



# A general theory of intertemporal decision-making and the perception of time

Vijay M. K. Namboodiri<sup>1</sup>, Stefan Mihalas<sup>2</sup>, Tanya M. Marton<sup>1</sup> and Marshall G. Hussain Shuler<sup>1\*</sup>

<sup>1</sup> Department of Neuroscience, Johns Hopkins University, Baltimore, MD, USA

<sup>2</sup> Allen Institute for Brain Science, Seattle, WA, USA

## Edited by:

Paul E. M. Phillips, University of Washington, USA

## Reviewed by:

Geoffrey Schoenbaum, University of Maryland School of Medicine, USA

Matthew S. Matell, Villanova University, USA

Fuat Balci, Koç University, Turkey

## \*Correspondence:

Marshall G. Hussain Shuler,  
Department of Neuroscience, Johns Hopkins University, 725 N Wolfe Street, 914 WBSB, Baltimore, MD 21205, USA  
e-mail: shuler@jhmi.edu

Animals and humans make decisions based on their expected outcomes. Since relevant outcomes are often delayed, perceiving delays and choosing between earlier vs. later rewards (intertemporal decision-making) is an essential component of animal behavior. The myriad observations made in experiments studying intertemporal decision-making and time perception have not yet been rationalized within a single theory. Here we present a theory—Training-Integrated Maximized Estimation of Reinforcement Rate (TIMERR)—that explains a wide variety of behavioral observations made in intertemporal decision-making and the perception of time. Our theory postulates that animals make intertemporal choices to optimize expected reward rates over a limited temporal window which includes a past integration interval—over which experienced reward rate is estimated—as well as the expected delay to future reward. Using this theory, we derive mathematical expressions for both the subjective value of a delayed reward and the subjective representation of the delay. A unique contribution of our work is in finding that the past integration interval directly determines the steepness of temporal discounting and the non-linearity of time perception. In so doing, our theory provides a single framework to understand both intertemporal decision-making and time perception.

**Keywords:** decision-making, discounting, intertemporal choice theory, time perception, impulsivity, scalar timing

## INTRODUCTION

Survival and reproductive success depends on beneficial decision-making. Such decisions are guided by judgments regarding outcomes, which are represented as expected reinforcement amounts. As actual reinforcements are often available only after a delay, measuring delays and attributing values to reinforcements that incorporate the cost of time is an essential component of animal behavior (Stephens and Krebs, 1986; Stephens, 2008). Yet, how animals perceive time and assess the worth of delayed outcomes—the quintessence of intertemporal decision-making—though fundamental, remains to be satisfactorily answered (Frederick et al., 2002; Kalenscher and Pennartz, 2008; Stephens, 2008). Rationalizing both the perception of time and the valuation of outcomes delayed in time in a unified framework would significantly improve our understanding of basic animal behavior, with wide-ranging applications in fields such as economics, ecology, psychology, cognitive disease, and neuroscience.

In the past, many theories including Optimal Foraging Theory (Stephens and Krebs, 1986; Stephens, 2008) (OFT), Discounted Utility Theory (Samuelson, 1937; Frederick et al., 2002; Kalenscher and Pennartz, 2008) (DUT), Ecological Rationality Theory (Bateson and Kacelnik, 1996; Stephens and Anderson, 2001; Stephens, 2008) (ERT), as well as other psychological models (Frederick et al., 2002; Kalenscher and Pennartz, 2008; Peters and Büchel, 2011; Van den Bos and McClure, 2013) have been proposed as solutions to the question of intertemporal choice. Of these, OFT, DUT, and ERT attempt to understand ultimate

causes of behavior through general optimization criteria, whereas psychological models attempt to understand its proximate biological implementation. The algorithms specified by these prior theories and models for intertemporal decision-making are all defined by their temporal discounting function—the ratio of subjective value of a delayed reward to the subjective value of the reward when presented immediately. These algorithms come in two major forms: hyperbolic (and hyperbolic-like) discounting functions (e.g., OFT and ERT) (Stephens and Krebs, 1986; Frederick et al., 2002; Kalenscher and Pennartz, 2008; Stephens, 2008), and exponential (and exponential-like, e.g.,  $\beta$ - $\delta$  Frederick et al., 2002; Peters and Büchel, 2011; Van den Bos and McClure, 2013) discounting functions (e.g., DUT) (Samuelson, 1937; Frederick et al., 2002; Kalenscher and Pennartz, 2008). Hyperbolic discounting functions have been widely considered to be better fits to behavioral data than exponential functions (Frederick et al., 2002; Kalenscher and Pennartz, 2008).

None of these theories and models can systematically explain the breadth of data on intertemporal decision-making; we argue that the inability of prior theories to rationalize behavior stems from the lack of biologically-realistic constraints on general optimization criteria (see next section). Further, while intertemporal decision-making necessarily requires perception of time, theories of intertemporal decision-making and time perception (Gibbon et al., 1997; Lejeune and Wearden, 2006) are largely independent and do not attempt to rationalize both within a single framework. The motivation for our present work was to create a biologically-realistic and parsimonious theory of

intertemporal decision-making and time perception which proposes an algorithmically-simple decision-making process to (1) maximize fitness and (2) to explain the diversity of behavioral observations made in intertemporal decision-making and time perception.

## PROBLEMS WITH CURRENT THEORIES AND MODELS

Intertemporal choice behavior has been modeled using two dissimilar approaches. The first approach is to develop theories that explore ultimate (Alcock and Sherman, 1994) causes of behavior through general optimization criteria (Samuelson, 1937; Stephens and Krebs, 1986; Bateson and Kacelnik, 1996; Stephens and Anderson, 2001; Frederick et al., 2002; Stephens, 2008). In ecology, there are two dominant theories of intertemporal choice, OFT and ERT. The statement of OFT posits that the choice behavior of animals should result from a global maximization of a “fitness currency” representing long-term *future* reward rate (Stephens and Krebs, 1986; Stephens, 2008). However, how animals could in principle achieve this goal is unclear, as they face at least two constraints: (1) they cannot know the future beyond the currently presented options, and (2) they have limited computational/memory capacity. Owing to these constraints, prior algorithmic implementations of OFT assume that the current trial structure repeats *ad-infinitum*. Therefore, maximizing reward rates over the indefinite future can be re-written as maximizing reward rates over an effective trial (including all delays in the trial) (Stephens and Krebs, 1986; Bateson and Kacelnik, 1996; Stephens and Anderson, 2001; Stephens, 2008). Thus, OFT predicts a hyperbolic discounting function. ERT, on the other hand, states that it is sufficient to maximize reward rates only over the delay to the reward in the choice under consideration, (i.e., locally) to attain ecological success (Bateson and Kacelnik, 1996; Stephens and Anderson, 2001; Stephens, 2008), also predicting a hyperbolic discounting function. In economics, DUT (Samuelson, 1937; Frederick et al., 2002) posits that animals maximize long-term exponentially-discounted *future* utility so as to maintain temporal consistency of choice behavior (Samuelson, 1937; Frederick et al., 2002).

The second approach, mainly undertaken by psychologists and behavioral analysts, is to understand the proximate (Alcock and Sherman, 1994) origins of choices by modeling behavior using empirical fits to data collected from standard laboratory tasks (Kalenscher and Pennartz, 2008). An overwhelming number of these behavioral experiments, however, contradict the above theoretical models. Specifically, animals exhibit hyperbolic discounting functions, inconsistent with DUT (Frederick et al., 2002; Kalenscher and Pennartz, 2008; Stephens, 2008; Pearson et al., 2010), and violate the postulate of global reward rate maximization, inconsistent with OFT (Stephens and Anderson, 2001; Kalenscher and Pennartz, 2008; Stephens, 2008; Pearson et al., 2010). Further, there are a wide variety of observations like (1) the variability of discounting steepness within and across individuals (Frederick et al., 2002; Schweighofer et al., 2006; Luhmann et al., 2008), and many “anomalous” behaviors including (2) “Magnitude Effect” (Frederick et al., 2002; Kalenscher and Pennartz, 2008) (the steepness of discounting becomes lower as the magnitude of the reward increases),

(3) “Sign Effect” (Frederick et al., 2002; Kalenscher and Pennartz, 2008) (gains are discounted more steeply than losses), and (4) differential treatment of punishments (Loewenstein and Prelec, 1992; Frederick et al., 2002; Kalenscher and Pennartz, 2008), that are not explained by ERT (nor OFT and DUT). It must also be noted that none of the above theories are capable of explaining how animals measure delays to rewards, nor do prior theories of time perception (Gibbon et al., 1997; Lejeune and Wearden, 2006) attempt to explain intertemporal choice. Though psychology and behavioral sciences attempt to rationalize the above observations by constructing proximate models invoking phenomena like attention, memory, and mood (Frederick et al., 2002; Kalenscher and Pennartz, 2008; Van den Bos and McClure, 2013), ultimate causes are rarely proposed. As a consequence, these models of animal behavior are less parsimonious, and often *ad-hoc*.

In order to explain behavior, an ultimate theory must consider appropriate proximate constraints. The lack of appropriate constraints might explain the inability of the above theories to rationalize experimental data. By merely stating that animals maximize indefinitely-long-term *future* reward rates or discounted utility, the optimization criteria of OFT and DUT requires animals to consider the effect of all possible future reward-options when making the current choice (Stephens and Krebs, 1986; Kalenscher and Pennartz, 2008). However, such a solution would be biologically implausible for at least three reasons: (1) animals cannot know all the rewards obtainable in the future; (2) even if animals knew the disposition of all possible future rewards, the combinatorial explosion of such a calculation would present it with an untenable computation (e.g., in order to be optimal when performing even 100 sequential binary choices, an animal will have to consider each of the  $2^{100}$  combinations); (3) animals cannot persist for indefinitely long intervals without food in the hope of obtaining an unusually large reward in the distant future, even if the reward may provide the highest long-term reward rate (e.g., option between 11,000 units of reward in 100 days vs. 10 units of reward in 0.1 day). On the other hand, ERT, although computationally-simple, expects an animal to ignore its past reward experience while making the current choice.

To contend with uncertainties regarding the future, an animal could estimate reward rates based on an expectation of the environment derived from its past experience. In a world that presents large fluctuations in reinforcement statistics over time, estimating reinforcement rate using the immediate past has an advantage over using longer-term estimations because the correlation between the immediate past and the immediate future is likely high. Hence, our TIMERR theory proposes an algorithm for intertemporal choice that aims to maximize expected reward rate based on, and constrained by, memory of past reinforcement experience. As a consequence, it postulates that time is subjectively represented such that subjective representation of reward rate accurately reflects objective changes in reward rate (see section TIMERR Theory: Time Perception). In doing so, we are capable of explaining a wide variety of fundamental observations made in intertemporal decision-making and time perception. These include hyperbolic discounting (Stephens and Krebs, 1986; Stephens and Anderson, 2001; Frederick et al., 2002; Kalenscher and Pennartz, 2008), “Magnitude” (Myerson and Green, 1995;

Frederick et al., 2002; Kalenscher and Pennartz, 2008) and “Sign” effects (Frederick et al., 2002; Kalenscher and Pennartz, 2008), differential treatment of losses (Frederick et al., 2002; Kalenscher and Pennartz, 2008), scaling of timing errors with interval duration (Gibbon, 1977; Gibbon et al., 1997; Matell and Meck, 2000; Buhusi and Meck, 2005; Lejeune and Wearden, 2006), and, observations that impulsive subjects (as defined by abnormally steep discounting) under-produce (Wittmann and Paulus, 2008) time intervals and show larger timing errors (Wittmann et al., 2007; Wittmann and Paulus, 2008) (see “Summary” for a full list). It thereby recasts the above-mentioned “anomalies” not as flaws, but as features of reward-rate optimization under experiential constraints.

## MOTIVATION BEHIND THE TIMERR ALGORITHM

To illustrate the motivation and reasoning behind our theory, we consider a simple behavioral task. In this task, an animal must make decisions on every trial between two randomly chosen (among a finite number of possible alternatives) known reinforcement-options. Having chosen an option on one trial, the animal is required to wait the corresponding delay to obtain the reward amount chosen. An example environment with three possible reinforcement-options is shown in **Figure 1A**. We assert that the goal of the animal is to gather the maximum total reward over a fixed amount of time, or equivalently, to attain the maximum total (global) reward rate over a fixed number of trials.

Assuming a stationary reinforcement-environment in which it is not possible to directly know the pattern of future reinforcements, an animal may yet use its past reinforcement experience to instruct its current choice. Provisionally, suppose also that an animal can store its entire reinforcement-history in the task in its memory. So rather than maximizing reward rates into the future as envisioned by OFT, the animal can then maximize the total reward rate that would be achieved so far (at the end of the current trial). In other words, the animal could pick the option that when chosen, would lead to the highest global reward rate over all trials until, and including, the current trial, i.e.,

$$\text{Pick option with the highest value for } \frac{R + r_i}{T + t_i} \quad (1)$$

where  $T$  is the total time elapsed in the session so far,  $R$  is the total reward accumulated so far and  $(r_i, t_i)$  is the reward magnitude and delay, respectively, for the various reinforcement-options on the current trial. This ordered pair notation will be followed throughout the paper.

