11 Cost of time in motor control

Perhaps in no area of neuroscience is the cooperation between theory and data more fruitful than in the study of movements. The reason has much to do with the fact that movements of people and other animals tend to be highly reproducible. For example, reaching movements, eye movements, and walking each have distinct features that are present in all healthy people. A teenager may actively choose to dress very differently than his parents, but will not exhibit walking, reaching, or eye movements that are very different. Indeed, a persistent difference in movements is often a sign of a neurological disorder. Our aim here is to describe some of this regularity, and attempt to build a theory that may account for it.

Over the last 25 years, a large body of experimental and computational work has been directed towards offering theories that attempt to explain why there is regularity in our movements. The approaches are reminiscent of physics and its earliest attempts to explain regularity in motion of celestial objects. Similar to physics, the search in the field of motor control has been for normative laws, or cost functions, that govern our behavior. Here, we will focus on a theory that assumes that the purpose of movements is to achieve a rewarding state at a minimum effort. That is, we are going to assume that we make movements in order to change the state of our body to something more valuable, and effort is the cost that we pay to acquire reward.

Consider the behavior of the eyes during saccadic eye movements: when we are looking at an image, the brain shifts our gaze from location to location with saccades. During each saccade, the eyes move in a stereotypical trajectory, with temporal and spatial characteristics that are described in Fig. 10.3. These characteristics include a nearly linear relationship between saccade amplitude and duration (Fig. 10.3B), and a symmetric trajectory of eye velocities for small amplitude saccades but asymmetric velocities during large amplitude saccades (Fig. 10.3D). However, these characteristics are variable and the variability is not entirely due to random noise. For example, when we make a saccade to look at a face, the eyes move slightly faster and arrive at the target slightly sooner than if the goal of our saccade was to foveate another kind of image (Fig. 10.7). When monkeys are rewarded with juice for moving their eyes to a location, their saccade has a shorter duration and higher peak speed than when the same location is unrewarded (Fig. 10.4). Therefore, while saccades show certain patterns in their temporal and spatial properties, the variability in these patterns appears to have something to do with the value of the stimulus that affords the movement. Possibly, the expected reward alters the motor commands that the brain programs to make a movement. In particular, movements seem to be faster (higher speed) and have a shorter reaction time (Fig. 10.4C) when the value of the stimulus is higher.

These results hint that the subjective value of a state that we wish to attain by our movements somehow discounts the motor commands, making the movement faster and with a shorter duration. In a sense, by spending a greater effort, the brain produces motor commands that attain a valuable stimulus sooner. Time, even in the range of milliseconds seems to matter because it provides the reward sooner. These ideas lead us toward a cost for movements: it seems that reward is something that is good and we should get as much of it as possible, but effort is something that is bad and we should spend as little of it as possible. A reward that is attained sooner is better than the same reward if it is attained later. By mathematically describing these ideas, a framework will be constructed that can account for the specific saccade durations and velocities of healthy humans. More interestingly, we will be able to consider changes in the reward system of the brain due to development, aging, and disease, and try and make sense of changes that occur in control of movements.

11.1 Temporal discounting of reward

Economists and psychologists have noted that humans (and other animals) tend to discount reward as a function of time. An example of this is an experiment by Joel Myerson and Leonard Green (Myerson and Green, 1995) in which undergraduate students were asked to make a series of hypothetical decisions between monetary rewards. They were shown one card that specified how much they would be paid now, and another card that specified a larger amount that they would be paid if they waited some amount of time. Based on the choice that each subject made, the authors constructed a function that estimated how value of the reward declined as a function of time (in this case, years). For example, say that a subject picks at roughly 50% probability between \$350 now and \$1000 at 5 years. This would suggest that for this person, \$350 now is equal to \$1000 in 5 years, which would imply that \$1000 loses its value as a function of time. Myerson and Green considered two kinds of functions to describe this temporal discounting of value: hyperbolic and exponential.

$$V_e(p) = \alpha \exp(-\beta p)$$

$$V_h(p) = \frac{\alpha}{1+\beta p}$$

The variable p represents the time delay to reward acquisition. The exponential function is one that is typically used in economics and is called a discounted utility model. It assumes that future rewards are discounted because of the risk involved in waiting for them. The hyperbolic function is one that has been used by psychologists, and its rationale is based on empirical grounds: it fits the data better (Fig. 11.1A). (However, as we will see later, there may be a good reason for why animals use hyperbolic discount functions, as they may be trying to maximize reward per unit of time.) In particular, as can be seen in Fig. 11.1A, exponential discounting underestimates the value of reward at longer time delays. People generally do not discount future reward as steeply as is predicted by an exponential --- the hyperbolic form of the discount function fits the data better.

Interestingly, the basic form of the temporal discount function remains hyperbolic when we go from time delay of years to time delay of seconds. Koji Jimura and colleagues (Jimura et al., 2009) asked undergraduates to not drink anything for a few hours and then gave them a choice between a small amount of water now, and a larger amount some seconds later. They found that the temporal discount function was a hyperbola, with a rate $\beta = 0.09$ per second (Fig. 11.1B). Shunsuke Kobayashi and Wolfram Schultz (2008) trained thirsty monkeys to choose between a small amount of juice now and a greater amount some seconds later and found a temporal discount rate of $\beta = 0.31$ per second for one monkey and $\beta = 0.17$ per second for another monkey (Fig. 11.2A). The larger value of β indicates that the subject is more *impulsive*, i.e., would rather take a smaller amount of reward now than wait for the larger amount. Kenway Louie and Paul Glimcher (2010) also trained thirsty monkeys to choose between a small amount of juice now and a greater amount some seconds later and found a temporal discount rate of $\beta = 0.16$ per second for one monkey and $\beta = 0.04$ per second for the another (Fig. 11.2C and 11.2D). Leonard Green and colleagues (Green et al., 2004) trained pigeons to choose between a small amount of food now or a greater amount a few seconds later and found a temporal discount rate of $\beta = 0.62$ per second. Although it is difficult to directly compare these results because the experiments are somewhat different, there are two basic ideas that emerge from this vast literature: 1) animals tend to discount reward hyperbolically as a function of time, and 2) while there are inter-subject differences within specie (e.g., some monkeys are more impulsive than others), there are also large differences between species, with pigeons generally more impulsive than rhesus monkeys, and rhesus monkeys more impulsive than humans.

A hyperbolic function of time is not only a good fit to choices that people and other animals make regarding the temporal discounting of valuable commodities, it is also a good fit to discharge of dopamine cells in the brain of monkeys that have been trained to associate visual stimuli with delayed reward. Shunsuke Kobayashi and Wolfram Schultz (2008) trained two monkeys to watch a monitor upon which visual stimuli appeared. In a given trial, one stimulus was followed by a short delay and then a small amount of liquid. In another trial, another stimulus was followed by a longer delay and the same amount of liquid. They then recorded from dopaminergic cells in the substantia nigra (a part of the basal ganglia) and found that in response to the stimulus that predicted reward at short delay, discharge was high. In contrast, in response to the stimulus that predicted the same amount of reward after a long delay, discharge was low. This response declined approximately hyperbolically as a function of delay (Fig. 11.2B). Intriguingly, for the monkey that was more impulsive (monkey A in Fig. 11.2A), the dopamine function was slightly steeper (monkey A in Fig. 11.2B). Therefore, dopaminergic cells tend to give a short burst of discharge in response to a stimulus that predicts future reward, and the magnitude of this burst declines hyperbolically as a function of the expected delay to reward.

