Title: The value of initiating a pursuit in temporal decision-making

Authors

Elissa Sutlief*, Charlie Walters*, Tanya Marton†, Marshall G Hussain Shuler†

Affiliations

Elissa Sutlief
Department of Neuroscience, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA
esutlie1@jhmi.edu
*Co-first Author

Charlie Walters
Kavli Neuroscience Discovery Institute
Department of Neuroscience, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA
ckwalters@jhmi.edu
*Co-first Author

Tanya Marton
Department of Neuroscience, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA
tanya.marton@gmail.com
†Co-corresponding Author

Marshall G Hussain Shuler
Kavli Neuroscience Discovery Institute
Department of Neuroscience, Johns Hopkins University School of Medicine, 725 N. Wolfe Street, Baltimore, MD 21205, USA
shuler@jhmi.edu
†Corresponding Author

Keywords

Temporal decision-making, Reward-rate maximization, Subjective Value, time's cost, opportunity cost, apportionment cost, equivalent immediate reward, Discounting Function, Delay Effect, Magnitude Effect, Sign Effect, Misestimation, Normative theory, Malapportionment Hypothesis

Abstract

Reward-rate maximization is a prominent normative principle commonly held in behavioral ecology, neuroscience, economics, and artificial intelligence. Here, we identify and compare equations for evaluating the worth of initiating pursuits that an agent could implement to enable reward-rate maximization. We identify two fundamental temporal decision-making categories requiring the valuation of the initiation of a pursuit—forgo and choice decision-making—over which we generalize and analyze the optimal solution for how to evaluate a pursuit in order to maximize reward rate. From this reward-rate-maximizing formulation, we derive expressions for the subjective value of a pursuit, i.e. that pursuit's equivalent immediate reward magnitude, and reveal that time's cost is composed of an *apportionment*, in addition to, an *opportunity* cost. By re-expressing subjective value as a temporal discounting function, we show precisely how the temporal discounting function of a reward-rate-optimal agent is sensitive not just to the properties of a considered pursuit, but to the time spent and reward acquired outside of the pursuit for every instance spent within it. In doing so, we demonstrate how the apparent discounting function of a reward-rate-optimizing agent depends on the temporal structure of the environment and is a combination of hyperbolic and linear components, whose contributions relate the apportionment and opportunity cost of time, respectively. We further then show how purported signs of suboptimal behavior (hyperbolic discounting, the Delay effect, the Magnitude effect, the Sign effect) are in fact consistent with reward-rate maximization. Having clarified what features are and *are not* signs of optimal decision-making, we analyze the impact of the misestimation of reward rate-maximizing parameters in order to better account for the pattern of errors actually observed in humans and animals. We find that error in agents' assessment of the apportionment of time that underweights the time spent outside versus inside a considered pursuit type is the likely driver of suboptimal temporal decision-making observed behaviorally. We term this the Malapportionment Hypothesis. This generalized form for reward-rate maximization and its relation to subjective value and temporal discounting allows the true pattern of errors exhibited by humans and animals to be more deeply understood, identified, and quantified, which is key to deducing the learning algorithms and representational architectures actually used by humans and animals to evaluate the worth of pursuits.

Introduction

What is the worth of a pursuit? At the most universal level, temporal decision-making regards weighing the return of pursuits against their cost in time. The fields of economics, psychology, behavioral ecology, neuroscience, and artificial intelligence have endeavored to understand how animals, humans, and learning agents evaluate the worth of pursuits: how they factor the cost of time in temporal decision-making. A central step in doing so is to identify a normative principle and then to solve for how an agent, abiding by that principle, would best invest time in pursuits that compose a world. A normative principle with broad appeal identified in behavioral ecology is that of reward-rate maximization, as expressed in Optimal Foraging Theory (OFT), where animals seek to maximize reward rate while foraging in an environment (Charnov, 1976a, 1976b; Krebs et al., 1977; Pyke et al., 1977; Pyke, 1984). Solving for the optimal decision-making behavior under this objective provides the means to examine the curious pattern of adherence and deviation that is exhibited by humans and animals with respect to that ideal behavior. This difference provides clues into the process that animals and humans use to learn the value of, and represent, pursuits. Therefore, it is essential to analyze reward-rate-maximizing solutions for the worth of initiating a pursuit to clarify what behavioral signs are—and are not—deviations from optimal performance in the identification of the process (and its sources of error) actually used by humans and animals.