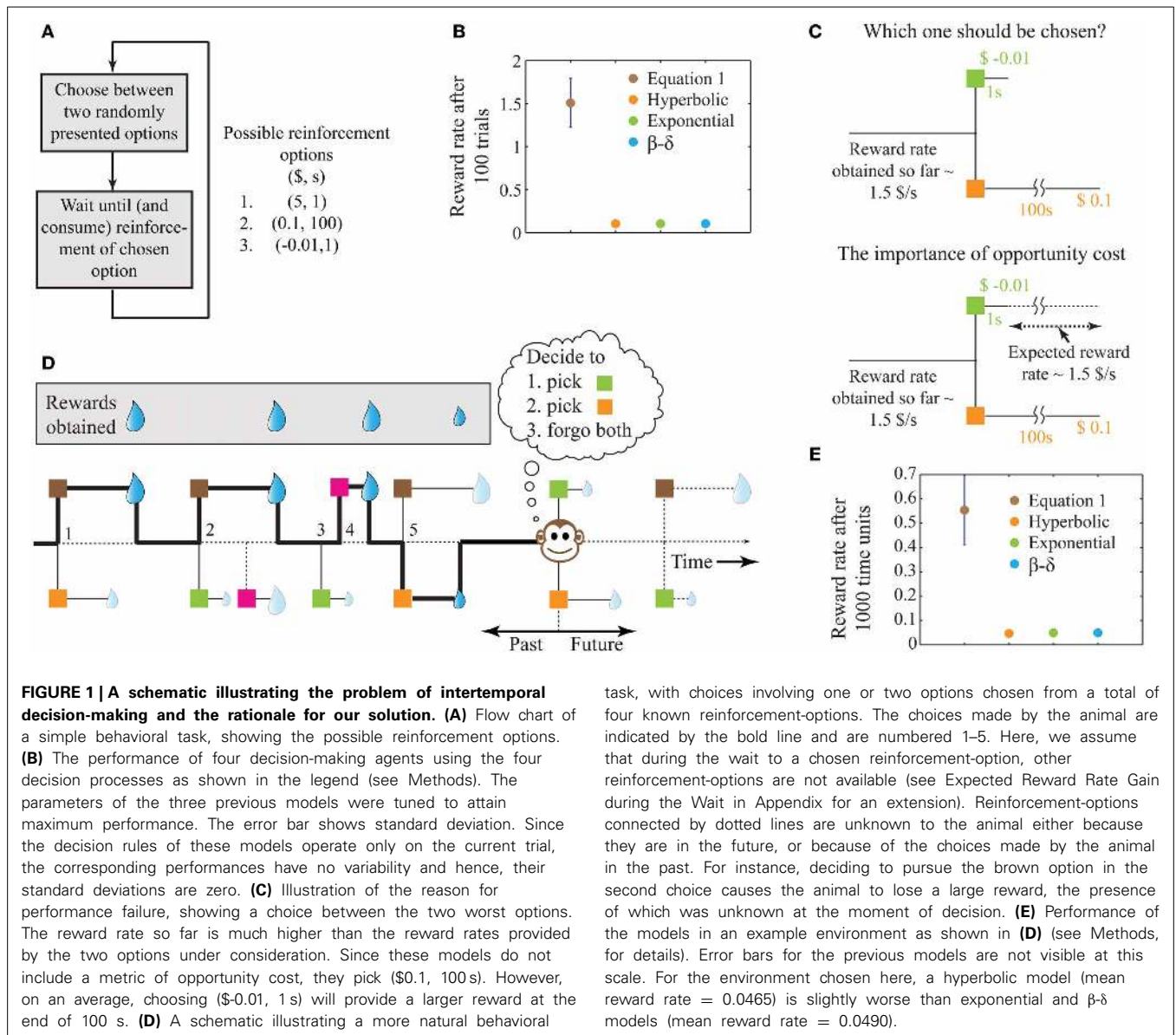
Under the above conditions, this algorithm yields the highest possible reward rate achievable at the end of any given number of trials. In contrast, previous algorithms for intertemporal decision-making (hyperbolic discounting, exponential discounting, two-parameter discounting), while being successful at fitting behavioral data, fail to maximize global reward rates. For the example reinforcement-environment shown in **Figure 1A**, simulations show that the algorithm in Equation (1) outperforms other extant algorithms by more than an order of magnitude (**Figure 1B**).

The reason why extant alternatives fare poorly is that they do not account for opportunity cost, i.e., the cost incurred in the lost opportunity to obtain better rewards than currently available. In the example considered, two of the reinforcement-options are significantly worse than the third (**Figure 1C**). Hence, in a choice between these two options, it is even worth incurring a small punishment ( $-\$0.01$ ) at a short delay for sooner opportunities of obtaining the best reward ( $\$5$ ) (**Figure 1C**). Previous models, however, pick the reward ( $\$0.1$ ) in favor of the punishment since they do not have an estimate of opportunity cost. In contrast, by storing the reinforcement history, Equation (1) accounts for the opportunity cost, and picks the punishment. Recent experimental evidence suggests that humans indeed accept small temporary costs in order to increase the opportunity for obtaining larger gains (Kolling et al., 2012).

The behavioral task shown in **Figure 1A** is similar to standard laboratory tasks studying intertemporal decisions (Frederick et al., 2002; Schweighofer et al., 2006; Kalenscher and Pennartz, 2008; Stephens, 2008). However, in naturalistic settings, animals commonly have the ability to forgo any presented option. Further, the number of options presented on a given trial can vary and could arise from a large pool of possible options. An illustration of such a task is displayed in **Figure 1D**, showing the outcomes of five past decisions. Decision 2 illustrates an instance of incurring an opportunity cost. Decision 3 shows the presentation of a single option that was forgone, leading to the presentation of a better option in decision 4. Though the options presented in decision 5 are those in decision 1, the animal’s choice behavior is the opposite, as a result of changing estimations of opportunity cost. Results of performance in such a simulated task (with no punishments) are shown in **Figure 1E**, again showing Equation (1) outperforming other models (see Methods).

## TIMERR THEORY: INTERTEMPORAL CHOICE

It is important to note that while the extent to which Equation (1) outperforms other models depends on the reinforcement-environment under consideration, its performance in a stationary environment will be greater than or equal to previous decision models. However, biological systems face at least three major constraints that limit the appropriateness of Equation (1): (1) their reinforcement-environments are non-stationary; (2) integrating reinforcement-history over arbitrarily long intervals is computationally implausible, and, (3) indefinitely long intervals without reward cannot be sustained by an animal (while maintaining fitness) even if they were to return the highest long-term reward rate (e.g., choice between 100,000 units of food in 100 days vs. 10 units of food in 0.1 day). Hence, in order to be biologically-realistic, TIMERR theory states that the interval over which reinforcement-history is evaluated, the past-integration-interval ( $T_{\text{ime}}$ ; *ime* stands for *in my experience*), is finite. Thus, the TIMERR algorithm states that animals maximize reward rates over an interval including  $T_{\text{ime}}$  and the learned expected delay to reward ( $t$ ) [Equation (2), **Figures 2A,B**]. This modification renders the decision algorithm shown in Equation (1) biologically-plausible.



**FIGURE 1 | A schematic illustrating the problem of intertemporal decision-making and the rationale for our solution. (A)** Flow chart of a simple behavioral task, showing the possible reinforcement options. **(B)** The performance of four decision-making agents using the four decision processes as shown in the legend (see Methods). The parameters of the three previous models were tuned to attain maximum performance. The error bar shows standard deviation. Since the decision rules of these models operate only on the current trial, the corresponding performances have no variability and hence, their standard deviations are zero. **(C)** Illustration of the reason for performance failure, showing a choice between the two worst options. The reward rate so far is much higher than the reward rates provided by the two options under consideration. Since these models do not include a metric of opportunity cost, they pick (\$0.1, 100 s). However, on an average, choosing (\$-0.01, 1 s) will provide a larger reward at the end of 100 s. **(D)** A schematic illustrating a more natural behavioral

task, with choices involving one or two options chosen from a total of four known reinforcement-options. The choices made by the animal are indicated by the bold line and are numbered 1–5. Here, we assume that during the wait to a chosen reinforcement-option, other reinforcement-options are not available (see Expected Reward Rate during the Wait in Appendix for an extension). Reinforcement-options connected by dotted lines are unknown to the animal either because they are in the future, or because of the choices made by the animal in the past. For instance, deciding to pursue the brown option in the second choice causes the animal to lose a large reward, the presence of which was unknown at the moment of decision. **(E)** Performance of the models in an example environment as shown in **(D)** (see Methods, for details). Error bars for the previous models are not visible at this scale. For the environment chosen here, a hyperbolic model (mean reward rate = 0.0465) is slightly worse than exponential and  $\beta$ - $\delta$  models (mean reward rate = 0.0490).

If the estimated average reward rate over the past integration window of  $T_{ime}$  is denoted by  $a_{est}$ , the TIMERR algorithm can be written as:

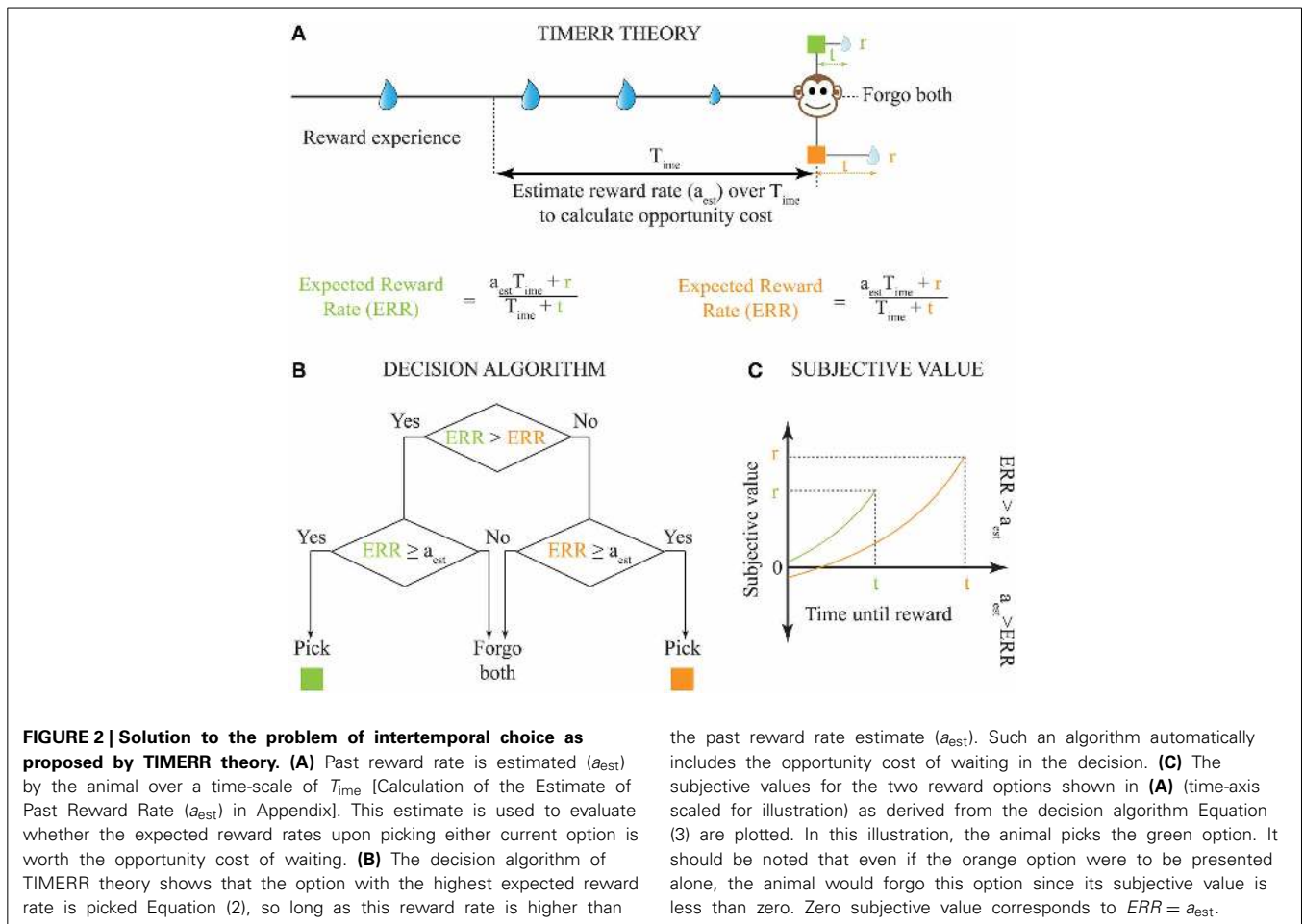
$$\text{Pick option with the highest value for } \frac{a_{est}T_{ime} + r_i}{T_{ime} + t_i} \quad (2)$$

Therefore, the TIMERR algorithm acts as a temporally-constrained, experience-based, solution to the optimization problem of maximizing reward rate. It is thus a better implementation of the statement of OFT than prior implementations. It requires that only experienced magnitudes and times of the rewards following conditioned stimuli are stored, therefore predicting that intertemporal decisions of animals will not incorporate post-reward delays due to limitations in associative learning (Kacelnik and Bateson, 1996; Stephens and Anderson, 2001; Pearson et al., 2010; Blanchard et al., 2013)

consistent with prior experimental evidence showing the insensitivity of choice behavior to post-reward delays (Stephens and Anderson, 2001; Kalenscher and Pennartz, 2008; Stephens, 2008; Pearson et al., 2010; Blanchard et al., 2013) (see Animals do not Maximize Long-Term Reward Rates in Appendix for a detailed discussion). It is important to note, however, that indirect effects of post-reward delays on behavior (Blanchard et al., 2013) can be explained as resulting from the implicit effect of post-reward delays on past reward rate; the higher the post-reward delays become, the lower will be the past reward rate.