This hyperbolic encoding of future reward is also present in the response of neurons in the lateral intraprietal area (LIP), an area of the brain that mediates sensorimotor transformations for control of eye movements. Cells in this area respond to stimuli in a selective region of the visual space (called a receptive field), firing before a saccade to that location. Kenway Louie and Paul Glimcher (2010) trained monkeys in a task in which two stimuli would appear (red and green LED), and the animal had the choice of making a saccade to one of the stimuli. In a given block of trials, the red stimulus was associated with an immediate reward (small magnitude), and the green stimulus was associated with a delayed reward (larger magnitude). The animal would saccade to indicate the choice of which reward it wanted. By manipulating the magnitude of the delayed reward the authors constructed the behavioral temporal discount function for each monkey (Fig. 11.2C and 11.2D). Monkey C was more impulsive than monkey D. Next, the authors recorded from cells in LIP. When a stimulus appeared in a cell's receptive field, the cell responded with a firing rate that increased with the subjective stimulus value (Fig. 11.2C and 11.2D). That is, for the more impulsive monkey (monkey C), LIP cells tended to fire much less for a stimulus that predicted a 12s delayed reward as compared to an immediate reward. Therefore, the discharge in LIP before the saccade was a reflection of the subjective value of the upcoming movement. It appears that the brain encodes the economic value of a given stimulus that promises future reward as a hyperbolic function of time.

11.2 Hyperbolic vs. exponential discounting of reward

We only have a few data points in each subplot of Figs. 11.1 and 11.2, and they are often missing error bars. Why has previous work preferred to use hyperbolas to represent temporal discounting of reward vs. exponentials? The reason has to do with a basic property of decision making: change of mind.

Suppose that on Friday morning you are informed that there will be a test in one of your classes on Monday. Later that morning a friend mentions that they are planning to go out to dinner and dancing on Saturday evening and it would be great if you could come. You are faced with a choice: stay home on Saturday and study for the test, or go out with friends. Suppose that the value that you associate with passing your test is greater than the value that you assign to going out with friends. Indeed, on Friday morning you evaluate your two options and decide to stay home on Saturday. However, as Saturday evening approaches, you change your mind and decide to go out with friends. Let us show that hyperbolic discounting of reward can explain this behavior, whereas exponential discounting cannot.

In Fig. 11.4 we have plotted the temporal discounting of two rewarding events that will take place in the future. One event has a greater value but will take place later (passing your test), the other event has a smaller value but will take place sooner (going out with friends). The two rewarding events are discounted in time at the same rate. So all that matters is the value that you assign to the events at the time that they will occur. Consider a hyperbolic form of temporal discounting of these two events (Fig. 11.4A). At time zero (Friday morning) you evaluate your choices and pick the later event (passing your test, black line), as this has a higher current value. However, as you get closer to the time of the sooner event (going out with friends), its value becomes greater than the later event (gray line crosses the black line). On Saturday evening you change your mind and decide to go out with friends. However, if you discounted reward exponentially (Fig. 11.4B), the two functions would never cross and you would not change your mind. The fact that people and other animals do occasionally change their mind is inconsistent with exponential temporal discounting of reward.

11.3 A cost for movements

Let us now return to the problem of making movements. Suppose that the objective of any movement is to place our body in a state that is more rewarding. Further suppose that the value of the state that we wish to acquire declines hyperbolically with time, meaning that it is better to get to this valuable state sooner rather than later. But how much better is it to get there sooner? Well, that depends on our subjective value of the stimulus and our temporal discount function. However, in order to acquire this state we will have to spend some effort. Usually, the greatest effort will be associated with acquiring the rewarding state sooner (move quickly), and the smaller effort will be associated with acquiring the rewarding state later (move slowly). Putting these ideas together, we get a natural balance between effortful behavior and maximizing reward: the movement that we perform is one that maximizes reward while minimizing effort.

Here, we will construct a simple model of the eye's dynamics, produce a saccade that minimizes a cost, and explore the influence of temporal discounting on the movement kinematics. Our objective is to ask why eye trajectories have their specific regularities, and why a change in the value associated with the target of the movement might alter this trajectory. As we will see, there appears to be a link between how the brain temporally discounts reward, and the trajectory that the eye follows during a saccade. The basic idea is that the motor commands that move our body are a reflection of an economic decision regarding reward and effort.

Suppose that in order to make a movement the brain solves the following problem: generate motor commands to acquire as much reward as possible, while expending as little effort as possible. Suppose that at time t, the state of our eye is described by vector $\mathbf{x}(t)$ (representing position, velocity, etc.), our motor commands are $u(t)$, and our target is a stimulus at position g (with respect to the fovea). Our brain assigns some value α to the visual stimulus that happens to be the target of our saccade. The reward is acquired when the image of the target is on the fovea, which will require a movement that will take time. The crucial question is how passage of time should affect the reward value of the target. That is, how much more valuable is it for us to place the face on our fovea within 50ms vs. 100ms? Shadmehr, Jean-Jacque Orban de Xivry, Minnan Xu-Wilson, and Ting-Yu Shih (Shadmehr et al., 2010) proposed that the same temporal discount function that described devaluation of reward in timescales of years and seconds, may also describe this devaluation in timescale of milliseconds. That is, the motor system will incur a cost for delaying the acquisition of reward because of the time p that it takes to place the valuable image on the fovea:

$$J_p = \alpha\left(1 - \frac{1}{1+\beta p}\right) \tag{11.1}$$

Therefore, the longer it takes to get the target on the fovea, the larger the loss of reward value. More importantly, note that the cost depends on the initial value α that we assign to the stimulus, and the rate β with which we temporally discount this value. If the stimulus was interesting (e.g., face of an attractive person), we incur a greater cost for the duration of the movement than if the stimulus was uninteresting (e.g., an image made of random noise). Similarly, if one subject is more impulsive than another (i.e., a larger β), then the passage of a given amount of time will cost more for the impulsive subject.

In order to move the eyes, we will have to spend some effort in terms of motor commands. We have little information about how the brain represents motor costs, but in the previous chapter we saw some evidence that this cost is approximately a quadratic function of force. For example, Fagg et al. (2002) suggested that the width of the tuning function of muscles that act on the wrist could be accounted for if one assumes that the brain activates the muscles in such as way as to minimize a cost composed of the sum of the squared force contribution of each muscle (Fig. 10.11). Similarly, O'Sullivan et al. (2009) estimated that in a task in which fingers of the two hands cooperate to produce a goal force, the brain generated a force in each finger so to minimize the sum of squared forces (Fig. 10.14). Based on this admittedly limited body of evidence, the best that we can currently do is to define motor costs during a movement as the sum of squared muscle activations:

$$J_u = \lambda \int_0^p u^2(t)dt \tag{11.2}$$

The parameter λ specifies the relative cost that we incur for the motor commands. Finally, when our movement ends at time $t = p$, the target at position g should be on the fovea. This constitutes an accuracy cost, and it is mathematically convenient to represent it as a quadratic function:

$$J_x = \tau\left(x(p) - g\right)^2 \tag{11.3}$$

In Eq. (11.3), the term $x(p)$ represents the position of our eye at the end of the movement. The parameter τ specifies the relative cost that we incur for being inaccurate.

In summary, we assume that in performing a movement, the brain attempts to produce motor commands that minimize a cost composed of accuracy cost, effort cost, and temporal cost:

$$J = J_x + J_u + J_p \tag{11.4}$$

Our first objective is to ask whether a hyperbolic temporal cost can account for the kinematics (i.e., duration, velocity, etc.) of saccades. Our second objective is to ask whether this temporal cost is related to reward processing in the brain. These objectives require solving an optimal control problem in which Eq. (11.4) serves as a cost function. The crucial prediction of the theory is that there should be specific changes in saccade durations and velocities due to changes in the reward discounting function (Eq. 11.1), for example, due to changes in stimulus reward value α or temporal discounting rate β.

11.4 Optimal control of eye movements

In this section we will describe a very simply model of the human eye plant (i.e., dynamics of the eyes), and then find a set of motor commands that bring the eyes to the target while minimizing a cost (Eq. 11.4). Our objective will be to find the 'best movement' that can be performed in order to be as lazy as possible (least effort), while getting as much reward as possible (least devaluation).

