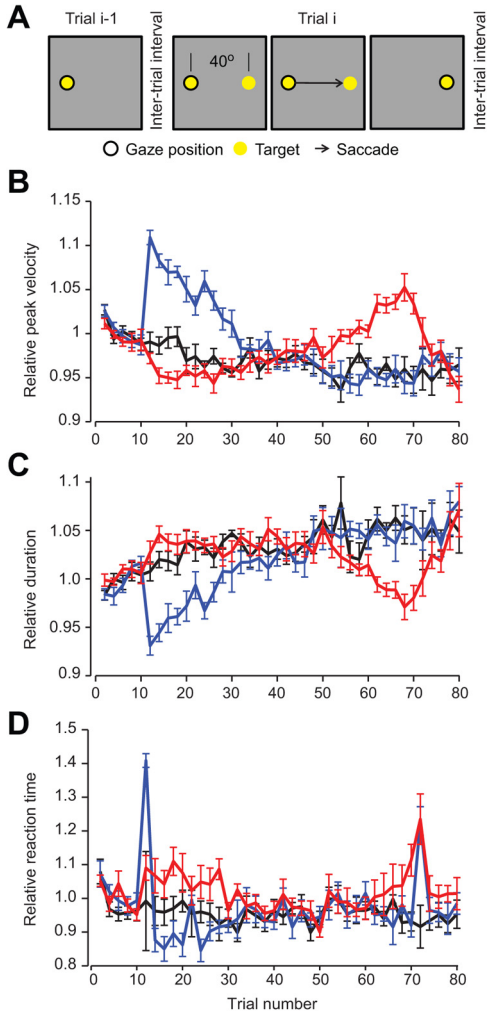

Figure 4. Control experiment. ***A***, Experimental protocol. This experiment was similar to that shown in Figure 3*A*, except that the previous target (current point of fixation) was not extinguished until after the saccade to the next target had begun. In this way, subjects could linger on the current target as long as they wanted. ***B***, Changes in peak velocity with respect to the first 10 saccades for each ITI block type (as in Fig. 3*B*). ***C***, Changes in saccade duration. ***D***, Changes in reaction time. Error bars are SEM.

marked difference in mean reaction time, the patterns of change (Fig. 4*D*) due to changes in ITI were qualitatively similar to the main experiment: the sudden change in ITI on the 11th trial produced an increase in reaction time, regardless of whether ITI was decreased or increased. Following sudden change, gradually reducing the ITI produced longer reaction times and increasing the ITI produced shorter reaction times. Although there was no statistically significant effect of ITI on reaction time (BLOCK × TRIAL interaction, $F_{(2,874)} = 1.24$, $p = 0.3$), this was largely caused by a single trial early in the block when an abrupt change in ITI caused unusually high reaction times. Overall, these observations are very similar to those in our main experiment. Therefore, the changes in vigor were unlikely to be due to subjects feeling rushed by the increased pace of the experiment.

Characteristics of the temporal discount function

We noted earlier that durations of saccades of different amplitudes could be accounted for by both hyperbolic and exponential temporal discount functions (Fig. 1*C*). However, we found that saccade durations not only depend on saccade amplitude, but also on the time since the last saccade, i.e., ITI (Fig. 3*F*). This experimental result confirms a prediction that we derived based on the premise of rate of reward, providing a rationale for hyperbolic temporal discounting (as in Eq. 13). However, let us now ask a more general question: in principle, what kinds of temporal discount functions could account for the data in Figure 3? For example, could exponential discounting account for these data?

In general we can imagine discount functions in which ITI combines additively with movement duration:

$$E[\text{reward}|\tau,\delta] = aP[\text{success}|\tau]F(\tau + \delta). \quad (14)$$

Figure 5 illustrates the influence of ITI on movement duration under a variety of temporal discount functions. Suppose that for some class of movements the probability of success increases with movement duration (i.e., the slower the movement, the more accurate), as displayed in Figure 5*A*. If the objective is to maximize rate of reward, then time carries a specific cost in which the probability of success is multiplicatively penalized by a hyperbolic temporal discount function: $F(\tau + \delta) = \frac{1}{\tau + \delta}$. Increasing the inter-movement interval δ shifts the temporal discount function to the left, altering its slope and shifting the peak of the discounted reward function to long duration movements. As a result, for hyperbolic discounting, an increase in ITI reduces the vigor of movements.

Now instead consider exponential discounting: $F(\tau + \delta) = \exp(-k(\tau + \delta))$. If reward is discounted exponentially, changing δ has no effect on the optimal duration because it simply leads to an overall scaling of the expected discounted reward (Fig. 5*B*). Therefore, the fact that we observed changes in saccadic vigor due to changes in ITI rejects the hypothesis that reward is discounted exponentially.