From the TIMERR algorithm, it is possible to derive the subjective value of a delayed reward (**Figure 2C**)—defined as the amount of immediate reward that is subjectively equivalent to the delayed reward.

This is calculated by asserting that reward rate for  $(SV(r, t), 0) = \text{reward rate for } (r, t)$



**FIGURE 2 | Solution to the problem of intertemporal choice as proposed by TIMERR theory. (A)** Past reward rate is estimated ( $a_{est}$ ) by the animal over a time-scale of  $T_{ime}$  [Calculation of the Estimate of Past Reward Rate ( $a_{est}$ ) in Appendix]. This estimate is used to evaluate whether the expected reward rates upon picking either current option is worth the opportunity cost of waiting. **(B)** The decision algorithm of TIMERR theory shows that the option with the highest expected reward rate is picked Equation (2), so long as this reward rate is higher than

the past reward rate estimate ( $a_{est}$ ). Such an algorithm automatically includes the opportunity cost of waiting in the decision. **(C)** The subjective values for the two reward options shown in **(A)** (time-axis scaled for illustration) as derived from the decision algorithm Equation (3) are plotted. In this illustration, the animal picks the green option. It should be noted that even if the orange option were to be presented alone, the animal would forgo this option since its subjective value is less than zero. Zero subjective value corresponds to  $ERR = a_{est}$ .

i.e.,

$$\frac{a_{est} + \frac{SV(r, t)}{T_{ime}}}{1 + \frac{0}{T_{ime}}} = \frac{a_{est} + \frac{r}{T_{ime}}}{1 + \frac{t}{T_{ime}}}$$

where  $SV(r, t)$  is the subjective value of reward  $r$  delayed by time  $t$ . Simplifying, the expression for  $SV(r, t)$  is given by

$$SV(r, t) = \frac{r - a_{est}t}{1 + \frac{t}{T_{ime}}} \tag{3}$$

where  $a_{est}$  is an estimate of the average reward rate in the past over the integration window  $T_{ime}$  with the reward option specified by a magnitude  $r$  and a delay  $t$ .

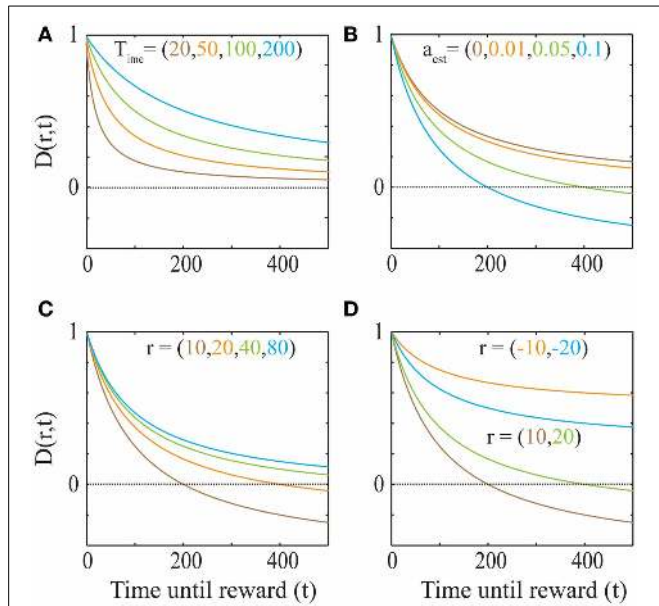
Equation (3) presents an alternative interpretation of the algorithm: the animal is estimating the net worth of pursuing each delayed reward by subtracting the opportunity cost incurred by forfeiting potential alternative reward options during the delay to a given reward and normalizing by the explicit temporal cost of waiting. This is because the numerator in Equation (3) represents the expected reward gain but subtracts this opportunity cost,  $a_{est}t$ , which corresponds to a baseline expected amount of reward that might be acquired over  $t$ . The denominator is the explicit temporal cost of waiting.

### THE TEMPORAL DISCOUNTING FUNCTION

The temporal discounting function—the ratio of subjective value to the subjective value of the reward when presented immediately—is given by [based on Equation (3)]

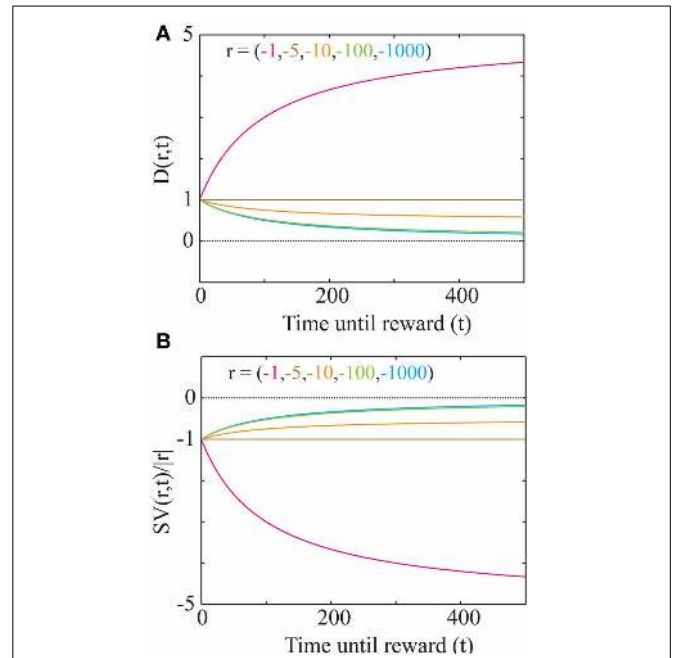
$$D(r, t) = \frac{SV(r, t)}{r} = \frac{1 - \frac{a_{est}}{r}t}{1 + \frac{t}{T_{ime}}} \tag{4}$$

This discounting function is hyperbolic with an additional, dynamical (changing with  $a_{est}$ ) subtractive term. The effects of varying the parameters, viz. the past integration interval ( $T_{ime}$ ), estimated average reward rate ( $a_{est}$ ) and reward magnitude ( $r$ ), on the discounting function are shown in **Figure 3**. The steepness of this discounting function is directly governed by  $T_{ime}$ , the past integration interval (**Figure 3A**). In other words, the longer one integrates over the past to estimate reinforcement history, the higher the tolerance to delays when considering future rewards, thus rationalizing abnormally steep discounting (characteristic of impulsivity) as resulting from abnormally low values of  $T_{ime}$ . As opportunity costs ( $a_{est}$ ) increase, delayed rewards are discounted more steeply (**Figure 3B**). Also, as the magnitude of the reward increases (**Figure 3C**), the steepness of discounting becomes lower, referred to as the “Magnitude Effect” (Myerson and Green, 1995; Frederick



**FIGURE 3 | The dependence of the discounting function on its parameters Equation (4).** (A) Explicit temporal cost of waiting: As the past integration interval ( $T_{ime}$ ) increases, the discounting function becomes less steep, i.e., the subjective value for a given delayed reward becomes higher ( $a_{est} = 0$  and  $r = 20$ ). (B) Opportunity cost affects discounting: As  $a_{est}$  increases, the opportunity cost of pursuing a delayed reward increases and hence, the discounting function becomes steeper. The dotted line indicates a subjective value of zero, below which rewards are not pursued, as is the case when the delay is too high. ( $r = 20$  and  $T_{ime} = 100$ ). (C) “Magnitude Effect”: As the reward magnitude increases, the steepness of discounting decreases (Myerson and Green, 1995; Frederick et al., 2002; Kalenscher and Pennartz, 2008) ( $T_{ime} = 100$  and  $a_{est} = 0.05$ ). (D) “Sign Effect” and differential treatment of losses: Gains (green and brown) are discounted steeper than losses (cyan and orange) of equal magnitudes (Frederick et al., 2002; Kalenscher and Pennartz, 2008) ( $T_{ime} = 100$  and  $a_{est} = 0.05$ ). Note that as the magnitude of loss decreases, so does the steepness of discounting (Figure 4). In fact, for losses with magnitudes lower than  $a_{est} T$ , the discounting function will be greater than 1, leading to a differential treatment of losses (Frederick et al., 2002; Kalenscher and Pennartz, 2008) (see text, Figure 4).

et al., 2002; Kalenscher and Pennartz, 2008) in prior experiments. Further, it is shown that gains are discounted more steeply than losses of equal magnitudes in net positive environments (Figure 3D), as shown previously and referred to as the “Sign Effect” (Frederick et al., 2002; Kalenscher and Pennartz, 2008). It must also be pointed out that the discounting function for a loss becomes steeper as the magnitude of the loss increases, observed previously as the reversal of the “Magnitude Effect” for losses (Hardisty et al., 2012) (Figure 4A). In fact, when forced to pick a punishment in a net positive environment, low-magnitude (below  $a_{est} \times T_{ime}$ ) losses will be preferred immediately while higher-magnitude losses will be preferred when delayed (Figure 4B), as has been experimentally observed (Frederick et al., 2002; Kalenscher and Pennartz, 2008; Hardisty et al., 2012) (for a full treatment of the effects of changes in variables, see Consequences of the Discounting Function in Appendix).

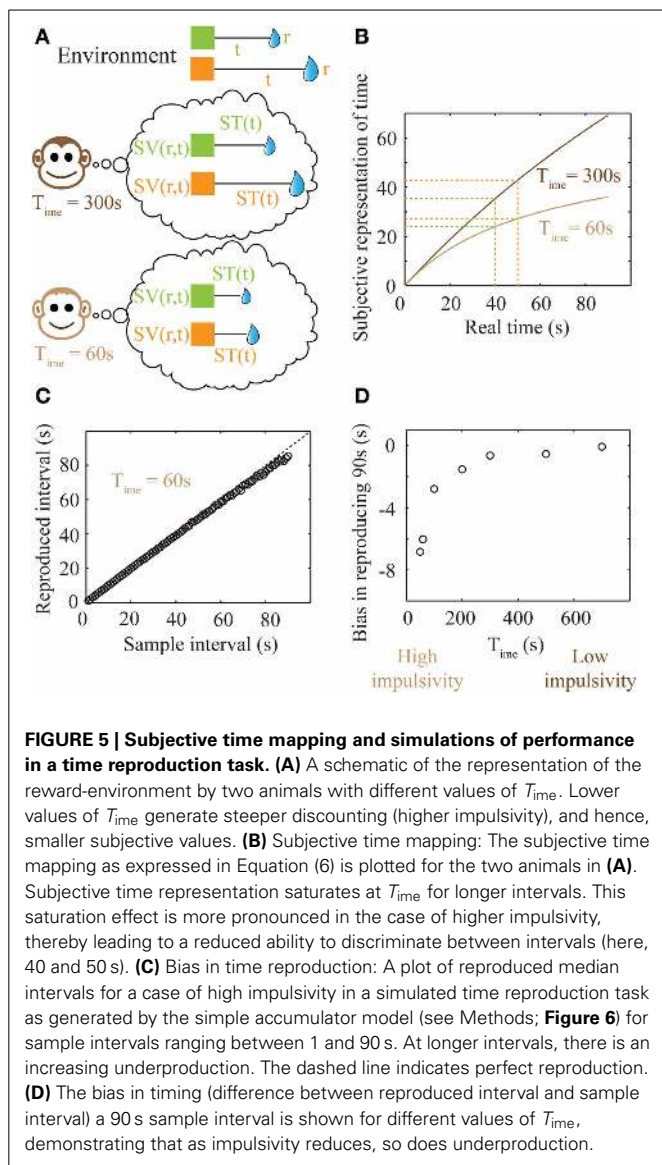


**FIGURE 4 | “Magnitude Effect” and Differential treatment of losses in a net positive environment.** (A) The discounting function plotted for losses of various magnitudes (as shown in Figure 3D;  $a_{est} = 0.05$  and  $T_{ime} = 100$ ). As the magnitude of a loss increases, the discounting function becomes steeper. However, the slope of the discounting steepness with respect to the magnitude is minimal for large magnitudes (100 and 1000; see Consequences of the Discounting Function in Appendix). At magnitudes below  $a_{est} T_{ime}$ , the discounting function becomes an increasing function of delay. (B) Plot of the signed discounting function for the magnitudes as shown in (A), showing that for magnitudes lower than  $a_{est} T_{ime}$ , a loss becomes even more of a loss when delayed. Hence, at low magnitudes ( $< a_{est} T_{ime}$ ), losses are preferred immediately. No curve crosses the dotted line at zero, showing that at all delays, losses remain punishing.