In our one dimensional model of the eye in Fig. 11.4A, there are two elastic elements that pull the eye in each direction, and a viscous element that resists this motion. Suppose that the stiffness of each spring is $k/2$. The force in the bottom spring is $\frac{k}{2}x$ and the force in the top spring is $\frac{k}{2}\left(x_0 - x\right)$. The dynamics of this system is:

$$m\ddot{x} = -\frac{k}{2}x - b\dot{x} + f + \frac{k}{2}\left(x_0 - x\right)$$

which can be simplified to:

$$m\ddot{x} = -k\left(x - \frac{x_0}{2}\right) - b\dot{x} + f \tag{11.5}$$

If we redefine x so that we measure it from $\frac{x_0}{2}$, then the equivalent system is shown in Fig. 11.4B, where the equilibrium point of the spring is at $x = 0$. As a result, the dynamics of our system becomes:

$$m\ddot{x} = -kx - b\dot{x} + f \tag{11.6}$$

Solving for acceleration we have:

$$\ddot{x} = -\frac{k}{m}x - \frac{b}{m}\dot{x} + \frac{1}{m}f \tag{11.7}$$

The term f represents the active forces produced by the muscles. A very simple muscle model is one that transforms 'activations' u into force:

$$\alpha_1 \frac{df}{dt} + \alpha_2 f = u \tag{11.8}$$

We can re-write this third order dynamics as a sequence of first order differential equations. Let us define $x_1 \equiv x$, $x_2 \equiv \dot{x}$, and $x_3 \equiv f$. We have:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -\frac{k}{m} & -\frac{b}{m} & \frac{1}{m} \\ 0 & 0 & -\frac{\alpha_2}{\alpha_1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\alpha_1} \end{bmatrix} u \tag{11.9}$$

David Robinson and colleagues (Keller, 1973;Robinson et al., 1986) perturbed the human eyes, measured the mechanical response, and found that the system could be approximated via the linear system of Eq. (11.9) with time constants of 224, 13, and 4 ms. [In Eq. (11.9), if we set $k = 1$, then $b = \tau_1 + \tau_2$ and $m = \tau_1\tau_2$, where $\tau_1 = 0.224$ and $\tau_2 = 0.013$. If we set $\alpha_2 = 1$, then $\alpha_1 = 0.004$.] Our goal is to find the motor commands $u(t)$ that move the eyes so that we get as much reward as possible (i.e., minimize the temporal discounting of reward), while being as lazy and accurate as possible (i.e., minimize the effort and accuracy costs). To accomplish our goal, we will reformulate the dynamics from continuous to discrete time, and then solve the optimal control problem.

Relating discrete and continuous representations of linear systems can be readily done using Euler's approximation. Suppose the continuous representation is of the form:

$$\dot{\mathbf{x}}(t) = A_c\mathbf{x}(t) + \mathbf{b}_c u(t) \tag{11.10}$$

We have:

$$\dot{\mathbf{x}}(t) \approx \frac{\mathbf{x}(t+\Delta t) - \mathbf{x}(t)}{\Delta t} \tag{11.11}$$

From Eq. (11.11), we have:

$$\begin{aligned}\mathbf{x}(t+\Delta t) &\approx \dot{\mathbf{x}}(t)\Delta t + \mathbf{x}(t)\\ &\approx \left(A_c\mathbf{x}(t)+\mathbf{b}_c u(t)\right)\Delta t + \mathbf{x}(t)\\ &\approx \left(I + A_c\Delta t\right)\mathbf{x}(t) + \mathbf{b}_c u(t)\Delta t\end{aligned} \tag{11.12}$$

Therefore, the discrete and continuous representations are related as follows:

$$\begin{aligned}\mathbf{x}^{(k+1)} &= A_d\mathbf{x}^{(k)} + \mathbf{b}_d u^{(k)}\\ A_d &\approx \left(I + A_c\Delta t\right)\\ \mathbf{b}_d &\approx \mathbf{b}_c\Delta t\end{aligned} \tag{11.13}$$

[Eq. (11.13) describes an approximate relationship between discrete and continuous forms. One can also find an exact relationship. The technique that provides an exact solution relies on matrix exponentials. We have provided a tutorial on the subject in the web based supplementary materials that accompany this text.] Our objective is to find a sequence of motor commands $\mathbf{u}_h = \left[u^{(0)}, u^{(1)}, \cdots, u^{(p-1)}\right]^T$ to minimize a cost of the form:

$$J = \left(\mathbf{y}^{(p)} - \mathbf{r}\right)^T T\left(\mathbf{y}^{(p)} - \mathbf{r}\right) + \mathbf{u}_h^T L \mathbf{u}_h + \alpha\left(1 - \frac{1}{1+\beta p}\right) \tag{11.14}$$

where:

$$T \equiv \begin{pmatrix} v_1 & 0 & 0\\ 0 & v_2 & 0\\ 0 & 0 & v_3\end{pmatrix} \qquad \mathbf{r} \equiv \begin{bmatrix} g\\ 0\\ 0\end{bmatrix} \qquad L \equiv \begin{pmatrix} \lambda^{(0)} & 0 & 0 & 0\\ 0 & \lambda^{(1)} & 0 & 0\\ 0 & 0 & \ddots & 0\\ 0 & \cdots & 0 & \lambda^{(p-1)}\end{pmatrix}$$

and:

$$\begin{aligned}\mathbf{x}^{(k+1)} &= A\mathbf{x}^{(k)} + \mathbf{b}u^{(k)}\\ \mathbf{y}^{(k)} &= C\mathbf{x}^{(k)}\end{aligned} \tag{11.15}$$

In our simulations, we set the terms C and L to be the identity matrix. The first term in Eq. (11.14) enforces our desire to have endpoint accuracy. Because accuracy matters only at the end of the movement, we penalize the squared difference between the state of the eye at movement end and the goal state (the goal state includes stimulus position, as well as zero velocity and acceleration). The second term penalizes effort. The third term enforces our desire to get to the target as soon as possible in order to minimize the temporal discounting of reward. The terms α and β describe how reward is valued and discounted in time. As we will see, different populations (e.g., children vs. adults, monkeys vs. humans) discount reward in different ways. Furthermore, certain diseases (e.g., schizophrenia and Parkinson's disease) alter how the brain processes reward. Eq. (11.14) predicts that there should be a correlation between this discounting of reward and eye movements in these different populations.

Notice that our model of the eye plant has no noise in it. We will solve the optimal control problem for a system without noise, and then solve it again with noise. We will find that if the system has signal dependent noise, the optimum set of motor commands are quite different than if it has no noise. We will see that the trajectory of the modeled eye with signal dependent noise looks quite similar to actual eye movements, but without this noise, the trajectory is quite different than the actual eye movement.

To minimize the cost in Eq. (11.14), we will make a critical assumption: that the motor commands during a saccade do not have access to sensory feedback regarding state of the eye. Indeed, typical saccades are too brief for visual feedback to influence saccade trajectory (a movement ends after about 100ms). Furthermore, proprioceptive signals from the eyes do not play a significant role in controlling saccade trajectories (Keller and Robinson, 1971;Guthrie et al., 1983). Technically, this means that we have an open-loop optimal control problem that we are trying to solve.

In order to minimize the cost in Eq. (11.14), given the constraint in Eq. (11.15), we will divide our problem into two parts: first, we select an arbitrary length of time p and find the optimal set of motor commands $\mathbf{u}_h^*(p)$ that minimize Eq. (11.14). We repeat this for all possible p. Second, given that we produce the optimal motor commands $\mathbf{u}_h^*(p)$, we search the space of p for the one movement time that provides us with the minimum total cost J.