There are of course other plausible forms of temporal discounting, such as exponentials with squared exponents $F(\tau + \delta) = \exp(-k(\tau + \delta)^2)$. These forms imply that the cost of time is fairly constant for short durations, but durations that are longer carry increasingly greater cost. Such forms can also be dissociated from hyperbolic discounting, as they predict a sensitivity to ITI opposite that of hyperbolic discounting. In this case, an increase in ITI leads to an increase in movement vigor (Fig. 5*C*), which is inconsistent with our experimental results.

Therefore, the fact that we observed reduced saccadic vigor with increased ITI implies that temporal discounting has a specific shape. What is the class of functions that in principle could account for our data? The objective function to be maximized is

$$J(\tau) = P(\text{success}|\tau)F(\tau + \delta). \quad (15)$$

Suppose that, for a given value of δ, the optimal movement duration is τ_1. This implies that the gradient of J at τ_1 is zero, i.e.,

$$\left.\frac{\partial J}{\partial \tau}\right|_{\tau_1,\delta} = J' = P'F + PF' = 0. \quad (16)$$

Suppose we now increase δ to some new value. The resulting change in the gradient J' at τ_1 reveals how the optimal duration will change. If the gradient becomes positive, this means that the peak of J must have shifted to a larger value of τ. Likewise, a negative gradient implies a decrease in the optimal duration. In other words, the optimum movement duration will increase with inter-movement interval if the gradient of J' with respect to δ is positive:

$$\frac{\partial J'}{\partial \delta} = P'F' + PF'' > 0. \quad (17)$$

We arrived at Equation 17 by noting that $\frac{\partial^2 F}{\partial \tau \partial \delta} \equiv \frac{\partial^2 F}{\partial \tau^2}$ because τ and δ appear additively in F, and that $\frac{\partial P}{\partial \delta} = 0$ (because the

probability of success for a given movement duration is independent of the intertrial interval). We can eliminate P from Equation 17 by dividing through by P, which is strictly positive, and substituting $\frac{P'}{P} = -\frac{F'}{F}$ (which follows from Eq. 16). If we further multiply through by F (which is also strictly positive), we obtain the following condition:

$$FF'' > (F')^2. \qquad (18)$$

Importantly, this condition on the discount function is independent of the probability of success P and thus independent of the particular plant model and control policy we assume (since these will affect P but not F). Any temporal discount function F that satisfies Equation 18 will lead to the prediction that movement duration will increase (i.e., movement vigor will decrease) as inter-movement interval increases. For example, hyperbolic temporal discount functions (Eq. 13) have the property of Equation 18, as do sums of exponentials. For exponential discounting with linear exponents (Eq. 2) the left and right hand sides of Equation 18 are equal. For exponential discounting with quadratic exponents (Fig. 5*C*), the inequality is reversed.