Equivalent immediate reward (subjective value, sv)

To ask, 'what is the value of a pursuit?' is to quantify by some metric the worth of a future state—the pursuit's outcome—at the time of a prior one, the pursuit's initiation. A sensible metric for the worth of a

pursuit is the magnitude of immediate reward that would be treated by an agent as equivalent to a policy of investing the requisite time in the pursuit and obtaining its reward. This *equivalent immediate reward,* as judged by the agent, is the pursuit's "Subjective Value" (sv), in the parlance of the field (Mischel et al., 1969). It is widely assumed that decisions about what pursuits should be taken are made on the basis of their subjective value (Niv, 2009). However, a decision-making algorithm needn't calculate subjective value in its evaluation of the worth of initiating a pursuit. It could, for instance, assess the reward rate of the pursuit relative to the reward rate received in the world as a whole. Indeed, algorithms leading to reward-rate optimization can arise from different underlying processes, each with their own controlling variables. Nonetheless, any algorithm's evaluation can be re-expressed in terms of equivalent immediate reward, providing a ready means to compare evaluation across different learning algorithms and representational architectures as biologically realized in animals and humans or as artificially implemented in silico.

Decisions to initiate pursuits

As decisions occur at branch points between pursuits, the value of initiating a pursuit is of particular importance as it is on this basis that an agent would decide 1) whether to accept or *forgo* an offered pursuit; or, 2) how to *choose* between mutually exclusive pursuits. Though 'Forgo' decisions are regarded as near-optimal, as in prey selection (Krebs et al., 1977; Stephens and Krebs, 1986; Blanchard and Hayden, 2014), 'Choice' decisions—as commonly tested in laboratory settings—reveal a suboptimal bias for smaller-sooner rewards when selection of later-larger rewards would maximize the reward rate while foraging in an environment, i.e. the "global reward rate" (Logue et al., 1985; Blanchard and Hayden, 2015; Carter and Redish, 2016; Kane et al., 2019). This curious pattern of behavior, wherein forgo decisions can present as optimal while choice decisions as suboptimal, poses a challenge to any theory purporting to rationalize temporal decision-making as observed in animals and humans.

Temporal Discounting Functions

Historically, temporal decision-making has been examined using a temporal discounting function to describe how delays in rewards influence their valuation. Temporal discounting functions describe the subjective value of an offered reward as a function of when the offered reward is realized. To isolate the form of discount rate from any difference in reward magnitude and sign, subjective value is commonly normalized by the reward magnitude when comparing subjective value-time functions (Strotz, 1956; Jimura et al., 2009). Therefore, we use the convention that temporal discounting functions are the magnitude-normalized subjective value-time function (Strotz, 1956). An understanding of the form of temporal discounting has important implications in life, as steeper temporal discounting has been associated with many negative life outcomes (Bretteville-Jensen, 1999; Critchfield and Kollins, 2001; Bickel et al., 2007, 2012; Story et al., 2014), most notably the risk of developing an addiction. Psychologists and behavioral scientists have long found that animals' temporal discounting in intertemporal choice tasks is well-fit by a hyperbolic discounting function (Ainslie, 1974; Mazur, 1987; Richards et al., 1997; Monterosso and Ainslie, 1999; Green and Myerson, 2004; Hwang et al., 2009; Louie and Glimcher, 2010). Other examples of motivated behavior also show hyperbolic temporal discounting (Haith et al., 2012).