### TIMERR THEORY: TIME PERCEPTION

Attributing values to rewards delayed in time necessitates representations of those temporal delays. These representations of time are subjective, as it is known that time perception varies within and across individuals (Gibbon et al., 1997; Matell and Meck, 2000; Buhusi and Meck, 2005; Lejeune and Wearden, 2006; Wittmann and Paulus, 2008), and that errors in representation of time increase with the interval being represented (Gibbon et al., 1997; Matell and Meck, 2000; Buhusi and Meck, 2005; Lejeune and Wearden, 2006). While there are many models that address how timing may be implemented in the brain (Gibbon, 1977; Killeen and Fetterman, 1988; Matell and Meck, 2000; Buhusi and Meck, 2005; Simen et al., 2011a,b), our aim in this section is to present an “ultimate” theory of time perception, i.e., a theory of the principles behind time perception.

Since TIMERR theory states that animals seek to maximize expected reward rates, we posit that time is represented subjectively (Figure 5A) so as to result in accurate representations of changes in expected reward rate. In other words, subjective time is represented so that subjective reward rate (subjective value/subjective time) equals the true expected reward rate less



the baseline expected reward rate ( $a_{est}$ ). Hence, if the subjective representation of time associated with a delay  $t$  is denoted by  $ST(t)$ ,

$$\frac{SV(r, t)}{ST(t)} = \left(\frac{r}{t} - a_{est}\right) \tag{5}$$

Combining Equation (5) with Equation (3), we get

$$ST(t) = \frac{t}{1 + \frac{t}{T_{ime}}} \tag{6}$$

Such a representation has the property of being bounded [ $ST(\infty) = T_{ime}$ ], thereby making it possible to represent very long durations within the finite dynamic ranges of neuronal firing rates. Plots of the subjective time representation of delays between 1 and 90 s are shown in **Figure 5B** for two different values of  $T_{ime}$ . As mentioned previously (**Figure 3A**), a lower value of  $T_{ime}$  corresponds to steeper discounting, characteristic of more impulsive

decision-making. It can be seen that the difference in subjective time representations between 40 and 50 s is smaller for a lower  $T_{ime}$  (high impulsivity). Hence, higher impulsivity corresponds to a reduction in the ability to discriminate between long intervals (a decrease in the precision of time representation) (**Figures 5A,B**).

Internal time representation has been previously modeled using accumulator models (Buhusi and Meck, 2005; Simen et al., 2011a,b) that incorporate the underlying noisiness in information processing. We used a simple noisy accumulator model (see Methods, **Figure 6A**) that represents subjective time according to Equation (6) to simulate a time interval reproduction task (Buhusi and Meck, 2005; Lejeune and Wearden, 2006). In this model, we assumed that the noise in the slope of the accumulator was proportional to the square root of the signal and that there is a constant read-out noise (see Methods for details). Such noise in the accumulator slope (i.e., proportional to the square root of the signal) occurs in spiking neuronal models that assume Poisson statistics, having been used in prior accumulator models (Simen et al., 2011b). The results of time interval reproduction simulations (see Methods) are shown in **Figures 5C,D**. Lower values of  $T_{ime}$  correspond to an underproduction of time intervals (i.e., decreased accuracy of reproduction), with the magnitude of underproduction increasing with increasing durations of the sample interval (**Figure 5C**). When attempting to reproduce a 90 s sample interval, the magnitude of underproduction decreases with increases in  $T_{ime}$ , or equivalently, with decreasing impulsivity (**Figure 5D**). These predictions are supported by prior experimental evidence (Wittmann and Paulus, 2008).

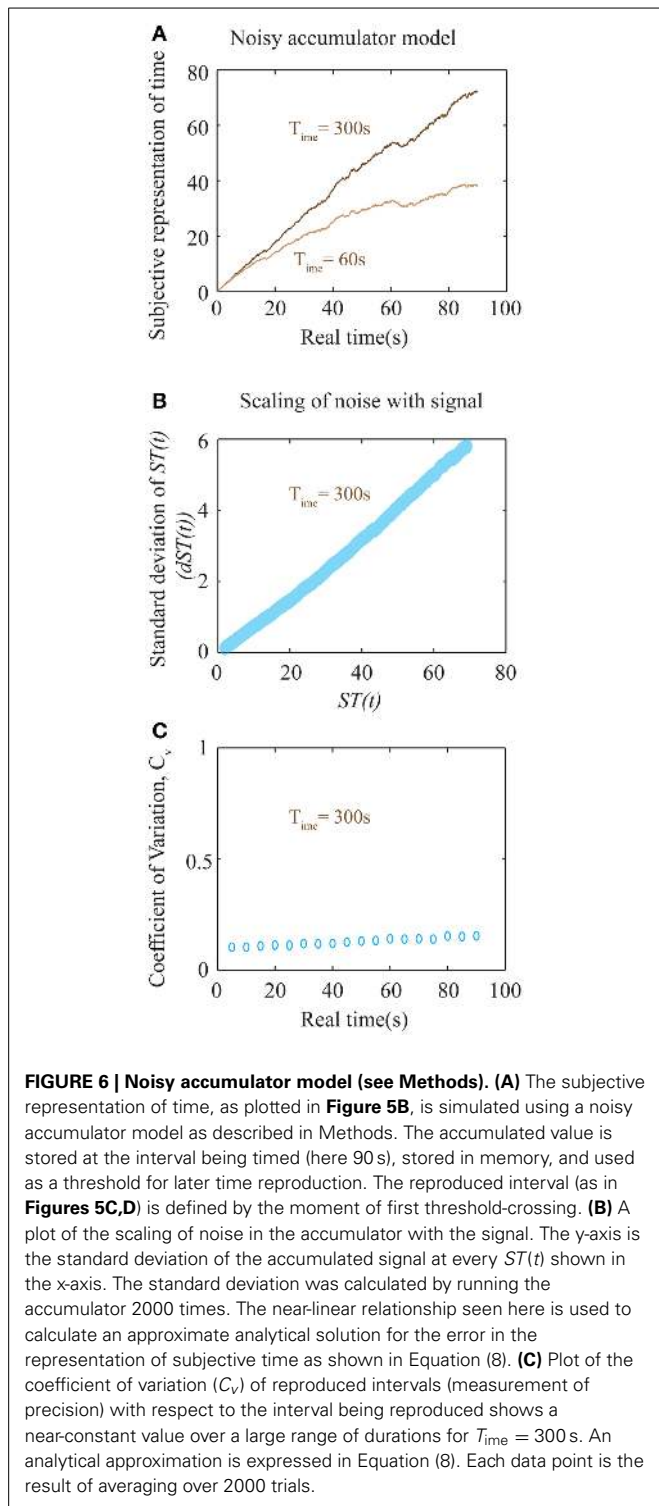
**ERRORS IN TIME PERCEPTION**

Prior studies have observed that the error in representation of intervals increases with their durations (Gibbon et al., 1997; Matell and Meck, 2000; Buhusi and Meck, 2005; Lejeune and Wearden, 2006). Such an observation is consistent with the subjective time representation presented here (**Figures 5A,B**). TIMERR theory predicts that the representation errors will be larger when  $T_{ime}$  is smaller (higher impulsivity) (**Figures 5A,B**), as observed experimentally (Wittmann et al., 2007; Wittmann and Paulus, 2008). Prior studies investigating the relationship between time duration and reproduction error have observed a linear scaling (“scalar timing”) within a limited range (Gibbon et al., 1997; Matell and Meck, 2000; Buhusi and Meck, 2005; Lejeune and Wearden, 2006).

Calculating the error in reproduced intervals by the accumulator model mentioned above cannot be done analytically. However, we present an approximate analytical solution below. Assuming that the representation of subjective time,  $ST(t)$ , has a constant infinitesimal noise of  $dST(t)$  associated with it, the noise in representation of a true interval  $t$ , denoted as  $dt$  will obey

$$\frac{dST(t)}{dt} = \frac{d}{dt} \left( \frac{t}{\left(1 + \frac{t}{T_{ime}}\right)} \right) = \frac{1}{\left(1 + \frac{t}{T_{ime}}\right)^2}$$

If one assumes that the neural noise in representing  $ST(t)$  is linearly related to the signal, with a term proportional to the signal in addition to a constant noise [i.e.,  $dST(t) = kST(t) + c$ ], then



the corresponding error in real time is

$$dt = k \frac{t}{\left(1 + \frac{t}{T_{ime}}\right)} \left(1 + \frac{t}{T_{ime}}\right)^2 + c \left(1 + \frac{t}{T_{ime}}\right)^2$$

$$= k \left(t + \frac{t^2}{T_{ime}}\right) + c \left(1 + \frac{t}{T_{ime}}\right)^2 \tag{7}$$

The coefficient of variation (error/central tendency) expected from such a model is then

$$C_v \approx \frac{dt}{t} = \frac{k \left(t + \frac{t^2}{T_{ime}}\right) + c \left(1 + \frac{t}{T_{ime}}\right)^2}{t}$$

This can be simplified as

$$C_v \approx k \left(1 + \frac{t}{T_{ime}}\right) + \frac{c \left(1 + \frac{t}{T_{ime}}\right)^2}{t} \tag{8}$$

In the above expression,  $c$  can be thought of as a constant additive noise in the memory of subjective representation of time,  $ST(t)$ , whereas the noise proportional to the signal could result from fluctuations in the slope of accumulation. In fact, for the accumulator mentioned above (that exhibits a square root dependence of the noise in slope with respect to the signal), the net relationship between the noise of the signal and the signal itself, is approximately linear (Figure 6B). Hence, our earlier assumption is a good approximation to the more realistic, yet analytically intractable, accumulator model considered above. The results of numerical simulations on  $C_v$  are shown in Figure 6C, showing a near-constant value for a large range of sample durations.

The above equation results in a U-shaped  $C_v$  curve. If the constant additive noise ( $c$ ) is small compared to the linear noise, the second term will dominate only for very low time intervals. At these very low time intervals, this will lead to a decrease in  $C_v$  as durations increase from zero. At longer intervals,  $C_v$  will appear to be a constant before a linearly increasing range. Importantly, the slope of the linear range will depend on the value of  $T_{ime}$ . Hence, though the accumulator model considered here predicts an increase in  $C_v$  at long intervals, it nonetheless will appear constant within a range determined by  $T_{ime}$ . For larger values of  $T_{ime}$ ,  $C_v$  will tend toward a constant. For the simulations shown in Figure 6C with a  $T_{ime}$  of 300 s,  $C_v$  is near constant over a very wide range of durations. While  $C_v$  is generally considered to be a constant, experimental evidence examining a wide range of sample durations analyzed across many studies (Gibbon et al., 1997; Bizo et al., 2006) accords with the specific prediction of a U-shaped coefficient of variation (spread/central tendency) for the production times Equation (8). We do note, however, that a more realistic model representing neural processing could lead to quantitative deviations from the simple approximations presented here. Such involved calculations are beyond the scope of this work. Nevertheless, the most important falsifiable prediction of our theory regarding timing is that the error in time perception will show quantitative deviations from Weber's law in impulsive subjects (with aberrantly low values of  $T_{ime}$ ). It must also be emphasized that the above equations only apply within an individual subject when  $T_{ime}$  can be assumed to be a constant, independent of the durations being tested. Pooling data across different subjects, as is common, would lead to averaging across different values of  $T_{ime}$ , and hence a flattening of the  $C_v$  curve.



**TEMPORAL BISECTION**

Time perception is also studied using temporal bisection experiments (Allan and Gibbon, 1991; Lejeune and Wearden, 2006; Baumann and Odum, 2012) in which subjects categorize a sample interval as closer to a short ( $t_s$ ) or a long ( $t_l$ ) reference interval. The sample interval at which subjects show maximum uncertainty in classification as short or long is called the point of subjective equality, or, the “bisection point.” The bisection point is of considerable theoretical interest. If subjects perceived time linearly with constant errors, the bisection point would be the arithmetic mean of the short and long intervals. On the other hand, if subjects perceived time in a scalar or logarithmic fashion or used a ratio-rule under linear mappings, it has been proposed that the bisection point would be at the geometric mean (Allan and Gibbon, 1991). However, experiments studying temporal bisection have produced ambiguous results. Specifically, the bisection point has been shown to vary between the geometric mean and the arithmetic mean and has sometimes even been shown to be below the geometric mean, closer to the harmonic mean (Killeen et al., 1997).