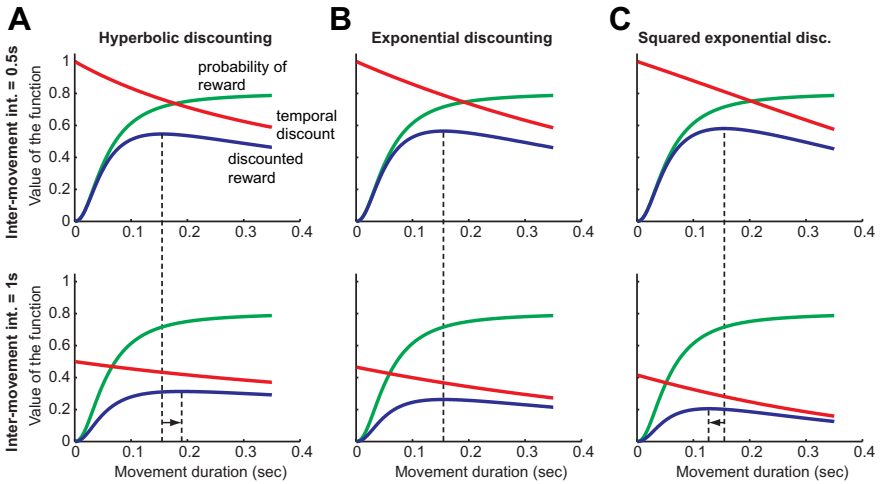

Figure 5. Effect of inter-movement interval on movement duration under various temporal discounting regimes. ***A***, Top, We consider an arbitrary class of movements for which probability of success (acquisition of reward) increases with movement duration (green line). Hyperbolic temporal discounting, plotted here by the red line, is the function $\frac{1}{\tau + \delta}$. τ is movement duration, and δ is inter-movement interval (here assumed to be 0.5 s). The blue line is the multiplication of probability of reward with the temporal discount function (Eq. 1). The movement duration that maximizes the discounted reward is noted by the dashed line. In this case, the discounted reward corresponds to reward rate. ***B***, ***C***, Top, Corresponding plots for an exponential discount function with linear exponents $\exp(-k(\tau + \delta))$, and an exponential discount function with squared exponents $\exp(-k(\tau + \delta)^2)$. All discount functions are scaled to be equal 1 at 0.5 s and parameters for the exponential discount functions were adjusted to predict the same optimal movement duration as rate of reward for $\delta = 0.5$ s. Bottom, The effect of increasing the inter-movement interval δ to 1 s. For hyperbolic discounting, as this delay is increased the optimum movement duration becomes longer, i.e., the movement vigor decreases. For an exponential temporal discount function with linear exponents there is no change in the optimum movement duration as inter-movement intervals are changed. For an exponential discount function with quadratic exponents, movement duration decreases as the inter-movement interval increases.

In summary, our experimental observations imply that control of saccades relies on a temporal discount function that satisfies Equation 18. The dependence of saccade vigor on intertrial interval emerges naturally in this mathematical framework and is unexplained by any previous model that we are aware of.

Discussion

We have proposed a new framework for control of movements. Our theory is founded on the principle of rate of reward—an idea that was previously invoked to explain aspects of decision-making in primates (Gold and Shadlen, 2002) and response intensities of rodents in free-operant tasks (Niv et al., 2007). We have shown that a rate of reward principle not only provides an explanation for vigor of saccades of varying amplitudes but also generates the novel prediction that vigor should be modulated by changes in ITI. This is in contrast to previous models which assume that motor commands that guide a movement are independent of movements that occurred previously. Our experiments confirm our predictions with remarkable precision: as the ITI changed, so did saccadic vigor. For example, we observed an increase in saccade peak velocity of 9.0 ± 0.9% (mean ± SD), compared with our theoretical prediction of 10%, and a decrease of −7.9 ± 1.1%, compared with our theoretical prediction of −5%. Changes in durations were in the predicted direction but somewhat smaller in magnitude, particularly for decreases in ITI. We are unaware of any previous motor control model that can explain such changes.

Our idea that time carries a cost in control of movements may explain a number of curious findings: (1) When a stimulus moves toward the fovea, saccades take longer to initiate than when the stimulus is moving away from the fovea (Segraves et al., 1987). Thus, subjects are not willing to wait for the stimulus to reach the fovea, highlight the idea that waiting even a few hundred milliseconds carries a cost. (2) An effective way to train monkeys to slow down their reaching movements is to impose a time penalty for overly fast movements (Churchland et al., 2006). The effectiveness of such a training protocol clearly illustrates the importance of the cost of time and is easily explained through a rate of reward framework, but is difficult to reconcile with models in which only the duration of the current movement is important. (3) The idea that changes in the available rate of reward can affect movement vigor is supported by a study by Ljungberg et al. (1992) who reported that in non-human primates, reaching movements made to collect a food reward were significantly slower when such rewards were available infrequently compared with when they were available frequently. This result could be considered analogous to our finding that increasing ITI decreases the vigor of saccades.

A fundamental question in neuroscience is why movements have characteristic kinematics. Why not move faster or slower? To approach this problem, previous works have sought to minimize weighted sums of rather ad hoc penalties for accuracy (quadratic distance to target) or time (linear or hyperbolic). We have suggested a different approach here: the decision regarding vigor of a movement depends on the probability of success of that movement multiplied by a function that represents temporal discounting of the value of the reward associated with that movement. These two approaches can in fact be related by applying a logarithmic transformation to the discounted reward (Eq. 14). This gives rise to an additive cost function that closely resembles those used in previous theories (Shadmehr et al., 2010), but with the quadratic accuracy term replaced by $-\log(P[\text{success}|\tau])$ and

the time penalty term replaced by the logarithm of the discount function. Notably, if the discount function is exponential, the logarithmic transformation gives rise to a linear time cost. Thus we can interpret previous models that have employed linear time costs (Harris and Wolpert, 2006) as tacitly assuming an exponential discounting of reward.