Step 1: the optimal motor commands for a given movement duration

It is useful to write the history of states and how they relate to the history of motor commands. We note that the state $\mathbf{x}$ at time step k is related to the history of motor commands as follows:

$$\begin{aligned}
\mathbf{x}^{(1)} &= A\mathbf{x}^{(0)} + \mathbf{b}u^{(0)} \\
\mathbf{x}^{(2)} &= A\mathbf{x}^{(1)} + \mathbf{b}u^{(1)} = A^2\mathbf{x}^{(0)} + A\mathbf{b}u^{(0)} + \mathbf{b}u^{(1)} \\
\mathbf{x}^{(3)} &= A\mathbf{x}^{(2)} + \mathbf{b}u^{(2)} = A^3\mathbf{x}^{(0)} + A^2\mathbf{b}u^{(0)} + A\mathbf{b}u^{(1)} + \mathbf{b}u^{(2)} \\
\mathbf{x}^{(k)} &= A^k\mathbf{x}^{(0)} + \sum_{j=0}^{k-1} A^{k-1-j}\mathbf{b}u^{(j)}
\end{aligned} \tag{11.16}$$

Therefore, the state at the end of our movement is:

$$\mathbf{x}^{(p)} = A^p \mathbf{x}^{(0)} + F\Gamma\left(\mathbf{u}_h + \boldsymbol{\varepsilon}_h\right) \tag{11.17}$$

where:

$$F \equiv \begin{bmatrix} A^{p-1} & A^{p-2} & A^{p-3} & \cdots & I \end{bmatrix} \qquad \Gamma \equiv \begin{bmatrix} \mathbf{b} & 0 & \cdots & 0 \\ 0 & \mathbf{b} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{b} \end{bmatrix} \tag{11.18}$$

From Eq. (11.14), we have:

$$J(p) = \mathbf{x}^{(p)T} C^T T C \mathbf{x}^{(p)} - 2\mathbf{x}^{(p)T} C^T T \mathbf{r} + \mathbf{r}^T T \mathbf{r} + \mathbf{u}_h^T L \mathbf{u}_h + \alpha\left(1 - \frac{1}{1+\beta p}\right) \tag{11.19}$$

Inserting Eq. (11.17) into above, we can compute the derivative of this cost with respect to the motor commands at a given p:

$$\begin{aligned} \frac{dJ}{d\mathbf{u}_h} &= 2\Gamma^T F^T C^T T C A^p \mathbf{x}^{(0)} - 2\Gamma^T F^T C^T T \mathbf{r}_h \\ &+ 2\Gamma^T F^T C^T T C F \Gamma \mathbf{u}_h + 2L\mathbf{u}_h \end{aligned} \tag{11.20}$$

Setting Eq. (11.20) to zero and solving for $\mathbf{u}_h$ we have:

$$\mathbf{u}_h^*(p) = \left(L + \Gamma^T F^T C^T T C F \Gamma\right)^{-1} \Gamma^T F^T C^T T\left(\mathbf{r}_h - C A^p \mathbf{x}^{(0)}\right) \tag{11.21}$$

Fig. 11.5 shows simulation results for a 50 deg saccade (thin line, marked ‘noise-free’), for which we assumed $p = 160\text{ms}$. The simulation time step is 1ms, $\mathbf{x}^{(0)} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$, motor costs are $\lambda^{(i)} = 1$, and the target is at 50 deg, $g = 50\pi/180$. The only variable in our system is tracking cost T, which has three parameters: $v = \begin{bmatrix} 5\times 10^9 & 1\times 10^6 & 80 \end{bmatrix}$. We find that the simulated eye arrives at the target, but the velocity profile is symmetric in time (the line labeled noise free in Fig. 11.5). In contrast to our simulation, the velocity profile of the eye during a real saccade (of 50 deg) is highly asymmetric, with the maximum speed attained near the start of the movement. That is, despite the fact that we picked a reasonable movement duration, and found the optimal motor commands for that duration, we could not replicate the observed trajectory. Why did our simulation fail to produce a realistic movement?

Importance of signal dependence noise

For a movement to show an asymmetric speed profile, it must be advantageous to produce large motor commands early into the movement rather than producing the same large motor commands near the end. Chris Harris and Daniel Wolpert (1998) noted that this advantage comes about when the system has a specific noise property: the relationship between motor commands and change in state depends on a noise that grows with the size of the motor commands, called signal-dependent noise. As we will show, our simulation failed to produce a realistic movement because we imagined the system to be noise free.

To consider the effect of noise in control of our system, suppose that motor commands u are affected by noise $\varepsilon \sim N\left(0,\kappa^2u^2\right)$. That is, the noise is a Gaussian distribution with mean zero and variance that is proportional to u^2. We saw examples of this physiological noise in Fig. 10.12B. The fact that the variance of our noise is dependent on the motor commands is the key assumption. If the noise variance was independent of the motor commands, addition of noise would still have no effect on our solution in Eq. 11.23 (this is because what matters is whether the noise has a non-zero derivative with respect to the motor commands). So let us now solve the optimal control problem by assuming that the dynamics of our eye is affected by signal dependent noise:

$$\begin{aligned}
\mathbf{x}^{(k+1)} &= A\mathbf{x}^{(k)} + \mathbf{b}\left(u^{(k)} + \varepsilon^{(k)}\right) \\
\varepsilon^{(k)} &\sim N\left(0, \kappa^2\left(u^{(k)}\right)^2\right) \\
\mathbf{y}^{(k)} &= C\mathbf{x}^{(k)}
\end{aligned} \tag{11.22}$$

Consider the endpoint variance of this system after a sequence of motor commands $u^{(0)},\cdots,u^{(p-1)}$:

$$\begin{aligned}
\mathbf{x}^{(p)} &= A^p\mathbf{x}^{(0)} + \sum_{k=0}^{p-1} A^{p-1-k}\mathbf{b}\left(u^{(k)} + \varepsilon^{(k)}\right) \\
\mathrm{var}\left[\mathbf{x}^{(p)}\right] &= \kappa^2 \sum_{k=0}^{p-1}\left(u^{(k)}\right)^2 A^{p-1-k}\mathbf{b}\mathbf{b}^T\left(A^{p-1-k}\right)^T
\end{aligned} \tag{11.23}$$

The variance at the end of the movement depends on the sum of squared motor commands. Why should we care about this variability? Because the variability affects our accuracy, which is part of the cost function that we are trying to minimize: when the state is a random variable, the expected value of the ‘squared-error’ (as in our cost function in Eq. 11.14) implicitly includes the sum of the squared bias and the variance of the state. If we choose large motor commands, they will produce a larger variance, and this will increase our cost. Interestingly, we see from Eq. (11.23) that the variance depends more strongly on the motor commands that arrive later: the power to which the matrix A is raised is smaller for larger k. Therefore, if we want reduced variance near the end of our movement, we should minimize the size of motor commands, particularly near the end of the movement. If we do this, we will reduce our endpoint variance, improve our accuracy, and as a result reduce our costs.

To express these ideas mathematically, let us see how the cost in Eq. (11.14) is affected by the addition of signal dependent noise. We define:

$$\boldsymbol{\varepsilon}_h = \begin{bmatrix} \varepsilon^{(0)} \\ \varepsilon^{(1)} \\ \vdots \\ \varepsilon^{(p-1)} \end{bmatrix} \qquad U = \begin{bmatrix} u^{(0)} & 0 & \cdots & 0 \\ 0 & u^{(1)} & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & u^{(p-1)} \end{bmatrix} \tag{11.24}$$

The mean and variance of our noise vector are:

$$E\left[\boldsymbol{\varepsilon}_h\right] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \operatorname{var}\left[\boldsymbol{\varepsilon}_h\right] = \kappa^2 UU \tag{11.25}$$

The state at the end of the movement is:

$$\mathbf{x}^{(p)} = A^p \mathbf{x}^{(0)} + F\Gamma\left(\mathbf{u}_h + \boldsymbol{\varepsilon}_h\right) \tag{11.26}$$

The expected value and variance of our state are:

$$\begin{aligned} E\left[\mathbf{x}^{(p)}\right] &= A^p \mathbf{x}^{(0)} + F\Gamma\mathbf{u}_h \\ \operatorname{var}\left[\mathbf{x}^{(p)}\right] &= \kappa^2 F\Gamma UU\Gamma^T F^T \end{aligned} \tag{11.27}$$

Because the cost in Eq. (11.14) is now a random variable, a reasonable thing to do is to minimize its expected value:

$$\begin{aligned} E[J] = E\left[\mathbf{x}^{(p)}\right]^T C^T T C E\left[\mathbf{x}^{(p)}\right] + tr\left[C^T T C \operatorname{var}\left[\mathbf{x}^{(p)}\right]\right] \\ -2E\left[\mathbf{x}^{(p)}\right]^T C^T T\mathbf{r} + \mathbf{r}^T T\mathbf{r} + \mathbf{u}_h^T L\mathbf{u}_h + \alpha\left(1 - \frac{1}{1+\beta p}\right) \end{aligned} \tag{11.28}$$