The bisection point as calculated by TIMERR theory is derived below. The calculation involves transforming both the short and long intervals into subjective time representations and expressing the bisection point in subjective time (subjective bisection point) as the mean of these two subjective representations. The bisection point expressed in real time is then calculated as the inverse of the subjective bisection point.

$$ST(t_s) = \frac{t_s}{1 + \frac{t_s}{T_{ime}}}; \quad ST(t_l) = \frac{t_l}{1 + \frac{t_l}{T_{ime}}}$$

Therefore, the bisection point in subjective time is given by

$$\begin{aligned} \text{Subjective bisection point (SBP)} &= \frac{ST(t_s) + ST(t_l)}{2} \\ &= \frac{\frac{t_s}{1 + \frac{t_s}{T_{ime}}} + \frac{t_l}{1 + \frac{t_l}{T_{ime}}}}{2} \end{aligned}$$

The value of the bisection point expressed in real time is given by the inverse of the subjective bisection point, viz.

$$\begin{aligned} \text{Bisection point in real time} &= \frac{SBP}{1 - \frac{SBP}{T_{ime}}} \\ &= \frac{T_{ime} \left( \frac{t_s + t_l}{2} \right) + t_s t_l}{T_{ime} + \left( \frac{t_s + t_l}{2} \right)} \quad (9) \end{aligned}$$

From the above expression, it can be seen that the bisection point can theoretically vary between the harmonic mean and the arithmetic mean as  $T_{ime}$  varies between zero and infinity, respectively.

Hence, TIMERR theory predicts that when comparing bisection points across individuals, individuals with larger values of  $T_{ime}$  will show bisection points closer to the arithmetic mean whereas individuals with smaller values of  $T_{ime}$  will show lower bisection points, closer to the geometric mean. If  $T_{ime}$  was smaller

still, the bisection point would be lower than the geometric mean, approaching the harmonic mean. This is in accordance with the experimental evidence mentioned above showing bisection points between the harmonic and arithmetic means (Allan and Gibbon, 1991; Killeen et al., 1997; Baumann and Odum, 2012). Further, we also predict that the steeper the discounting function, the lower the bisection point, as has been experimentally confirmed (Baumann and Odum, 2012). Predictions similar to ours have been made previously (Balci et al., 2011) regarding the location of the bisection point by assuming variability in temporal precision. If one assumes that impulsive subjects show larger timing errors, the previous model can also explain a reduction in the bisection point for subjects showing steeper discounting functions. However, it must be pointed out that the key contribution of our work is in deriving this result. This relationship is not an assumption in our work, but rather is an integral part of its contribution [see Equation (8) for relationship between impulsivity and  $C_v$ ].

**SUMMARY: PREDICTIONS OF TIMERR THEORY SUPPORTED BY EXPERIMENTS**

All the predictions mentioned below result from Equations (3) and (6).

1. The discounting function will be hyperbolic in form (Frederick et al., 2002; Kalenscher and Pennartz, 2008).
2. The discounting steepness could be labile within and across individuals (Loewenstein and Prelec, 1992; Frederick et al., 2002; Schweighofer et al., 2006; Luhmann et al., 2008; Van den Bos and McClure, 2013).
3. Temporal discounting could be steeper when average delays to expected rewards are lower (Frederick et al., 2002; Schweighofer et al., 2006; Luhmann et al., 2008) [see Effects of Plasticity in the Past Integration Interval ( $T_{ime}$ )].
4. “Magnitude Effect”: as reward magnitudes increase in a net positive environment, the discounting function becomes less steep (Frederick et al., 2002; Kalenscher and Pennartz, 2008) (Figure 3C).
5. “Sign Effect”: rewards are discounted steeper than punishments of equal magnitudes in net positive environments (Frederick et al., 2002; Kalenscher and Pennartz, 2008).
6. The “Sign Effect” will be larger for smaller magnitudes (Loewenstein and Prelec, 1992; Frederick et al., 2002) (see Consequences of the Discounting Function in Appendix).
7. “Magnitude Effect” for losses: as the magnitudes of losses increase, the discounting becomes steeper. This is in the reverse direction as the effect for gains (Hardisty et al., 2012). Such an effect is more pronounced for lower magnitudes (Hardisty et al., 2012) (see Consequences of the Discounting Function in Appendix).
8. Punishments are treated differently depending upon their magnitudes. Higher magnitude punishments are preferred at a delay, while lower magnitude punishments are preferred immediately (Loewenstein and Prelec, 1992; Frederick et al., 2002; Kalenscher and Pennartz, 2008) (Figure 4).
9. “Delay-Speedup” asymmetry: Delaying a reward that you have already obtained is more punishing than speeding up

- the delivery of the same reward from that delay is rewarding. This is because a received reward will be included in the current estimate of past reward rate ( $a_{est}$ ) and hence, will be included in the opportunity cost (Frederick et al., 2002; Kalenscher and Pennartz, 2008).
10. Time perception and temporal discounting are correlated (Wittmann and Paulus, 2008).
  11. Timing errors increase with the duration of intervals (Gibbon et al., 1997; Matell and Meck, 2000; Buhusi and Meck, 2005; Lejeune and Wearden, 2006).
  12. Timing errors increase in such a way that the coefficient of variation follows a U-shaped curve (Gibbon et al., 1997; Bizo et al., 2006).
  13. Impulsivity (as characterized by abnormally steep temporal discounting) leads to abnormally large timing errors (Wittmann et al., 2007; Wittmann and Paulus, 2008).
  14. Impulsivity leads to underproduction of time intervals, with the magnitude of underproduction increasing with the duration of the interval (Wittmann and Paulus, 2008).
  15. The bisection point in temporal bisection experiments will be between the harmonic and arithmetic means of the reference durations (Allan and Gibbon, 1991; Killeen et al., 1997; Baumann and Odum, 2012).
  16. The bisection point need not be constant within and across individuals (Baumann and Odum, 2012).
  17. The bisection point will be lower for individuals with steeper discounting (Baumann and Odum, 2012).
  18. The choice behavior for impulsive individuals will be more inconsistent than for normal individuals (Evenden, 1999). This is because their past reward rate estimates will show larger fluctuations due to a lower past integration interval.
  19. Post-reward delays will not be directly included in the intertemporal decisions of animals during typical laboratory tasks (Stephens and Anderson, 2001; Kalenscher and Pennartz, 2008; Stephens, 2008; Pearson et al., 2010). Variants of typical laboratory tasks may, however, lead to the inclusion of post-reward delays in decisions (Stephens and Anderson, 2001; Kalenscher and Pennartz, 2008; Stephens, 2008; Pearson et al., 2010). Post-reward delays can further indirectly affect decisions as they affect the past reward rate (Blanchard et al., 2013).

## DISCUSSION

Our theory provides a simple algorithm for decision-making in time. The algorithm of TIMERR theory, in its computational simplicity, could explain results on intertemporal choice observed across the animal kingdom (Stephens and Krebs, 1986; Frederick et al., 2002; Kalenscher and Pennartz, 2008), from insects to humans. Higher animals, of course, could evaluate subjective values with greater sophistication to build better models of the world including predictable statistical patterns of the environment and estimates of risks involved in waiting (Extensions of TIMERR Theory in Appendix). It must also be noted that other known variables influencing subjective value like satiety (Stephens and Krebs, 1986; Doya, 2008), the non-linear utility of reward magnitudes (Stephens and Krebs, 1986; Doya, 2008) and the non-linear dependence of health/fitness on reward rates

(Stephens and Krebs, 1986) have been ignored. Such factors, however, can be included as part of an extension of TIMERR theory while maintaining its inherent computational simplicity. We derived a generalized expression of subjective value that includes such additional factors Equation (A7), capturing even more variability in observed experimental results (Frederick et al., 2002; Kalenscher and Pennartz, 2008) (Non-Linearities in Subjective Value Estimation to Generalized Expression for Subjective Value in Appendix). It must also be noted that while we have ignored the effects of variability in either delays or magnitudes, explanations of such effects have previously been proposed (Gibbon et al., 1988; Kacelnik and Bateson, 1996) and are not in conflict with our theory. Also, since the exclusion of post-reward delays in decisions in TIMERR theory is borne out of limitations of associative learning, it allows for the inclusion of these delays in tasks where they can be learned. Presumably, an explicit cue indicating the end of post-reward delays could foster a representation and inclusion of these delays in decisions. Accordingly, it has been shown in recent experiments that monkeys include post-reward delays in their decisions when they are explicitly cued (Pearson et al., 2010; Blanchard et al., 2013).

In environments with time-dependent changes of reinforcement statistics, animals should have an appropriately sized past integration interval depending on the environment so as to appropriately estimate opportunity costs [e.g., integrating reward-history from the onset of winter would be highly maladaptive in order to evaluate the opportunity cost associated with a delay of an hour in the summer; also see Effects of Plasticity in the Past Integration Interval ( $T_{ime}$ ) in Appendix]. In keeping with the expectation that animals can adapt past integration intervals to their environment, it has been shown that humans can adaptively assign different weights to previous decision outcomes based on the environment (Behrens et al., 2007; Rushworth and Behrens, 2008). As Equations (3) and (4) show (Figure 3A), changes in  $T_{ime}$  would correspondingly affect the steepness of discounting. This novel prediction has two major implications for behavior: (1) the discounting steepness of an individual need not be a constant, as has sometimes been implied in prior literature (Frederick et al., 2002); (2) the longer the past integration interval, the higher the tolerance to delays when considering future rewards. In accordance with the former prediction, several recent reviews have suggested that discounting rates are variable within and across individuals (Loewenstein and Prelec, 1992; Frederick et al., 2002; Schweighofer et al., 2006; Luhmann et al., 2008; Van den Bos and McClure, 2013). The latter prediction states that impulsivity (Evenden, 1999), as characterized by abnormally steep discounting, could be the result of abnormally short windows of past reward rate integration. This may explain the observation that discounting becomes less steep as individuals develop in age (Peters and Büchel, 2011), should the longevity of memories increase over development. Past integration intervals could also be related to and bounded by the span of working memory. In fact, recent studies have shown that working memory and temporal discounting are correlated within subjects (Shamosh et al., 2008; Bickel et al., 2011) and also that improving working memory capacity decreases the steepness of discounting in

stimulant addicts (Bickel et al., 2011). Further, Equation (6) states that changes in  $T_{\text{ime}}$  would lead to corresponding changes in subjective representations of time. Hence, we predict that perceived durations may be linked to experienced reward environments, i.e., “time flies when you’re having fun.”

It is important to point out that the TIMERR algorithm for decision-making only depends on the calculation of the expected reward rate, as shown in **Figure 2B**. While this algorithm is mathematically equivalent to picking the option with the highest subjective value Equation (3), the discounting of delayed rewards results purely from the effect of those delays on the expected reward rate. Hence, as has been previously proposed (Pearson et al., 2010; Blanchard et al., 2013), we do not think of the discounting steepness as a psychological constant of an individual. Instead, we posit that apparent discounting functions are the consequence of maximizing temporally-constrained expected reward rates, and that abnormalities in temporal discounting result from abnormal adaptations of  $T_{\text{ime}}$ .

Reward magnitudes and delays have been shown to be represented by neuromodulatory and cortical systems (Platt and Glimcher, 1999; Shuler and Bear, 2006; Kobayashi and Schultz, 2008), while neurons integrating cost and benefit to represent subjective values have also been observed (Kalenscher et al., 2005; Kennerley et al., 2006). Recent reward rate estimation ( $a_{\text{est}}$ ) has been proposed to be embodied by dopamine levels over long time-scales (Niv et al., 2007). Interestingly, it has been shown that administration of dopaminergic agonists (antagonists) leads to underproduction (overproduction) (Matell et al., 2006) of time intervals, consistent with a relationship between recent reward rate estimation and subjective time representation as proposed here. Average values of foraging environment have also been shown to be represented in the anterior cingulate cortex (Kolling et al., 2012). In light of these experimental observations neurobiological models have previously proposed that decisions, similar to our theory, result from the net balance between values of the options currently under consideration and the environment as a whole (Kennerley et al., 2006; Kolling et al., 2012). However, these models do not propose that the effective interval ( $T_{\text{ime}}$ ) over which average reward rates are calculated directly determines the steepness of temporal discounting.