Often, this perspective assumes that the delay in and of itself devalues a pursuit's reward, failing to carefully distinguish the impact of its delay from the impact of the time required and reward obtained *outside* the considered pursuit. As a result, the discounting function tends to be treated as a process unto itself rather than the consequence of a process. Consequently, the field has concerned itself with the form of the discounting function—exponential (Glimcher et al., 2007; McClure et al., 2007), hyperbolic (Rachlin et al., 1972; Ainslie, 1975; Thaler, 1981; Mazur, 1987; Benzion et al., 1989; Green et al., 1994; Frederick et al., 2002; Kobayashi and Schultz, 2008; Calvert et al., 2010), pseudo-hyperbolic (Laibson, 1997; Montague et al., 2006; Berns et al., 2007), etc., as either derived from some normative principle, or as fit to behavioral observation. An exponential discounting function, for instance, was derived by Samuelson from the normative principle of time consistency (Samuelson 1937) and is widely held as rational (Samuelson, 1937; Koopmans, 1960; Laibson, 1997; Montague and Berns, 2002; McClure et al., 2004;

Mazur, 2006; Schweighofer et al., 2006; Berns et al., 2007; McClure et al., 2007; Nakahara and Kaveri, 2010; Kane et al., 2019), and by implication, reward-rate maximizing. Observed temporal decision-making behavior, however, routinely exhibits time inconsistencies (Strotz, 1956; Ainslie, 1975; Laibson, 1997; Frederick et al., 2002) and is better fit by a hyperbolic discounting function (Ainslie, 1975; Mazur et al., 1985; Frederick et al., 2002; Green and Myerson, 2004), and on that contrasting basis, humans and animals have commonly been regarded as *irrational* (Thaler, 1981; Loewenstein and Thaler, 1989; Loewenstein and Prelec, 1992; Frederick et al., 2002; Baker et al., 2003; Estle et al., 2006; Kalenscher and Pennartz, 2008). In addition, the case that humans and animals are irrational is, ostensibly, furthered by the observation of the 'Magnitude Effect' (Benzion et al., 1989; Green et al., 1994, 1997; Frederick et al., 2002; Baker et al., 2003; Estle et al., 2006; Yi et al., 2006; Grace et al., 2012; Kinloch and White, 2013) and the 'Sign Effect' (Thaler, 1981; Loewenstein and Thaler, 1989; Loewenstein and Prelec, 1992; Frederick et al., 2002; Baker et al., 2003; Estle et al., 2006; Kalenscher and Pennartz, 2008) (where the apparent discounting function is affected by the magnitude and the sign of the offered pursuit's outcome, respectively), and the 'Delay Effect', where preference between pursuits can switch as the time required for their obtainment changes (Rachlin et al., 1972; Ainslie, 1974; Bateson and Kacelnik, 1996; Stephens and Anderson, 2001; Frederick et al., 2002; Hayden and Platt, 2007; McClure et al., 2007; Carter et al., 2015; Carter and Redish, 2016).

Here, we aim to identify equations for evaluating the worth of initiating pursuits that an agent could implement to enable reward-rate maximization. We wish to gain deeper insight into how a considered pursuit, with its defining features (its reward and time), relates to the world of pursuits in which it is embedded in determining the pursuit's worth. Specifically, we investigate how pursuits and the pursuit-to-pursuit structure of a world interact with policies of investing time in particular pursuits to determine the global reward rate reaped from an environment. We aim to provide greater clarity into what constitutes time's cost and how it can be understood with respect to the reward and temporal structure of an environment and to counterfactual time investment policies. We propose that, by determining optimal decision-making equations and converting them to their equivalent subjective value and temporal discounting functions, actual (rather than asserted) deviations from optimality exhibited by humans and animals can be truly determined. We speculate that purported anomalies deviating from ostensibly 'rational' decision-making may in fact be consistent with reward-rate optimization. Further, by identifying parameters enabling reward-rate maximization and assessing resulting errors in valuation caused by their misestimation, we aim to gain insight into which parameters humans and animals may (mis)-represent that most parsimoniously explains the pattern of temporal decision-making actually observed.