The trace operator in Eq. (11.28) comes about because of the identity:

$$E\left[\mathbf{x}^T A\mathbf{x}\right] = E[\mathbf{x}]^T AE[\mathbf{x}] + tr\left[A\operatorname{var}[\mathbf{x}]\right]$$

The trace in Eq. (11.28) can be simplified:

$$\begin{aligned} tr\left[C^T T C \operatorname{var}\left[\mathbf{x}^{(p)}\right]\right] &= \kappa^2 tr\left[C^T TCF\Gamma UU\Gamma^T F^T\right] \\ &= \kappa^2 tr\left[U\Gamma^T F^T C^T TCF\Gamma U\right] \\ &= \kappa^2 \mathbf{u}_h^T diag\left[\Gamma^T F^T C^T TCF\Gamma\right]\mathbf{u}_h \end{aligned} \tag{11.29}$$

The term $diag\left[M\right]$ in Eq. (11.29) is the diagonal operator that generates a matrix with only the diagonal elements of the square matrix M. We see that the cost for a system with signal dependent noise (Eq. 11.28) is similar to the cost for a noise-less system (Eq. 11.14), with the crucial exception of an additional term (Eq. 11.29) that penalizes the 'squared' motor commands by an amount proportional to the variance of the signal dependent noise.

Let us now solve the optimal control problem for this system that has signal dependent noise. We pick a movement duration p and find the optimal motor commands by setting the derivative of Eq. (11.28) with respect to $\mathbf{u}_h$ to zero and then solve for $\mathbf{u}_h$:

$$\begin{aligned} S &= diag\left[\Gamma^T F^T C^T TCF\Gamma\right] \\ \mathbf{u}_h^*(p) &= \left(L + \Gamma^T F^T C^T TCF\Gamma + \kappa^2 S\right)^{-1}\Gamma^T F^T C^T T\left(\mathbf{r} - CA^p\mathbf{x}^{(0)}\right) \end{aligned} \tag{11.30}$$

Fig. 11.5 shows simulation results when the standard deviation of our signal dependent noise was set at $\kappa = 0.009$, with all other terms in the simulation unchanged. With this modest amount of noise the speed profile becomes asymmetric, particularly for large amplitude saccades. The simulated saccade now has its greatest speed early in the movement, closely resembling the trajectory of a real saccade (dashed line, Fig. 11.5). The reason for this is because for a system with signal dependent noise, it is better to produce large motor commands early rather than late, as the large noise in the early motor commands will be naturally dissipated by the viscous dynamics of the eye, resulting in little variability by the end of the movement.

Let us summarize the ideas thus far. Control of saccadic eye movements can be viewed as an open-loop process in the sense that motor commands do not rely on sensory feedback. Saccades exhibit characteristic timing and speed profiles. For example, large amplitude saccades tend to have an asymmetric speed profile, with the largest speeds attained early in the movement. If we assume a noise free plant, and a cost function that penalizes endpoint inaccuracy (squared endpoint error), as well as effort (squared motor commands), then the motor commands that minimize such costs do not produce the asymmetric speed profiles. However, if we assume that the motor commands suffer from signal dependent noise, then the expected value of the squared endpoint error implicitly incorporates a measure of endpoint variance. This variance depends more strongly on the motor commands that arrive later in the movement than those that arrive earlier. As a result, the policy (i.e., the sequence of motor commands) that minimizes the cost produces larger speeds earlier in the movement. The asymmetry in speed profiles of saccadic eye movements suggests that the brain is minimizing a cost that includes endpoint accuracy in the face of signal dependent noise.

Step 2: optimal duration and the influence of reward

In step 1 we assumed that the movement should have a specific duration, and then found the optimal motor commands for that duration $\mathbf{u}_h^*(p)$. In our second step, the question that we wish to explore is as follows: if we see something interesting at 15 degrees, why should our brain produce a saccade that is about 70ms in duration? If all that mattered were endpoint accuracy and motor costs, then we should make a very slow movement, something resembling smooth pursuit, not a saccade. However, if time to reward is itself a cost because time discounts reward, then this cost balances the effort and accuracy costs, encouraging us to move faster. The movement duration that we are looking for is one that minimizes the total cost J.

Fig. 11.6A shows the effort, accuracy, and reward costs for a saccade to a target at 20 deg. For a given movement period p, we generate the optimal policy $\mathbf{u}_h^*(p)$ in Eq. (11.30). Finally, we compute the costs in Eq. (11.28). The plot includes the total cost $E\left[J\right]$, accuracy costs $E\left[\left(\mathbf{x}^{(p)}-\mathbf{r}\right)^T T\left(\mathbf{x}^{(p)}-\mathbf{r}\right)\right]$, motor costs $\left(\mathbf{u}_h^*\right)^T L\mathbf{u}_h^*$, and costs of delaying the reward, J_p. On the left side of this figure we have plotted the costs when J_p is of a hyperbolic form, as in Eq. (11.1). Shorter duration saccades have a large cost $E\left[J\right]$ because the costs associated with inaccuracy and effort increase as saccade duration decreases. With increasing saccade duration, the cost of delaying the reward increases. The function $E\left[J\right]$ has a minimum, and this is the optimum movement duration. At this minimum, the movement balances the need to be accurate and lazy vs. the need to maximize reward.

Is there anything special about our hyperbolic cost, or would any function that penalizes time produce similar results? To answer this question, let us consider a quadratic temporal discounting of reward $J_p = \alpha p^2$. We see that for both hyperbolic and quadratic temporal discounting there exist parameter values such that a 20 deg saccade will have its minimum total cost at around 75ms. However, we see that a quadratic discounting of time implies that the loss as a function of movement duration increases rapidly. Therefore, in a quadratic regime there is little loss of reward when one compares two movements of 50 and 60ms in duration, but much greater loss when one compares two movements of 200 and 210ms in duration. In contrast, a hyperbolic discounting of time works exactly opposite. There is greater reward loss for short duration saccades than for long duration saccades. That is, as saccade durations increase, the sensitivity to passage of time decreases. As we will see, this is a crucial property of hyperbolic discounting, and one that allows us to account for the fact that saccade duration as a function of saccade amplitude grows faster than linearly (Fig. 10.3B).

We kept parameter values unchanged and computed optimal saccade durations for movements of various amplitudes in Fig. 11.6B. We see that for a quadratic temporal cost, increasing movement amplitudes accompany smaller and smaller changes in saccade durations. In contrast, for a hyperbolic temporal cost, increasing movement amplitudes accompany a faster than linear increase in saccade durations. The optimal movement duration is a balance between the desire to get to the goal state as soon as possible, while minimizing effort and inaccuracy. A hyperbolic cost of time produces a very different amplitude-duration relationship than a quadratic cost of time. To see why this is true, let us compute the condition that provides us with the optimal movement duration p^*: from Eq. (11.4), the derivative of our effort and accuracy costs must be equal to the derivative of our temporal cost:

$$-\left.\frac{d\left(J_x + J_u\right)}{dp}\right|_{p^*} = \left.\frac{dJ_p}{dp}\right|_{p^*} \tag{11.31}$$

Fig. 11.7 plots the above two derivatives for hyperbolic, linear, and quadratic time costs. Because the effort and accuracy costs depend on stimulus location (saccade amplitude), there is a separate curve for each amplitude. The point at which the derivative of the effort/accuracy cost crosses the temporal cost is the optimal movement duration. For example, for a 20deg saccade all three temporal costs produce an optimal duration of around 75ms. However, as saccade amplitudes increase, the three temporal costs predict very different saccade durations. When we compare the predicted pattern of saccade durations with actual data, we find that the hyperbolic function is an excellent fit. Fig. 11.8 summarizes these results for three kinds of temporal cost functions: quadratic, linear, and hyperbolic. This figure also includes data from actual saccades (data from

Fig. 10.3B). A quadratic cost of time produces reasonable estimates of saccade parameters for small amplitudes, but fails for larger amplitudes. The reason is that with a quadratic temporal cost, with passage of time the cost of time grows (derivative is increasing, as in Fig. 11.7). If we consider a linear cost of time, an approach that was suggested by Chris Harris and Daniel Wolpert (2006), once again we can produce reasonable estimates of saccade parameter for small amplitude saccades, but the simulations fail for larger amplitudes. However, in the case of a hyperbolic cost of time we can account for durations of both small amplitude as well as large amplitude saccades. Indeed, the fact that movement durations increase faster than linearly as a function of movement amplitudes is consistent with a hyperbolic cost of time, but not linear or quadratic.