While there have been previous models that connect time perception to temporal decision making (Staddon and Cerutti, 2003; Takahashi, 2006; Balci et al., 2011; Ray and Bossaerts, 2011), TIMERR theory is the first unified theory of intertemporal choice and time perception to capture such a wide array of experimental observations including, but not limited to, hyperbolic discounting (Stephens and Krebs, 1986; Stephens and Anderson, 2001; Frederick et al., 2002; Kalenscher and Pennartz, 2008), “Magnitude” (Myerson and Green, 1995; Frederick et al., 2002; Kalenscher and Pennartz, 2008) and “Sign” effects (Frederick et al., 2002; Kalenscher and Pennartz, 2008), differential treatment of losses (Frederick et al., 2002; Kalenscher and Pennartz, 2008), as well as correlations between temporal discounting, time perception (Wittmann and Paulus, 2008), and timing errors (Gibbon et al., 1997; Matell and Meck, 2000; Buhusi and Meck, 2005; Lejeune and Wearden, 2006; Wittmann et al., 2007; Wittmann and Paulus, 2008) (see “Summary” for a full list).

While the notion of opportunity cost long precedes TIMERR, TIMERR’s unique contribution is in stating that the past integration interval over which opportunity cost is estimated directly determines the steepness of temporal discounting and the non-linearity of time perception. This is the major falsifiable prediction of TIMERR. As a direct result, TIMERR theory suggests that the spectra of aberrant timing behavior seen in cognitive/behavioral disorders (Buhusi and Meck, 2005; Wittmann et al., 2007; Wittmann and Paulus, 2008) (Parkinson’s disease, schizophrenia, and stimulant addiction) can be rationalized as a consequence of aberrant integration over experienced reward history. Hence, TIMERR theory has major implications for the study (see Implications for Intertemporal Choice in Appendix) of decision-making in time and time perception in normal and clinical populations.

## METHODS

All simulations were run using MATLAB R2010a.

### SIMULATIONS FOR FIGURE 1

**Figure 1B:** Each of the four decision-making agents ran a total of 100 trials. This was repeated 10 times to get the mean and standard deviation. Every trial consisted of the presentation of two reinforcement-options randomly chosen from the three possible alternatives as shown in **Figure 1A**.

**Figure 1E:** The following four possible reward-options were considered, expressed as  $(r, t)$ : (0.1, 100), (0.0001, 2), (5, 2), (5, 150). The units are arbitrary. To create the reinforcement-environment, a Poisson-process was generated for the availability-times of each of the four options. These times were binned into bins of size 1 unit, such that each time bin could consist of zero to four reward-options. The rate of occurrence for each option was set equally to 0.2 events/unit of time. For the three previous decision-making models, the parameters were tuned for maximum performance by trial and error. Forgoing an available reward-option was not possible for these models since their subjective values are always greater than zero for rewards.

### SIMULATIONS FOR FIGURES 5, 6

An accumulator model described by the following equation was used for simulations of a time reproduction task.

$$dST(t) = \frac{dt}{\left(1 + \frac{t}{T_{\text{ime}}}\right)^2} + \sigma \sqrt{ST(t)} dW_t$$

where  $W_t$  is a standard Wiener process and  $\sigma$  is the magnitude of the noise.  $\sigma$  was set to 10%. Without the noise term in the R.H.S, this equation is consistent with the subjective time expression shown in Equation (6) since integrating for  $ST(t)$  exactly yields Equation (6). This equation can also be rewritten to be in terms of  $ST(t)$  as below.

$$dST(t) = \left(1 - \frac{ST(t)}{T_{\text{ime}}}\right)^2 dt + \sigma \sqrt{ST(t)} dW_t$$

The above equation was integrated using the Euler-Maruyama method. In this method,  $ST(t)$  is updated using the following

equation for a random walk

$$ST(t + \Delta t) = ST(t) + \left(1 - \frac{ST(t)}{T_{\text{time}}}\right)^2 \Delta t + \sigma \sqrt{ST(t)} \sqrt{\Delta t} N(0, 1)$$

where  $N(0, 1)$  is the standard normal distribution. The step size for integration,  $\Delta t$ , was set so that there were 1000 steps for every simulated duration in the time interval reproduction task (Figures 5, 6).

Every trial in the time reproduction task consisted of two phases: a time measurement phase and a time production phase. During the time measurement phase, the accumulator integrates subjective time until the expiration of the sample duration (Figure 6A). The subjective time value at the end of the sample duration is stored in memory after the addition of a constant Gaussian noise as the threshold for time production, i.e.,

$$\text{Threshold}(t) = ST(t) + c N(0, 1)$$

During the time production phase, the accumulator integrates subjective time until the threshold is crossed for the first time. This moment of first crossing represents the action response indicating the end of the sample duration, i.e.,

$$\text{Reproduced interval} = t : ST(t) \geq \text{Threshold}(t)$$

For the simulations resulting in Figures 5C,D, 6,  $\sigma = 0.1$  and  $c = 0.001$ . For Figure 5C, sample interval durations ranged between 1 and 90 s over bins of 1 s. A total of 2000 trials were performed for each combination of sample duration and  $T_{\text{time}}$  to calculate the median production interval as shown in Figures 5C,D. While calculating the moment of reproduction, the integration was carried out up to a maximum time equaling 10 times the sample duration.

## AUTHOR CONTRIBUTIONS

Vijay M. K. Namboodiri, Stefan Mihalas, and Marshall G. Hussain Shuler conceived of the study. Vijay M. K. Namboodiri and Stefan Mihalas developed TIMERR theory and its extensions. Vijay M. K. Namboodiri ran the simulations comparing the performance of Equation (1) with other models shown in Figures 1B,E and simulations of the time interval reproduction task used to generate Figures 5, 6. Tanya M. Marton and Marshall G. Hussain Shuler were involved in intellectual discussions throughout the work. Vijay M. K. Namboodiri wrote the manuscript with assistance from Stefan Mihalas, Tanya M. Marton, and Marshall G. Hussain Shuler.

## ACKNOWLEDGMENTS

We thank Dr. Peter Holland, Dr. Veit Stuphorn, Dr. James Knierim, Dr. Emma Roach, Dr. Camila Zold, Josh Levy, Gerald Sun, Grant Gillary, Jeffrey Mayse, Naween Anand, Arnab Kundu, and Kyle Severson for discussions and comments on the manuscript. This work was funded by NIMH (R01 MH084911 and R01 MH093665) to Marshall G. Hussain Shuler.

## REFERENCES

- Alcock, J., and Sherman, P. (1994). The utility of the proximate-ultimate dichotomy in ethology. *Ethology* 96, 58–62. doi: 10.1111/j.1439-0310.1994.tb00881.x
- Allan, L. G., and Gibbon, J. (1991). Human bisection at the geometric mean. *Learn. Motiv.* 22, 39–58. doi: 10.1016/0023-9690(91)90016-2
- Balci, F., Freestone, D., Simen, P., Desouza, L., Cohen, J. D., and Holmes, P. (2011). Optimal temporal risk assessment. *Front. Integr. Neurosci.* 5:56. doi: 10.3389/fnint.2011.00056
- Bateson, M., and Kacelnik, A. (1996). Rate currencies and the foraging starting: the fallacy of the averages revisited. *Behav. Ecol.* 7, 341–352. doi: 10.1093/beheco/7.3.341
- Baumann, A. A., and Odum, A. L. (2012). Impulsivity, risk taking, and timing. *Behav. Processes* 90, 408–414. doi: 10.1016/j.beproc.2012.04.005
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., and Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nat. Neurosci.* 10, 1214–1221. doi: 10.1038/nn1954
- Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F., and Baxter, C. (2011). Remember the future: working memory training decreases delay discounting among stimulant addicts. *Biol. Psychiatry* 69, 260–265. doi: 10.1016/j.biopsych.2010.08.017
- Bizo, L. A., Chu, J. Y. M., Sanabria, F., and Killeen, P. R. (2006). The failure of Weber's law in time perception and production. *Behav. Processes* 71, 201–210. doi: 10.1016/j.beproc.2005.11.006
- Blanchard, T. C., Pearson, J. M., and Hayden, B. Y. (2013). Postreward delays and systematic biases in measures of animal temporal discounting. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15491–15496. doi: 10.1073/pnas.1310446110
- Buhusi, C. V., and Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nat. Rev. Neurosci.* 6, 755–765. doi: 10.1038/nrn1764
- Doya, K. (2008). Modulators of decision making. *Nat. Neurosci.* 11, 410–416. doi: 10.1038/nn2077
- Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology (Berl.)* 146, 348–361. doi: 10.1007/PL00005481
- Frederick, S., Loewenstein, G., Donoghue, T. O., and Donoghue, T. E. D. O. (2002). Time discounting and time preference: a critical review. *J. Econ. Lit.* 40, 351–401. doi: 10.1257/002205102320161311
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* 84, 279–325. doi: 10.1037/0033-295X.84.3.279
- Gibbon, J., Church, R. M., Fairhurst, S., and Kacelnik, A. (1988). Scalar expectancy theory and choice between delayed rewards. *Psychol. Rev.* 95, 102–114. doi: 10.1037/0033-295X.95.1.102
- Gibbon, J., Malapani, C., Dale, C. L., and Gallistel, C. R. (1997). Toward a neurobiology of temporal cognition: advances and challenges. *Curr. Opin. Neurobiol.* 7, 170–184. doi: 10.1016/S0959-4388(97)80005-0
- Hardisty, D. J., Appelt, K. C., and Weber, E. U. (2012). Good or bad, we want it now: fixed-cost present bias for gains and losses explains magnitude asymmetries in intertemporal choice. *J. Behav. Decis. Making.* 26, 348–361. doi: 10.1002/bdm.1771
- Kacelnik, A., and Bateson, M. (1996). Risky theories—the effects of variance on foraging decisions. *Integr. Comp. Biol.* 36, 402–434. doi: 10.1093/icb/36.4.402
- Kalenscher, T., and Pennartz, C. M. A. (2008). Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making. *Prog. Neurobiol.* 84, 284–315. doi: 10.1016/j.pneurobio.2007.11.004
- Kalenscher, T., Windmann, S., Diekamp, B., Rose, J., Güntürkün, O., and Colombo, M. (2005). Single units in the pigeon brain integrate reward amount and time-to-reward in an impulsive choice task. *Curr. Biol.* 15, 594–602. doi: 10.1016/j.cub.2005.02.052
- Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J., and Rushworth, M. F. S. (2006). Optimal decision making and the anterior cingulate cortex. *Nat. Neurosci.* 9, 940–947. doi: 10.1038/nn1724
- Killeen, P. R., and Fetterman, J. G. (1988). A behavioral theory of timing. *Psychol. Rev.* 95, 274–295. doi: 10.1037/0033-295X.95.2.274
- Killeen, P. R., Fetterman, J. G., and Bizo, L. (1997). "Time's causes," in *Advances in Psychology: Time and Behavior: Psychological and Neurobehavioral Analyses*, eds C. M. Bradshaw and E. Szabadi (Amsterdam: Elsevier), 79–131.
- Kobayashi, S., and Schultz, W. (2008). Influence of reward delays on responses of dopamine neurons. *J. Neurosci.* 28, 7837–7846. doi: 10.1523/JNEUROSCI.1600-08.2008
- Kolling, N., Behrens, T. E. J., Mars, R. B., and Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science* 336, 95–98. doi: 10.1126/science.1216930