Results

To gain insight into the manner by which animals and humans attribute value to pursuits, it is essential to first understand how a reward-rate-maximizing agent would evaluate the worth of any pursuit within a temporal decision-making world. Here, by considering **Forgo** and **Choice** temporal decisions, we re-conceptualize how an ideal reward-rate-maximizing agent ought to evaluate the worth of initiating pursuits. We begin by formalizing temporal decision-making worlds as constituted of pursuits, with pursuits described as having reward rates and weights (their relative occupancy). Then, we examine **Forgo** decisions to examine what composes the cost of time and how a policy of taking/forgoing pursuits factors into the global reward rate of an environment and thus the worth of a pursuit. Having done so, we derive two equivalent expressions for the worth of a pursuit and from them re-express the worth of a pursuit as its equivalent immediate reward (its 'subjective value') in terms of the global reward rate achieved under policies of 1) accepting or 2) forgoing the considered pursuit type. We next examine **Choice** worlds to investigate the apparent nature of a reward-rate optimizing agent's temporal discounting function. Finally, having identified reward-rate maximizing equations, we examine what parameter misestimation leads to suboptimal pursuit evaluation that best explains behavior observed in humans and animals. Together, by considering the temporal structure of a time investment world as one composed of pursuits described by

their rates and weights, we seek to identify equations for how a reward-rate-maximizing agent could evaluate the worth of any pursuit composing a world and how those evaluations would be affected by misestimation of enabling parameters.

Temporal decision worlds are composed of pursuits with reward rates and weights

A temporal decision-making world is one composed of pursuits. A *pursuit* is a defined path over which an agent can traverse by investing time that often (but not necessarily) results in reward but which always leads to a state from which one or more potential other pursuits are discoverable. Pursuits have a *reward magnitude* (r) and a *time* (t). A pursuit therefore has 1) a *reward rate* (ρ, rho) and 2) a *weight* (w), being its relative occupancy with respect to all other pursuits. Here we refer to the defining features of a pursuit (its reward, time, reward rate, or weight) by subscripting the pursuit's name or type to each feature's symbol (r, t, rho or w, respectively), e.g. (ρ_{Pursuit}, w_{Pursuit}). In this way, the pursuit structure of temporal decision-making worlds, and the qualities defining pursuits, can be adequately referenced.

The temporal decision-making worlds considered are recurrent in that an agent traversing a world under a given policy will eventually return back to its current location. As pursuits constitute an environment, the environment itself then has a reward rate, the 'global reward rate' ρ_g, achieved under a given decision policy, $\rho_g Policy$. Whereas the global reward rate realized under a given policy of choosing one or another pursuit path may or may not be reward-rate optimal, the global reward rate achieved under a reward-rate maximal policy will be denoted as ρ_g*.

Forgo and Choice decision topologies

Having established a nomenclature for the properties of a temporal decision-making world, we now identify two fundamental types of decisions regarding whether to initiate a pursuit: "Forgo" decisions, and "Choice" decisions. In a Forgo decision **(Figure 1, left)**, the agent is presented with one of possibly many pursuits (and see **Supplemental 1 & 2**) that can either be accepted or rejected. After either the conclusion of the pursuit, if accepted, or immediately after rejection, the agent returns to a pursuit by default (the "default" pursuit). This default pursuit effectively can be a waiting period over which reward could be received and reoccurs until the next pursuit opportunity becomes available. Rejecting the offered pursuit constitutes a policy of spending less time to make a traversal across that decision-making world, whereas accepting the offered pursuit constitutes a policy of spending more time to make a traversal. In a Choice decision **(Figure 1, right)**, the agent is presented with a choice between at least two simultaneous and mutually exclusive pursuits, typically differing in their respective rewards' magnitudes and delays. Under any decision, upon exit from a pursuit, the agent returns to the same environment that it would have entered were the pursuit rejected. In the Forgo case in **Figure 1**, a policy of spending less time to traverse the world by rejecting the purple pursuit to return to the gold pursuit—and thus obtaining a smaller amount of reward (left)—must be weighed against a policy of acquiring more reward by accepting the purple pursuit at the expense of spending more time to traverse the world (right). In the Choice case in **Figure 1**, a policy of spending less time to traverse the world (left) by taking the smaller-sooner pursuit (aqua) must be weighed against a policy of spending more time to traverse the world (right) by accepting the larger-later pursuit (purple).