11.5 Cost of time and temporal discounting of reward

Why should the brain impose a hyperbolic cost on duration of movements? The answer, in our opinion, is that this cost expresses how the brain temporally discounts reward. We saw examples of temporal discounting of reward in the response of dopamine neurons (Fig. 11.2B), in the response of LIP neurons (Fig. 11.2C), and in the decisions that people made regarding money (Fig. 11.1A) or water (Fig. 11.1B). Suppose that the brain penalizes movement durations because passage of time delays the acquisition of reward. If this hypothesis is true, then it follows that movement kinematics should vary as a function of the amount of reward. For example, if we make a movement in response to a stimulus that promises little reward, α in Eq. (1) is small and the motor and accuracy costs become relatively more important. As a consequence, when our brain assigns a low value to the stimulus, our movement toward that movement should be slow. To explore this idea, let us consider what happens to saccades when we alter the value of the stimulus α. Movement durations depend on the rate at which reward value is discounted in time (Eq. 11.31). That is, movement duration depends on the derivate of cost J_p. This derivative is:

$$\frac{dJ_p}{dp} = \frac{\alpha\beta}{\left(1+\beta p\right)^2} \tag{11.32}$$

As α decreases, so does the derivative of the reward discount function. Fig. 11.9A plots this derivative for various α. We see that the optimal movement duration increases as stimulus value α decreases. An increase in movement duration coincides with a decrease in peak movement speed.

For example, the opportunity to look at a face is a valued commodity, and physical attractiveness is a dimension along which value rises (Hayden et al., 2007). As α increases, durations of simulated saccades decrease, resulting in higher velocities. This potentially explains why people make faster saccades to look at faces (Xu-Wilson et al., 2009).

A hyperbolic function is a good fit to discharge of dopamine cells in the brain of monkeys that have been trained to associate visual stimuli with delayed reward (Kobayashi and Schultz, 2008). That is, the response of these cells to stimuli is a good predictor of the temporally discounted value of these stimuli (Fig. 11.2B). In Parkinson's disease (PD), many of the dopaminergic cells die. Let us hypothesize that this is reflected in a devaluation of the stimulus, i.e., a smaller than normal α. In Fig. 11.9B we have plotted velocity-amplitude data from a number of studies that have examined saccades of people with moderate to severe PD. The saccades of PD patients exhibit an intriguing property: the peak speeds are normal for small amplitudes, but become much slower than normal for large amplitudes. If we simply reduce stimulus value α, the model reproduces velocity-amplitude characteristics of PD patients (Fig. 11.9B).

If an abnormally small stimulus value can produce slow saccades, then an abnormally large value should produce fast saccades. In schizophrenia, saccade velocities are faster than in healthy controls (Mahlberg et al., 2001). Schizophrenia is a complex disease that likely involves dysfunction of generation and uptake of many neurotransmitters including dopamine, glutamate, and GABA. Stone et al. (2007) suggested that in the striatum of schizophrenic patients, there is greater than normal dopamine synthesis. Kapur (2003) noted schizophrenics assign an unusually high salience to stimuli so that "every stimulus becomes loaded with significance and meaning". Indeed, currently available antipsychotic medications have one common feature: they block dopamine D2 receptors. The reward temporal discount function in schizophrenia has a higher slope with respect to controls (Heerey et al., 2007;Kloppel et al., 2008), implying a greater discount rate. In our framework, this produces a faster rise in the cost of time, increasing saccade speeds.

Consider another curious fact regarding saccades: as we age, the kinematics of our saccades change. Children produce faster saccades than young adults (Fioravanti et al., 1995;Munoz et al., 2003). According to our theory, the differences in saccade kinematics should be a consequence of the way the child's brain temporally discounts reward. The availability of dopamine in the brain declines with age. For example, rhesus monkeys exhibit a 50% decline in dopamine concentrations in the caudate and putamen from youth to old age (Collier et al., 2007), and squirrel monkeys exhibit a 20% decline (McCormack et al., 2004). Green et al. (1999) measured the temporal discount rate of reward in both young children and adults and found that the initial slope of the discount function was 2-3 times larger in children than adults. They would rather take a single cookie now, than wait for a brief period in order to receive two cookies. Shadmehr et al. (Shadmehr et al., 2010) showed that by increase the slope of the temporal cost function (via parameter α) by a factor of two, the resulting saccades share the velocity-amplitude relationship found in children's saccades.

Let us now consider the fact that saccade velocities differ across species. For example, rhesus monkeys exhibit velocities that are about twice as fast as humans (Straube et al., 1997;Chen-Harris et al., 2008). Although there are small differences in the eye plants of monkeys and humans, such differences cannot account for the remarkably faster saccades in monkeys (Shadmehr et al., 2010). One possibility is that the differences in saccades are related to inter-species differences in valuation of stimuli and temporal discounting of reward. Indeed, some rhesus monkeys exhibit a much greater temporal discount rate than humans: when making a choice between stimuli that promise reward (juice) over a range of tens of seconds, thirsty adult rhesus monkeys (Kobayashi and Schultz, 2008;Hwang et al., 2009) exhibit discount rates that are many times that of thirsty undergraduate students (Jimura et al., 2009) (compare Fig. 11.2A with Fig. 11.1B). If we take into account this higher slope of the temporal discount rate, the simulated monkey saccades will have velocities that were fairly consistent with the velocities that have been recorded from this specie (Shadmehr et al., 2010).

11.6 State dependent value of a stimulus

Animals do not assign a value to a stimulus based on its inherent properties, but based on their own state when the stimulus was encountered. For example, birds that are initially trained to obtain equal rewards after either large or small effort, and are then offered a choice between the two rewards without the effort, generally choose the reward previously associated with the greater effort. Tricia Clement and colleagues (Clement et al., 2000) trained pigeons to peck at circles. During training, on some trials a white circle would appear and after the pigeon pecked it once, two circles appeared (red and yellow), as shown in Fig. 11.10A. If the red circle was pecked, food was given. On other trials, the white circle had to be pecked 20 times before green and blue circles appeared. Pecking the green circle resulted in food. So the idea was to test whether after pecking 20 times would the green circle have greater value than the red, even though they both gave the same food. On test trials, some pigeons saw a white circle and after pecking it once they saw red and green circles. In about 70% of the trials, they pecked the green circle (Fig. 11.10B). Other pigeons saw a white circle and after pecking it 20 times they saw red and green circles. They chose the green circle about 65% of the time. Pigeons that were only shown red and green circles during test trials also preferred the green circle (Fig. 11.10B). This paradoxical result can be understood in terms of a greater utility (i.e., relative usefulness, rather than absolute value) for the reward that was attained following a more effortful action. The green circle resulted in the same food as the red circle, but the food associated with the green circle was attained during training after 20 pecks. This phenomenon is called state-dependent valuation learning, and is present in a wide variety of species from mammals to invertebrates (Pompilio and Kacelnik, 2010).

The state-dependent valuation of stimuli allows us to consider a curious fact: kinematics of saccades to target of a reaching movement are affected by the load on the arm. For example, the peak speed of a saccade is higher when there is a load that resists the reach, and lower when the load assists the reach (van Donkelaar et al., 2004). Why should varying the effort required to perform a reach to a target affect saccade velocities to that target? A reaching movement that is resisted by a load arrives at the target after a larger effort than one that is assisted. The more effortful state in which the reward is encountered favors assignment of a greater utility for that stimulus. This greater utility may contribute to a faster saccade.