One aspect of the data which we could not explain through our model in its present form is the fact that saccade vigor tends to decline over the course of a block, even if the intertrial interval remains constant. This decline in vigor is not due to muscular fatigue (Prsa et al., 2010), but instead is likely due to the fact that repetition devalues the stimulus (Xu-Wilson et al., 2009). In monkeys, saccades to targets that predict a juice reward are significantly more vigorous than unrewarded targets (Takikawa et al., 2002). Saccades that accompany reaching to a stimulus are also more vigorous than saccades without the reach (Snyder et al., 2002; van Donkelaar et al., 2004). An increased probability of reward also increases the vigor of wrist movements made by monkeys to acquire reward (Opris et al., 2011). Our previous formulation (Shadmehr et al., 2010) was able to explain this dependence of movement vigor on reward value by including an effort penalty that was independent of reward value and a time penalty that scaled with reward value. In our present formulation, we did not include such an effort cost as we found that it was not necessary to generate strong predictions about behavior in the tasks we considered. Expanding our framework to include an effort penalty in the net reward, before applying the temporal discount factor, could enable us to explain these aspects of behavior through the rate of reward framework.

In addition to changes in movement speed, we also observed changes in reaction time. Although a low reaction time can be associated with a general increase in vigor, a quantitative prediction of this effect is beyond the scope of our model. One way in which we might be able to account for such an effect would be to view reaction time as a period in which a decision must be made about the goal of the upcoming saccade. Computational models have described such decision-making in terms of a stochastic accumulation of evidence until a pre-determined threshold is reached, at which point the decision is made and an action triggered. Modulation of reaction time could therefore be interpreted as a change in the height of this threshold. It has been suggested that the level of the threshold might itself be set based on a rate of reward principle (Gold and Shadlen, 2002). Indeed, experiments directly analogous to our own that vary ITI during perceptual discrimination tasks find that decreases in ITI lead to faster, less accurate decisions (Cisek et al., 2009; Simen et al., 2009). Similarly, decreasing ITI causes monkeys to adopt a more risky policy when gambling for a juice reward (Hayden and Platt, 2007).

In reinforcement learning, future rewards are typically discounted exponentially, largely due to mathematical convenience. Infinite horizon control problems, however, are commonly formulated in terms of minimizing the average-cost per stage, exactly analogous to the rate of reward cost function we proposed here. The average cost per stage framework has been invoked to model behavior when animals face a sequence of choices between discrete actions (Daw, 2003). This idea has even been extended to include a basic notion of movement vigor, thereby offering an explanation for differences in the intensity of a rat's free-operant responses (such as lever press frequency) across different motivational states (Niv et al., 2007). That work posited a conceptual link between dopamine, rate of reward, and response vigor through the idea that tonic activity of dopamine neurons encodes the background average rate of reward. When rewards are plentiful, it is worthwhile choosing the more costly (either energetically or in terms of risk) policy and acting more vigorously. In Parkinson's disease, reduced tonic dopamine leads to less vigorous actions (Mazzoni et al., 2007), consistent with the idea that tonic dopamine encodes a rate of reward. Phasic activity of dopamine neurons may also be linked to rate of reward. Stimuli that predict reward at various delays elicit phasic responses in dopamine neurons that decline hyperbolically with the delay duration (Kobayashi and Schultz, 2008). Phasic dopamine activity may thus reflect a reward rate prediction error, rather than an error in total predicted reward.

If we view ITI as a factor that alters rate of reward, then a sudden change in ITI introduces a reward rate prediction error. When ITI is reduced, the prediction error is positive, implying that the brain is receiving a greater amount of reward than anticipated. Similarly, when ITI is increased, the prediction error is negative. We consistently observed that the change in vigor was faster when the prediction error was positive compared with negative. This implies that there may be a differential sensitivity to positive and negative reward prediction errors in the population we sampled—an idea that is consistent with basal ganglia neurophysiology (Maia and Frank, 2011). Viewed in this way, our experiment may provide a way to assess reward-dependent learning—namely, monitoring changes in behavior in response to altered rate of reward.

The fact that a single optimization principle seems to be shared by such a broad variety of tasks suggests that it may offer a unifying normative view of temporal discounting in decision-making and motor control. In effect, the long term behavioral goal of reward rate optimization is achieved through the short-term mechanism of temporal discounting. Hyperbolic discounting of reward may therefore be an obligatory phenomenon that has evolved because it tends to optimize reward rate in most ecologically relevant scenarios.

References

Bendiksby MS, Platt ML (2006) Neural correlates of reward and attention in macaque area LIP. Neuropsychologia 44:2411–2420.

Chen-Harris H, Joiner WM, Ethier V, Zee DS, Shadmehr R (2008) Adaptive control of saccades via internal feedback. J Neurosci 28:2804–2813.

Churchland MM, Santhanam G, Shenoy KV (2006) Preparatory activity in premotor and motor cortex reflects the speed of the upcoming reach. J Neurophysiol 96:3130–3146.