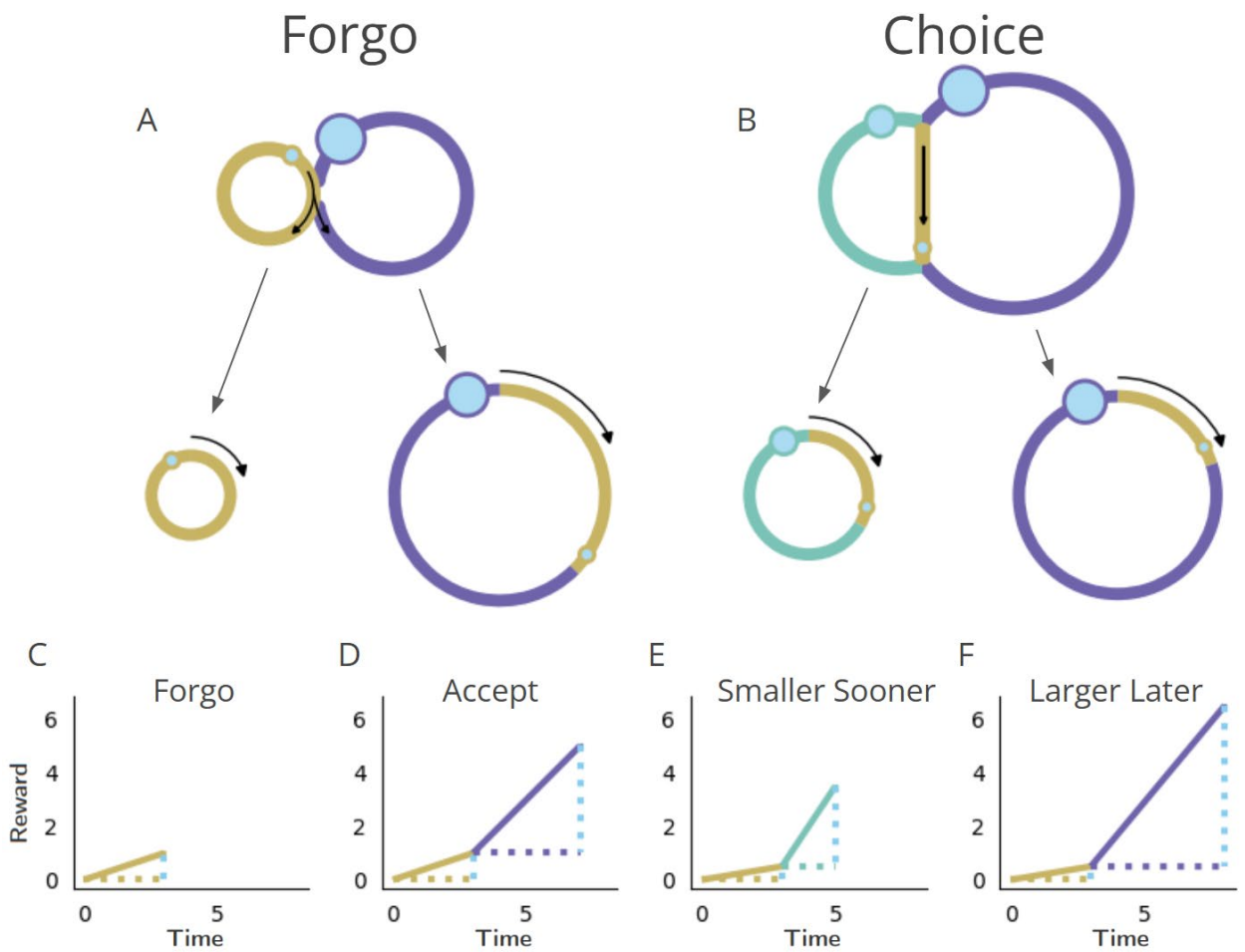


Figure 1. Fundamental classes of temporal decision-making regarding initiating a pursuit: "Forgo" and "Choice". 1st row- Topologies. The temporal structure of worlds exemplifying Forgo (left) and Choice (right) decisions mapped as their topologies. Forgo: A forgo decision to accept or reject the purple pursuit. When exiting the gold pursuit having obtained its reward (small blue circle), an agent is faced with 1) a path to re-enter gold, or 2) a path to enter the purple pursuit, which, on its completion, re-enters gold. Choice: A choice decision between an aqua pursuit, offering a small reward after a short amount of time, or a purple pursuit offering a larger amount of reward after a longer time. When exiting the gold pursuit, an agent is faced with a path to enter 1) the aqua or 2) the purple pursuit, both of which lead back to the gold pursuit upon their completion. **2nd row- Policies**. Decision-making policies chart a course through the pursuit-to-pursuit structure of a world. Policies differ in the reward obtained and in the time required to complete a traversal of that world under that policy. Policies of investing less (left) or more (right) time to traverse the world are illustrated for the considered Forgo and Choice worlds. Forgo: A policy of rejecting the purple pursuit to re-enter the gold pursuit (left) acquires less reward though it requires less time to make a traversal of the world than a policy of accepting the purple option (right). Choice: A policy of choosing the aqua pursuit (left) results in less reward though it requires less time to traverse the world than a policy of choosing the purple pursuit (right). **3rd row- Time/reward investment.** The times (stippled horizontal lines, colored by pursuit) and rewards (stippled vertical blue lines) of pursuits, and their associated reward rates (solid lines) acquired under a policy of forgo or accept in the Forgo world, or, of choosing the sooner smaller or later larger pursuit in the Choice world.