- Lejeune, H., and Wearden, J. H. (2006). Scalar properties in animal timing: conformity and violations. *Q. J. Exp. Psychol.* 59, 1875–1908. doi: 10.1080/17470210600784649
- Loewenstein, G., and Prelec, D. (1992). Anomalies in intertemporal choice: evidence and an interpretation. *Q. J. Econ.* 107, 573–597. doi: 10.2307/2118482
- Luhmann, C. C., Chun, M. M., Yi, D.-J., Lee, D., and Wang, X.-J. (2008). Neural dissociation of delay and uncertainty in intertemporal choice. *J. Neurosci.* 28, 14459–14466. doi: 10.1523/JNEUROSCI.5058-08.2008
- Matell, M. S., Bateson, M., and Meck, W. H. (2006). Single-trials analyses demonstrate that increases in clock speed contribute to the methamphetamine-induced horizontal shifts in peak-interval timing functions. *Psychopharmacology (Berl.)* 188, 201–212. doi: 10.1007/s00213-006-0489-x
- Matell, M. S., and Meck, W. H. (2000). Neuropsychological mechanisms of interval timing behavior. *Bioessays* 22, 94–103. doi: 10.1002/(SICI)1521-1878(200001)22:1<94::AID-BIES14>3.0.CO;2-E
- Myerson, J., and Green, L. (1995). Discounting of delayed rewards: models of individual choice. *J. Exp. Anal. Behav.* 64, 263–276. doi: 10.1901/jeab.1995.64-263
- Niv, Y., Daw, N. D., Joel, D., and Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl.)* 191, 507–520. doi: 10.1007/s00213-006-0502-4
- Pearson, J. M., Hayden, B. Y., and Platt, M. L. (2010). Explicit information reduces discounting behavior in monkeys. *Front. Psychol.* 1:237. doi: 10.3389/fpsyg.2010.00237
- Peters, J., and Büchel, C. (2011). The neural mechanisms of inter-temporal decision-making: understanding variability. *Trends Cogn. Sci.* 15, 227–239. doi: 10.1016/j.tics.2011.03.002
- Platt, M. L., and Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature* 400, 233–238. doi: 10.1038/22268
- Ray, D., and Bossaerts, P. (2011). Positive temporal dependence of the biological clock implies hyperbolic discounting. *Front. Neurosci.* 5:2. doi: 10.3389/fnins.2011.00002
- Rushworth, M. F. S., and Behrens, T. E. J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat. Neurosci.* 11, 389–397. doi: 10.1038/nn2066
- Samuelson, P. A. (1937). A note on measurement of utility. *Rev. Econ. Stud.* 4, 155–161. doi: 10.2307/2967612
- Schweighofer, N., Shishida, K., Han, C. E., Okamoto, Y., Tanaka, S. C., Yamawaki, S., et al. (2006). Humans can adopt optimal discounting strategy under real-time constraints. *PLoS Comput. Biol.* 2:e152. doi: 10.1371/journal.pcbi.0020152
- Shamosh, N. A., Deyoung, C. G., Green, A. E., Reis, D. L., Johnson, M. R., Conway, A. R. A., et al. (2008). Individual differences in delay discounting: relation to intelligence, working memory, and anterior prefrontal cortex. *Psychol. Sci.* 19, 904–911. doi: 10.1111/j.1467-9280.2008.02175.x
- Shuler, M. G., and Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science* 311, 1606–1609. doi: 10.1126/science.1123513
- Simen, P., Balci, F., Desouza, L., Cohen, J. D., and Holmes, P. (2011a). Interval timing by long-range temporal integration. *Front. Integr. Neurosci.* 5:28. doi: 10.3389/fnint.2011.00028
- Simen, P., Balci, F., de Souza, L., Cohen, J. D., and Holmes, P. (2011b). A model of interval timing by neural integration. *J. Neurosci.* 31, 9238–9253. doi: 10.1523/JNEUROSCI.3121-10.2011
- Staddon, J. E. R., and Cerutti, D. T. (2003). Operant conditioning. *Annu. Rev. Psychol.* 54, 115–144. doi: 10.1146/annurev.psych.54.101601.145124
- Stephens, D. W. (2008). Decision ecology: foraging and the ecology of animal decision making. *Cogn. Affect. Behav. Neurosci.* 8, 475–484. doi: 10.3758/CABN.8.4.475
- Stephens, D. W., and Anderson, D. (2001). The adaptive value of preference for immediacy: when shortsighted rules have farsighted consequences. *Behav. Ecol.* 12, 330–339. doi: 10.1093/beheco/12.3.330
- Stephens, D. W., and Krebs, J. R. (1986). *Foraging Theory*. Princeton, NJ: Princeton University Press.
- Takahashi, T. (2006). Time-estimation error following Weber-Fechner law may explain subadditive time-discounting. *Med. Hypotheses* 67, 1372–1374. doi: 10.1016/j.mehy.2006.05.056
- Van den Bos, W., and McClure, S. M. (2013). Towards a general model of temporal discounting. *J. Exp. Anal. Behav.* 99, 58–73. doi: 10.1002/jeab.6
- Wittmann, M., Leland, D. S., Churan, J., and Paulus, M. P. (2007). Impaired time perception and motor timing in stimulant-dependent subjects. *Drug Alcohol Depend.* 90, 183–192. doi: 10.1016/j.drugalcdep.2007.03.005
- Wittmann, M., and Paulus, M. P. (2008). Decision making, impulsivity and time perception. *Trends Cogn. Sci.* 12, 7–12. doi: 10.1016/j.tics.2007.10.004

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 November 2013; paper pending published: 23 January 2014; accepted: 12 February 2014; published online: 28 February 2014.

Citation: Nambodiri VMK, Mihalas S, Marton TM and Hussain Shuler MG (2014) A general theory of intertemporal decision-making and the perception of time. *Front. Behav. Neurosci.* 8:61. doi: 10.3389/fnbeh.2014.00061

This article was submitted to the journal *Frontiers in Behavioral Neuroscience*.

Copyright © 2014 Nambodiri, Mihalas, Marton and Hussain Shuler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

**APPENDIX**

**RESULTS**

**Extensions of TIMERR theory**

**Alternative version (store reward rates evaluated upon the receipt of reward in memory).** An alternative version of TIMERR theory could be appropriate for very simple life forms with limited computational resources that are capable of intertemporal decision making (e.g., insects). Rather than representing both the magnitude and delay to rewards separately and making decisions based on real time calculations, upon the receipt of reward, such animals could store subjective value directly in memory. In such a case, the reward rate at the time of reward receipt would be calculated over  $T_{ime} + t$  and converted to subjective value. The decision between reward options is then simply described as picking the option with the highest stored subjective value. Mathematically, such a calculation is exactly equivalent to the calculation presented in Extensions of TIMERR Theory in Appendix.

While the advantage of this model is that it is computationally less expensive, the disadvantages for the model are that (1) subjective values in memory are not generalizable, i.e., the subjective value in memory for an option will fundamentally depend on the reward environment in which it was presented; and (2) representations of the reward delays could be useful for anticipatory behaviors.

**Evaluation of risk.** Until now, we have assumed that a delayed reward will be available for consumption, provided the animal waits the delay, i.e., there are no explicit risks in obtaining the reward. In many instances in nature, however, such an assumption is not true. If the animal could build a model of the risks involved in obtaining a delayed reward, it could do better by including such a model in its decision making. Given information about a delayed reward ( $r, t$ ), if the animal could predict the expected reward available for consumption after having waited the delay [ $ER(r, t)$ ], the subjective value of such a reward could be written as

$$SV(r, t) = \frac{ER(r, t) - a_{est}t}{1 + \frac{t}{T_{ime}}} \tag{A1}$$

This is based on Equation (3).

It is important to note that this equation can still be expressed in terms of subjective time as defined in the Main Text, viz.

$$SV(r, t) = \left( \frac{ER(r, t)}{t} - a_{est} \right) \frac{t}{1 + \frac{t}{T_{ime}}}$$

Generally speaking, building such risk models is difficult, especially since they are environment-specific. However, there could be statistical patterns in environments for which animals have acquired corresponding representations over evolution. Specifically, decay of rewards arising from factors like natural decay (rotting, for instance) or due to competition from other foragers could have statistical patterns. During the course of travel to a food source, competition poses the strongest cause for decay since natural decay typically happens over a longer time-scale, viz.

days to months. In such an environment with competition from other foragers, a forager could estimate how much a reward will decay in the time it takes it to travel to the food source.

Suppose the forager sees a reward of magnitude  $r$  at time  $t = 0$ , the moment of decision. The aim of the forager is to calculate how much value will be left by the time it reaches the food source, and to use this estimate in its current decision. Let us denote the time taken by the forager to travel to the food source by  $t$ .

We assume that the rate of decay of a reward in competition is proportional to a power of its magnitude, implying that larger rewards are more sought-after in competition and hence, would decay at a faster rate. We denote the survival time of a typical reward by  $t_{sur}$  and consider that after time  $t_{sur}$ , the reward is entirely consumed. If, as stated above, one assumes that  $t_{sur}$  is inversely related to a power  $\alpha$  of the magnitude of a reward at any time [ $r(t)$ ], we can write that  $t_{sur} = \frac{1}{kr(t)^\alpha}$  where  $k$  is a constant of proportionality.

Hence, the rate of change of a value with initial magnitude  $r$ , will be

$$\frac{dr(t)}{dt} = -\frac{r(t)}{t_{sur}} = -(kr(t)^\alpha) r(t)$$

Solving this differential equation for  $r(t)$ ,

$$r(t) = \frac{r}{(1 + k\alpha r^\alpha t)^{1/\alpha}}$$

Here we set  $r(0) = r$ .

A forager could estimate the parameters  $k$  and  $\alpha$  based on the density of competition and other properties of the environment. In such a case, the subjective value of a delayed reward ( $r, t$ ) should be calculated as

$$SV(r, t) = \frac{\frac{r}{(1 + k\alpha r^\alpha t)^{1/\alpha}} - a_{est}t}{1 + \frac{t}{T_{ime}}} \tag{A2}$$

The discounting function in this case is

$$D(r, t) = \frac{1}{(1 + k\alpha r^\alpha t)^{1/\alpha}} - \frac{a_{est}t}{r} \tag{A3}$$

Such a discounting function can be thought of as a quasi-hyperbolic discounting function, and is a more general form than Equation (3) since  $k = 0$  returns Equation (3).

**Non-linearities in subjective value estimation.** Animals do not perceive rewards linearly (e.g., 20 L of juice is not 100 times more valuable than 200 mL). Non-linear reward perception may reflect the non-linear utility of rewards: too little is often insufficient while too much is unnecessary. Further, the value of a reward depends on the internal state of an animal (e.g., 200 mL of juice is more valuable to a thirsty animal than a satiated animal). We address such non-linearities as applied to TIMERR theory here.

If the non-linearities and state-dependence of magnitude perception can be expressed by a function  $f(r, \text{state})$ , then this function can be incorporated into Equation (3) to give

$$SV(r, t) = \frac{f(r, \text{state}) - a_{\text{est}}t}{1 + \frac{t}{T_{\text{ime}}}} \quad (\text{A4})$$

The introduction of such state-dependence and non-linearities may account for the anomalous “preference for spread” (Frederick et al., 2002; Kalenscher and Pennartz, 2008) and “preference for improving sequences” (Frederick et al., 2002; Kalenscher and Pennartz, 2008) seen in human choice behavior.

**Expected reward rate gain during the wait.** We have not yet considered the possibility that animals could expect to receive additional rewards during the wait to delayed rewards, i.e., while animals expect to lose an average reward rate of  $a_{\text{est}}$  during the wait, there could be a different reward rate that they might, nevertheless, expect to gain. If we denote that this additional expected reward rate is a fraction  $f$  of  $a_{\text{est}}$ , then we can state that the net expected loss of reward rate during the wait is  $(1 - f)a_{\text{est}}$ . This factor can also be added to expressions of subjective value calculated above in Equations (3), (A2), and (A4). Specifically, Equation (3) becomes

$$SV(r, t) = \frac{r - (1 - f)a_{\text{est}}t}{1 + \frac{t}{T_{\text{ime}}}} \quad (\text{A5})$$

Such a factor is especially important in understanding prior human experiments. In abstract questions like “\$100 now or \$150 a month from now?” human subjects expect an additional reward rate during the month and are almost certainly not making decisions with the assumption that the only reward they can obtain during the month is \$150.

**State-dependence of discounting steepness.** In the basic version of TIMERR theory, the time window over which the algorithm aims to maximize reward rates is the past integration interval ( $T_{\text{ime}}$ ) plus the time to a delayed reward. However, non-linearities in the relationship between reward rates and fitness levels [as discussed in Effects of Plasticity in the Past Integration Interval ( $T_{\text{ime}}$ ) in Appendix] could lead to state-dependent consumption requirements. For example, in a state of extreme hunger, it might be appropriate for the decision rule to apply a very short time scale of discounting so as to avoid dangerously long delays to food. However, integrating past reward rates over such extremely short timescales could compromise the reliability of the estimated reward rate. Hence, as a more general version of TIMERR theory, the window over which reward rate is maximized could incorporate a scaled down value of the interval over which past reward rate is estimated, with the scaling factor governed by consumption requirements. If such a scaling factor is represented by  $s(\text{state})$ , Equation (3) would become

$$SV(r, t) = \frac{r - a_{\text{est}}t}{1 + \frac{t}{T_{\text{ime}}s(\text{state})}} \quad (\text{A6})$$

**Generalized expression for subjective value.** Combining Equations (A2), (A4)–(A6), we can write a more general expression for the subjective value of a delayed reward, including a model of risk along with additional reward rates, state dependences, and non-linearities in the perception of reward magnitude

$$SV(r, t) = \frac{\frac{f(r, \text{state})}{(1 + k\alpha f(r, \text{state})^\alpha t)^{1/\alpha}} - (1 - f)a_{\text{est}}t}{1 + \frac{t}{T_{\text{ime}}s(\text{state})}} \quad (\text{A7})$$

Equation (A7) is a more complete expression for the subjective value of delayed rewards. Such an expression could capture almost the entirety of experimental results, considering its inherent flexibility. However, it should be noted that even with as simple an expression as Equation (3), many observed experimental results can be explained.