Cisek P, Puskas GA, El-Murr S (2009) Decisions in changing conditions: the urgency-gating model. J Neurosci 29:11560–11571.

Collewijn H, Erkelens CJ, Steinman RM (1988) Binocular co-ordination of human horizontal saccadic eye movements. J Physiol 404:157–182.

Daw ND (2003) Reinforcement learning models of the dopamine system and their behavioral implications. Ph.D. thesis. Carnegie Mellon University.

Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47:381–391.

Gold JI, Shadlen MN (2002) Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. Neuron 36:299–308.

Harris CM, Wolpert DM (1998) Signal-dependent noise determines motor planning. Nature 394:780–784.

Harris CM, Wolpert DM (2006) The main sequence of saccades optimizes speed-accuracy trade-off. Biol Cybern 95:21–29.

Hayden BY, Platt ML (2007) Temporal discounting predicts risk sensitivity in rhesus macaques. Curr Biol 17:49–53.

Kacelnik A (1997) Normative and descriptive models of decision making: time discounting and risk sensitivity. In: Characterizing human psychological adaptations (Bock G, Cardew G, eds), pp 51–70. Chichester: Wiley.

Kobayashi S, Schultz W (2008) Influence of reward delays on responses of dopamine neurons. J Neurosci 28:7837–7846.

Ljungberg T, Apicella P, Schultz W (1992) Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67:145–163.

Maia TV, Frank MJ (2011) From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 14:154–162.

Mazzoni P, Hristova A, Krakauer JW (2007) Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 27:7105–7116.

Milstein DM, Dorris MC (2007) The influence of expected value on saccadic preparation. J Neurosci 27:4810–4818.

Myerson J, Green L (1995) Discounting of delayed rewards: models of individual choice. J Exp Anal Behav 64:263–276.

Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl) 191: 507–520.

Opris I, Lebedev M, Nelson RJ (2011) Motor planning under unpredictable reward: modulations of movement vigor and primate striatum activity. Front Neurosci 5:61.

Prsa M, Dicke PW, Thier P (2010) The absence of eye muscle fatigue indicates that the nervous system compensates for non-motor disturbances of oculomotor function. J Neurosci 30:15834–15842.

Robinson DA (1986) The systems approach to the oculomotor system. Vision Res 26:91–99.

Schmidt RA, Zelaznik H, Hawkins B, Frank JS, Quinn JT Jr (1979) Motor-output variability: a theory for the accuracy of rapid motor acts. Psychol Rev 47:415–451.

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

Schweighofer N, Shishida K, Han CE, Okamoto Y, Tanaka SC, Yamawaki S, Doya K (2006) Humans can adopt optimal discounting strategy under real-time constraints. PLoS Comput Biol 2:e152.

Segraves MA, Goldberg ME, Deng SY, Bruce CJ, Ungerleider LG, Mishkin M (1987) The role of striate cortex in the guidance of eye movements in the monkey. J Neurosci 7:3040–3058.

Shadmehr R, Orban de Xivry JJ, Xu-Wilson M, Shih TY (2010) Temporal discounting of reward and the cost of time in motor control. J Neurosci 30:10507–10516.

Simen P, Contreras D, Buck C, Hu P, Holmes P, Cohen JD (2009) Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. J Exp Psychol Hum Percept Perform 35:1865–1897.

Snyder LH, Calton JL, Dickinson AR, Lawrence BM (2002) Eye-hand coordination: saccades are faster when accompanied by a coordinated arm movement. J Neurophysiol 87:2279–2286.

Sutton RS, Barto AG (1981) Toward a modern theory of adaptive networks: expectation and prediction. Psychol Rev 88:135–170.

Takikawa Y, Kawagoe R, Itoh H, Nakahara H, Hikosaka O (2002) Modulation of saccadic eye movements by predicted reward outcome. Exp Brain Res 142:284–291.

van Beers RJ (2007) The sources of variability in saccadic eye movements. J Neurosci 27:8757–8770.

van Beers RJ (2008) Saccadic eye movements minimize the consequences of motor noise. PLoS One 3:e2070.

van Donkelaar P, Siu KC, Walterschied J (2004) Saccadic output is influenced by limb kinetics during eye-hand coordination. J Mot Behav 36:245–252.

Watanabe K, Hikosaka O (2005) Immediate changes in anticipatory activity of caudate neurons associated with reversal of position-reward contingency. J Neurophysiol 94:1879–1887.

Xu-Wilson M, Zee DS, Shadmehr R (2009) The intrinsic value of visual information affects saccade velocities. Exp Brain Res 196:475–481.