11.7 Why hyperbolic discounting of reward?

Why should passage of time discount reward in a hyperbolic fashion? Other than the fact that a hyperbolic function seems to fit the empirical data better (Myerson and Green, 1995), is there a rationale for this pattern? Alex Kacelnik (1997) has considered this question and made the following suggestion: perhaps the choices that people make regarding time and reward is a reflection of a more fundamental normative law in which the objective is to maximize reward per unit of time. For example, suppose that you are given a choice between reward α_1 now and reward α_2 at delay t. Further suppose that you are asked to make these choices at intervals γ (this would represent an inter-trial interval). If your objective is to maximize reward per unit of time, then you would compare two quantities: $J_1 = \dfrac{\alpha_1}{\gamma}$ and $J_2 = \dfrac{\alpha_2}{\gamma + t}$. You should pick the choice that gives you the larger reward rate. Let us show that if you choose in this way, you are implicitly discounting hyperbolically. Suppose that given the choice between α_1 now and α_2 at t, you choose each at around 50% probability. This would imply that:

$$\frac{\alpha_1}{\gamma} = \frac{\alpha_2}{\gamma + t}$$

Which implies that α_1 declines with time at a rate of γ^{-1}:

$$\alpha_1 = \frac{\alpha_2}{1 + \gamma^{-1} t} \tag{11.33}$$

Therefore, a policy that maximizes the rate of reward would result in a hyperbolic temporal discounting of reward.

We can use this idea to reformulate the costs in our motor control problem. Suppose that we perform actions in such a way as to maximize reward per unit of time, while minimizing the effort expended per unit of time. That is, what matters is the difference between the reward that we hope to attain and the effort we expect to expend, per unit of time. If $\Pr\left(r = 1 \middle| \mathbf{u}(p)\right)$ represents the probability of acquiring reward that has value α, given motor commands $\mathbf{u}$ that produce a movement of duration p, and J_u represents the effort costs associated with these commands, then the motor commands should be chosen to maximize the following:

$$J = \frac{\alpha \Pr\left(r = 1 \middle| \mathbf{u}(p)\right) - J_u}{1 + \gamma^{-1} p} \tag{11.34}$$

In the numerator, the probability of success would generally increase as movement duration p increases. The term $\left(1 + \gamma^{-1} p\right)^{-1}$ is a decreasing function of p, resulting in a cost J in Eq. (11.34) that will have a maximum at a particular movement duration. It remains to be seen whether maximizing the rate of reward is sufficient to explain movement patterns in biology, as well as the economic decision making processes that are reflected in temporal discounting of reward.

11.8 Summary

Suppose that the objective of any voluntary movement is to place the body at a more valuable state. Further suppose that the value associated with this state is not static, but is discounted in time: we would rather receive the reward now than later. The value that one assigns a stimulus, and the rate at which this value declines, forms a reward temporal discount function. The temporal discounting of reward forms an implicit cost of time, i.e., a penalty for the duration of the movement. This penalty is a hyperbolic function. If one assumes that motor commands are programmed to minimize effort while maximizing reward, and if one further assumes that reward loses value hyperbolically as a function of movement duration, then one can mathematically reproduce the relationship between movement duration, amplitude, and velocity in saccades.

Research over the last two decades indicates that the reward temporal discount function is affected by disease, is affected by development and aging, and is affected by evolution (inter-species differences in the brain). There appears to be a correlation between changes in the reward temporal discount function and changes in duration and velocities of saccadic eye movements. This correlation suggests that the motor commands that move the eyes reflect a specific cost of time, one in which passage of time discounts reward. Here, we used open-loop optimal control to describe a cost function that includes effort and reward, and then modeled control of eye movements to ask why saccades have their particular duration and velocities, and why these kinematics change as stimulus value or discounting of reward changes in the brain. The motor commands that move our body may be a reflection of an economic decision regarding reward and effort.

Figure 11.1. Temporal discounting of reward in people. **A**. Value of $1000 as a function of time to acquisition of the money. For example, on average, undergraduate students would accept $400 now or $1000 at 4 years at equal probability. The best fits to the data via a hyperbolic function or an exponential function are shown. Exponential functions tend to fit poorly particularly for long delays. (Data from (Myerson and Green, 1995)). **B.** Temporal discounting of water for thirsty undergraduate students. The volunteers were given a choice between a small amount of water now and a larger amount some seconds later. The points show subjective equality. For example, students would take about 8ml of water now than wait 30s for 16ml. (Data from (Jimura et al., 2009)).

Figure 11.2. Behavioral and neural correlates of temporal discounting of reward in rhesus monkeys. **A**. Monkeys were trained to fixate a center dot on a monitor while two visual stimuli appeared on the left and right side of the fixation point. One stimulus was associated with a small reward that would be available now, and the other stimulus was associated with a larger reward that would be available some seconds later. The animal made a saccade to one of the stimuli to indicate its choice. The data summarizes the choices made by the animals via a temporal discounting function. (Data from (Kobayashi and Schultz, 2008)) **B**. Discharge of dopaminergic cells in the substantia nigra in response to stimuli that were paired with a given amount of reward at some time in the future. In the task, the animals were trained on five visual stimuli that each was paired with a constant amount of reward delivered immediately, at 2s, 4s, etc. The plot shows the response of the population of dopamine cells to presentation of the visual stimulus. (Data from (Kobayashi and Schultz, 2008)) **C**. Behavioral temporal discounting of liquid reward for a monkey, along with normalized activity of cells in the LIP area of the same monkey during the decision making task. The animal made a choice between small reward now vs. a larger reward that would be available some seconds later. The animals indicated its choice by making a saccade to the visual stimulus associated with that reward/delay. In the period before the saccade, LIP cells that had their receptive field on the eventual choice responded by discharging proportional to the subjective value of the stimulus. (Data from (Louie and Glimcher, 2010)) **D**. Same as in part C, but for a different monkey.

Figure 11.3. Change of mind and hyperbolic vs. exponential discounting of reward. **A**. Hyperbolic discounting. At time zero, the discounted value associated with the later event (black line) is higher than the sooner event (gray line). Therefore, at time zero you pick the later event. However, as time passes and we get closer to the time of the sooner event, its discounted value

becomes greater than the later event. This produces a change of mind. **B**. Exponential discounting. If at time zero the discounted value of the sooner event is less than the later event, its value will remain smaller of the two choices at all times, preventing a change of mind.

Figure 11.4. A one dimensional model of the eye plant. Models in A and B are equivalent.

Figure 11.5. Saccades to a target at 50deg were generated for a noise free system (Eq. 11.15), and for a system with signal dependent noise (Eq. 11.22), under a policy that minimized endpoint variance and effort. The velocity profile is asymmetric in the case of a system with signal dependent noise, showing a high initial velocity and then a slower tail. This is because endpoint accuracy depends more strongly on the motor commands that arrive later in the movement than those that arrive earlier. Asymmetric velocity profiles during large amplitude saccades arise in optimal control of a system that suffers from signal dependent noise.

Figure 11.6. The effect of quadratic and hyperbolic discounting of reward on saccade durations. **A.** Costs of a saccade to a target at 20 deg. For the quadratic discount, $J_p = \alpha p^2$, $\alpha = 1.8\times10^5$. For the hyperbolic discount, $\alpha = 1.15\times10^4$ and $\beta = 3$. All other parameters are equal for the two simulations. Optimum saccade duration for both simulations is around 75ms. **B.** Expected value of the cost for various saccade amplitudes. The optimum saccade duration is noted by the gray line. Parameter values are unchanged from part A.

Figure 11.7. The optimal movement duration is one that satisfies Eq. (11.31), i.e., a duration at which the derivative of the effort/accuracy cost is equal to the derivative of the temporal cost. The intersection between the derivative of the effort/accuracy cost and temporal cost is the optimal movement duration. When we set the quadratic, linear, and hyperbolic temporal costs to produce similar optimal movement times for one saccade amplitude (in this case 20 deg), they produce very different duration-amplitude relationships for other amplitudes. Quadratic: $\alpha = 1.8\times10^5$. Linear: $\alpha = 2.4\times10^4$. Hyperbolic: $\alpha = 1.15\times10^4$, $\beta = 3$.