**DISCUSSION**

**Implications for intertemporal choice**

**Consequences of the discounting function.** We rewrite Equation (4) below followed by its implications for intertemporal choice in environments with positive and negative past reward rate estimates.

$$D(r, t) = \frac{SV(r, t)}{r} = \frac{1 - \frac{a_{\text{est}}t}{r}}{1 + \frac{t}{T_{\text{ime}}}}$$

In an environment with positive  $a_{\text{est}}$ , the following predictions can be made

1. “Magnitude Effect” for gains: as noted in the Main Text, as  $r$  increases, the numerator increases in value, effectively making the discounting less steep (Figure 3C). This effect has been experimentally observed and has been referred to as the “magnitude” effect (Frederick et al., 2002; Kalenscher and Pennartz, 2008). TIMERR theory makes a further prediction, however, that the size of the “magnitude” effect will depend on the size of  $a_{\text{est}}$  and  $t$ . Specifically, as  $a_{\text{est}}$  and  $t$  increase, so does the size of the effect.
2. “Magnitude Effect” for losses/punishments: if  $r$  is negative (i.e., loss/punishment), the discounting function will become more steep as the magnitude of  $r$  increases (Figures 3D, 4). Hence, in a rewarding environment ( $a_{\text{est}} > 0$ ), the “magnitude” effect for punishments is in the opposite direction as the “magnitude” effect for gains.
3. “Sign Effect”: gains are discounted more steeply than punishments of equal magnitudes. A further prediction is that this effect will be larger for smaller reward magnitudes. This prediction has been proven experimentally (Loewenstein and Prelec, 1992; Frederick et al., 2002).
4. Differential treatment of losses/punishments: As the “magnitude” of the punishment decreases below  $a_{\text{est}}T_{\text{ime}}(r > -a_{\text{est}}T_{\text{ime}})$ , the discounting function becomes a monotonically increasing function of delay. This means that the punishment would be preferred immediately when the magnitude of

punishment is below this value. Above this value, a delayed punishment would be preferred to an immediate punishment. This prediction has experimental support (Frederick et al., 2002; Kalenscher and Pennartz, 2008).

5. A reward of  $r$  delayed beyond  $t = r/a_{est}$  will lead to a negative subjective value. Hence, given an option between pursuing or forgoing this reward, the animal would only pursue (forgo) the reward at shorter (longer) delays.

When understanding the reversal of the “Magnitude Effect” for losses, it is important to keep in mind that as  $|r| \rightarrow \infty$ , both losses and gains approach the same asymptote.

$$D(r, t; |r| \rightarrow \infty) = \frac{1}{1 + \frac{t}{T_{ime}}}$$

Hence, as the magnitude of a loss increases, the size of the “Magnitude Effect” becomes lower and harder to detect (Figure 4).

In an environment with negative  $a_{est}$  (i.e., net punishing environment), all the predictions listed above would reverse trends. Specifically,

1. “Magnitude Effect” for gains: as  $r$  increases, the discounting becomes steeper
2. “Magnitude Effect” for losses: as the magnitude of a punishment increases, the discounting function becomes less steep.
3. “Sign Effect”: Punishments are discounted more steeply than gains of equal magnitudes.
4. Differential treatment of gains: as the magnitude of the gain decreases below  $a_{est}T_{ime}$  ( $r < -a_{est}T_{ime}$ ), it would be preferred at a delay. Beyond this magnitude, the gain would be preferred immediately.
5. A punishment of magnitude  $r$  will be treated with positive subjective value if it is delayed beyond  $t = r/a_{est}$ .

**Animals do not maximize long-term reward rates.** In typical animal intertemporal choice experiments, in order to ensure that different reward options do not lead to a marked difference in overall experiment duration, a post-reward delay is introduced for all options such that the net duration of each trial is constant. In such experiments, a global-reward-rate-maximizing agent should always choose the larger reward, irrespective of the cue-reward delay, since the net time spent per trial in collecting any reward equals the constant trial duration. However, a preponderance of experimental evidence shows that animals deviate from such ideal behavior of maximizing reward rates over the entire session (Stephens and Anderson, 2001; Kalenscher and Pennartz, 2008; Stephens, 2008). Such experimental results are typically interpreted to signify that animals do not, in fact, act as reward-rate-maximizing agents (Stephens and Anderson, 2001; Kalenscher and Pennartz, 2008; Stephens, 2008). TIMERR theory proposes that even though animals are maximizing reward rates, albeit under constraints of experience, post-reward delays are not incorporated into their decision process due to limitations

of associative learning (Kacelnik and Bateson, 1996). As a consequence, animal choice behavior in such laboratory tasks would appear not to maximize global reward rates.

TIMERR theory, however, allows for the possibility that in a variant of standard laboratory tasks that makes a post-reward delay immediately precede another reward included in the choice behavior would result in animals not ignoring post-reward delays. Prior experiments evince this possibility (Stephens and Anderson, 2001). Specifically, post-reward delays are included in the decision process by birds performing a patch leave-stay task that is economically equivalent to standard laboratory tasks on intertemporal choice (Stephens and Anderson, 2001). Also, as mentioned in the main text, TIMERR theory also allows for the inclusion of these delays in tasks where they can be learned e.g., when they are explicitly cued (Pearson et al., 2010; Blanchard et al., 2013).

**Effects of plasticity in the past integration interval ( $T_{ime}$ )**

The most important implication of the TIMERR theory is that the steepness of discounting of future rewards will depend directly on the past integration interval, i.e., the longer you integrate over the past, the more tolerant you will be to delays, and vice-versa. In the above sections, the past integration interval ( $T_{ime}$ ) was treated as a constant. However, the purpose of the past integration interval is to reliably estimate the baseline reward rate expected through the delay until a future reward. Further, since  $T_{ime}$  determines the temporal discounting steepness, it will also affect the rate at which animals obtain rewards in a given environment. Hence, depending on the reinforcement statistics of the environment, it would be appropriate for animals to adaptively integrate reward history over different temporal windows so as to maximize rates of reward.

In this section, we qualitatively address the problem of optimizing  $T_{ime}$ . We consider that an optimal  $T_{ime}$  would satisfy four criteria: (1) obtain rewards at magnitudes and intervals that maximize the fitness of an animal, which is accomplished partially through (2) reliable estimation of past reward rates leading to (3) appropriate estimations of opportunity cost for typical delays faced by the animal with (4) minimal computational/memory costs.

Before considering the general optimization problem for  $T_{ime}$ , it is useful to consider an illustrative example. This example ignores the last three criteria listed above and only considers the impact of  $T_{ime}$  on the fitness of an animal. Consider a hypothetical animal that typically obtains rewards at a rate of 1 unit per hour. Suppose such an animal is presented with a choice between (a) 2 units of reward available after an hour, and (b) 20 units of reward available after 15 h. The subjective values of options “a” and “b” are calculated below for four different values of  $T_{ime}$ , as per Equation (3).

	Subjective value of “a”	Subjective value of “b”	Chosen option
$T_{ime} = \infty$ h	1	5	<i>b</i>
$T_{ime} = 10$ h	0.91	2	<i>b</i>
$T_{ime} = 2.5$ h	0.71	0.71	Both equal
$T_{ime} = 1$ h	0.50	0.31	<i>a</i>



As is apparent, larger  $T_{ime}$  biases the choice toward option “b.” This is appropriate in order to maximize long-term reward rate since the long term reward rate is higher for option “b,” as shown below.

Reward rate having chosen option “b” = 20 units in 15 h = 20/15 units/h.

Reward rate having chosen option “a” = 2 units in 1 h + 14 units in the remaining 14 h = 16/15 units/h.

However, if we presume that this animal evolved so as to require a minimum reward of 2 units within every 10 h in order to function in good health, choosing option “b” would be inappropriate. Hence, it is clear that for this hypothetical animal,  $T_{ime}$  should be much lower than 10 h. In summary, so as to meet consumption requirements, it is inappropriate to integrate past reward rate history over very long times even if the animal has infinite computational/memory resources. Keeping in mind the above example and the four criteria listed for an optimal  $T_{ime}$ , we enumerate the following disadvantages for setting inappropriately large or inappropriately small  $T_{ime}$ .

Integrating over inappropriately large  $T_{ime}$  has at least four disadvantages to the animal: (1) a very long  $T_{ime}$  is inappropriate given consumption requirements of an animal, as illustrated above; (2) the computational/memory costs involved in this integration are high; (3) integrating over large time scales in a dynamically changing environment could make the estimate of past reward rate inappropriate for the delay to reward (e.g., integrating over the winter and spring seasons as an estimate of baseline reward rate expected over a delay of an hour in the summer might prove very costly for foragers); (4) the longer the  $T_{ime}$ , the harder it is to update  $a_{est}$  in a dynamic environment.

Integrating over inappropriately small  $T_{ime}$ , on the other hand, presents the following disadvantages: (1) estimate of baseline reward rate would be unreliable since integration must be carried out over a long enough time-scale so as to appreciate the stationary variability in an environment; (2) estimate of baseline reward rate might be highly inappropriate for the future delay (e.g., integrating over the past 1 min might be very inappropriate when the delay to a future reward is a day); (3) the animal would more greatly deviate from global optimality [as is clear from Equation (3)].

In light of the above discussion, we argue that the following relationships should hold for  $T_{ime}$ . In each of these relationships, all factors other than the one considered are assumed constant.

- R1. Time-dependent changes in environmental reinforcement statistics: if an environment is unstable, i.e., the reinforcement statistics of the environment are time-dependent, we predict that  $T_{ime}$  would be lower than the timescale of the dynamics of changes in environmental statistics.
- R2. Variability of estimated reward rate: if an environment is stable and has very low variability in the estimated reward rate it provides to an animal, integrating over a long  $T_{ime}$  would not provide a more accurate estimate of past reward rate than integrating over a short  $T_{ime}$ . Hence, in order to be better at adapting to potential changes in the environment

and minimize computational/ memory costs, we predict that in a stable environment,  $T_{ime}$  will reduce (increase) as the variability in the estimated reward rate reduces (increases).

- R3. Mean of estimated reward rate: in a stable environment with higher average reward rates, the benefit of integrating over a long  $T_{ime}$  will be smaller when weighed against the computational/memory cost involved. As an extreme example, when the reward rate is infinity, the benefit of integrating over long windows is infinitesimal. This is because the benefit of integrating over a longer  $T_{ime}$  can be thought of as the net gain in average reward rate over that achieved when decisions are made with the lowest possible  $T_{ime}$ . If the increase in average reward rate is solely due to an increase in the mean (constant standard deviation) of reward magnitudes, the proportional benefit of integrating over a large  $T_{ime}$  reduces. If the increase in average reward rate is solely due to an increase in frequency of rewards, the integration can be carried out over a lower time to maintain the estimation accuracy. Hence, we predict that, in general, as average reward rates increase (decrease),  $T_{ime}$  will decrease (increase).
- R4. Average delays to rewards: as the average delay between the moment of decision and receipt of rewards increases (decreases),  $T_{ime}$  should increase (decrease) correspondingly. This is because reward history calculated over a low  $T_{ime}$  might be inappropriate as an estimate of baseline reward rate for the delays until future reward.

In human experiments, it is common to give abstract questionnaires to study preference (e.g., “which do you prefer: \$100 now or \$150 a month from now?”). In such tasks, setting  $T_{ime}$  to be of the order of seconds or minutes might be very inappropriate to calculate a baseline expected reward rate over the month to a reward (R4 above). Hence, we predict that  $T_{ime}$  might increase so as to match the abstract delays to allow humans to discount less steeply as these delays increase. Similarly, when the choice involves delays of the order of seconds, integrating over hours might not be appropriate and therefore, the discounting steepness would be predicted to be higher in such experiments. Thus, in prior experimental results (Loewenstein and Prelec, 1992; Frederick et al., 2002; Schweighofer et al., 2006; Luhmann et al., 2008),  $T_{ime}$  might have changed to reflect the delays queried.

#### **Calculation of the estimate of past reward rate ( $a_{est}$ )**

It must be noted that even though the calculation of  $a_{est}$  is performed over a time-scale of  $T_{ime}$ , yet unspecified is the particular form of memory for past reward events. The simplest form of a memory function is one in which rewards that were received within a past duration of  $T_{ime}$  are recollected perfectly while any reward that was received beyond this duration is completely forgotten. A more realistic memory function will be such that a reward that was received will be remembered accurately with a probability depending on the time in the past at which it was received, with the dependence being a continuous and monotonically decreasing function. For such a function,  $T_{ime}$

will be defined as twice the average recollected duration over the probability distribution of recollection. The factor of two is to ensure that in the simplest memory model presented above, the longest duration at which rewards are recollected (twice the average duration) is  $T_{\text{ime}}$ .

If we define local updating as updating  $a_{\text{est}}$  based solely on the memory of the last reward (both magnitude and time elapsed since its receipt), the constraint of local updating when placed on

such a general memory function necessitates it to be exponential in time. In this case,  $a_{\text{est}}$  is updated as:

$$a_{\text{est}} \rightarrow a_{\text{est}} + \frac{2r}{T_{\text{ime}}}; \text{ upon receipt of reward}$$

$$a_{\text{est}} \rightarrow a_{\text{est}} \exp\left(-\frac{2t_{\text{lastreward}}}{T_{\text{ime}}}\right); \text{ otherwise}$$

where  $t_{\text{lastreward}}$  is the time elapsed since the receipt of the last reward.