Figure 11.8. Optimal saccade durations as a function of amplitude for various temporal discount functions (quadratic, linear, and hyperbolic). The data points are mean and SD of saccade durations in healthy people (Collewijn et al., 1988). The gray line is for reference. The thin black lines are simulation results for various stimulus values, with the largest value producing

shorter duration saccades. While linear and quadratic discounting schemes can account for durations of short amplitude saccades, they underestimate durations of long amplitude saccades. Simulation values are as follows. Quadratic: $\alpha = (1, 2, 3) \times 10^5$. Linear: $\alpha = (2, 2.5, 3.0) \times 10^4$. Hyperbolic: $\alpha = (8.5, 9.5, 11.5) \times 10^3$, $\beta = 3$.

Figure 11.9. Effects of changing the stimulus value α on saccade durations and saccade peak velocities for a hyperbolic temporal discounting of reward. **A**. Optimal saccade duration is the intersection of the derivative of the accuracy/effort cost (dashed lines) and the derivative of the temporal cost (solid lines). As α decreases (indicated by the arrow), the optimal saccade duration increases. Simulated values are: $\alpha = (8, 10, 12, 14) \times 10^3$ and $\beta = 3$. **B.** Peak saccade velocity as a function of amplitude for various stimulus values α. Simulation results are shown by the lines for $\alpha = (6 - 13.5) \times 10^3$ and $\beta = 3$. For each line, the stimulus value α was kept constant. Reducing stimulus value decreased saccade speeds, but the changes in speed were much bigger for large amplitude saccades than small amplitudes. Data from various papers are shown by the numbers in healthy controls and in patients with Parkinson's disease. These data are from Shibasaki et al. (1979), Collewijn et al. (1988), White et al. (1983), Blekher et al. (2000), and Nakamura et al. (1991).

Figure 11.10. The value of a given reward depends on the state of the animal at the time that reward was attained. **A**. Pigeons were trained to peck at circles. During training, on some trials a white circle would appear and after the pigeon pecked it once, two circles appeared (red and yellow). If the red circle was pecked, food was given. On other trials, the white circle had to be pecked 20 times before green and blue circles appeared. Pecking the green circle resulted in food. **B**. On test trials, some pigeons saw a white circle and after pecking it once they saw red and green circles. In about 70% of the trials, they pecked the green circle. Other pigeons saw a white circle and after pecking it 20 times they saw red and green circles. They chose the green circle about 65% of the time. Pigeons that were only shown red and green circles during test trials also preferred the green circle. Data from (Clement et al., 2000).

Reference List

Blekher T, Siemers E, Abel LA, Yee RD (2000) Eye movements in Parkinson's disease: before and after pallidotomy. Invest Ophthalmol Vis Sci 41:2177-2183.

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

Clement TS, Feltus JR, Kaiser DH, Zentall TR (2000) Work ethic in pigeons: reward value is directly related to the effort or time required to obtain the reward. Psychonomic Bulletin Review 7:100-106.

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

Collier TJ, Lipton J, Daley BF, Palfi S, Chu Y, Sortwell C, Bakay RA, Sladek JR, Jr., Kordower JH (2007) Aging-related changes in the nigrostriatal dopamine system and the response to MPTP in nonhuman primates: diminished compensatory mechanisms as a prelude to parkinsonism. Neurobiol Dis 26:56-65.

Fagg AH, Shah A, Barto AG (2002) A computational model of muscle recruitment for wrist movements. J Neurophysiol 88:3348-3358.

Fioravanti F, Inchingolo P, Pensiero S, Spanio M (1995) Saccadic eye movement conjugation in children. Vision Res 35:3217-3228.

Green L, Myerson J, Holt DD, Slevin JR, Estle SJ (2004) Discounting of delayed food rewards in pigeons and rats: is there a magnitude effect? J Exp Anal Behav 81:39-50.

Green L, Myerson J, Ostaszewski P (1999) Discounting of delayed rewards across the life span: age differences in individual discounting functions. Behavioural Processes 46:89-96.

Guthrie BL, Porter JD, Sparks DL (1983) Corollary discharge provides accurate eye position information to the oculomotor system. Science 221:1193-1195.

Harris CM, Wolpert DM (2006) The Main Sequence of Saccades Optimizes Speed-accuracy Trade-off. Biol Cybern 95:21-29.

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

Hayden BY, Parikh PC, Deaner RO, Platt ML (2007) Economic principles motivating social attention in humans. Proc Biol Sci 274:1751-1756.

Heerey EA, Robinson BM, McMahon RP, Gold JM (2007) Delay discounting in schizophrenia. Cogn Neuropsychiatry 12:213-221.

Hwang J, Kim S, Lee D (2009) Temporal discounting and inter-temporal choice in rhesus monkeys. Front Behav Neurosci 3:9.

Jimura K, Myerson J, Hilgard J, Braver TS, Green L (2009) Are people really more patient than other animals? Evidence from human discounting of real liquid rewards. Psychon Bull Rev 16:1071-1075.

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

Kapur S (2003) Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry 160:13-23.

Keller EL (1973) Accommodative vergence in the alert monkey. Motor unit analysis. Vision Res 13:1565-1575.

Keller EL, Robinson DA (1971) Absence of a stretch reflex in extraocular muscles of the monkey. J Neurophysiol 34:908-919.

Kloppel S, Draganski B, Golding CV, Chu C, Nagy Z, Cook PA, Hicks SL, Kennard C, Alexander DC, Parker GJ, Tabrizi SJ, Frackowiak RS (2008) White matter connections reflect changes in voluntary-guided saccades in pre-symptomatic Huntington's disease. Brain 131:196-204.

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

Louie K, Glimcher PW (2010) Separating value from choice: delay discounting activity in the lateral intraparietal area. J Neurosci 30:5498-5507.

Mahlberg R, Steinacher B, Mackert A, Flechtner KM (2001) Basic parameters of saccadic eye movements--differences between unmedicated schizophrenia and affective disorder patients. Eur Arch Psychiatry Clin Neurosci 251:205-210.

McCormack AL, Di Monte DA, Delfani K, Irwin I, DeLanney LE, Langston WJ, Janson AM (2004) Aging of the nigrostriatal system in the squirrel monkey. J Comp Neurol 471:387-395.

Munoz DP, Armstrong IT, Hampton KA, Moore KD (2003) Altered control of visual fixation and saccadic eye movements in attention-deficit hyperactivity disorder. J Neurophysiol 90:503-514.

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

Nakamura T, Kanayama R, Sano R, Ohki M, Kimura Y, Aoyagi M, Koike Y (1991) Quantitative analysis of ocular movements in Parkinson's disease. Acta Otolaryngol Suppl 481:559-562.

O'Sullivan I, Burdet E, Diedrichsen J (2009) Dissociating variability and effort as determinants of coordination. PLoS Comput Biol 5:e1000345.

Pompilio L, Kacelnik A (2010) Context-dependent utility overrides absolute memory as a determinant of choice. Proc Natl Acad Sci U S A 107:508-512.

Robinson DA, Gordon JL, Gordon SE (1986) A model of the smooth pursuit eye movement system. Biol Cybern 55:43-57.

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.

Shibasaki H, Tsuji S, Kuroiwa Y (1979) Oculomotor abnormalities in Parkinson's disease. Arch Neurol 36:360-364.

Stone JM, Morrison PD, Pilowsky LS (2007) Glutamate and dopamine dysregulation in schizophrenia--a synthesis and selective review. J Psychopharmacol 21:440-452.

Straube A, Fuchs AF, Usher S, Robinson FR (1997) Characteristics of saccadic gain adaptation in rhesus macaques. J Neurophysiol 77:874-895.

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.

White OB, Saint-Cyr JA, Tomlinson RD, Sharpe JA (1983) Ocular motor deficits in Parkinson's disease. II. Control of the saccadic and smooth pursuit systems. Brain 106 (Pt 3):571-587.

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