Behavioral observations under Forgo and Choice decisions

These classes of temporal decisions have been investigated by ecologists, behavioral scientists, and psychologists for decades. Forgo decisions describe instances likened to prey selection (Krebs et al., 1977; Stephens and Krebs, 1986; Blanchard and Hayden, 2014). Choice decisions have extensively been examined in intertemporal choice experiments (Rachlin et al., 1972; Ainslie, 1974; Bateson and Kacelnik, 1996; Stephens and Anderson, 2001; Frederick et al., 2002; Hayden and Platt, 2007; McClure et al., 2007; Carter et al., 2015; Carter and Redish, 2016). Experimental observation in temporal decision-making demonstrates that animals are optimal (or virtually so) in Forgo (Krebs et al., 1977; Stephens and Krebs,

1986; Blanchard and Hayden, 2014), taking the offered pursuit when its rate exceeds the "background" reward rate, and are as if sub-optimally impatient in Choice, selecting the smaller-sooner (SS) pursuit when the larger-later (LL) pursuit is reward-rate-maximizing (Logue et al., 1985; Blanchard and Hayden, 2015; Carter and Redish, 2016; Kane et al., 2019).

Deriving optimal policy from forgo decision-making worlds

We begin our examination of how to maximize the global reward rate reaped from a landscape of rewarding pursuits by examining forgo decisions. A general formula for the global reward rate of an environment in which agents must invest time in obtaining rewards is needed in order to formally calculate a policy's ability to accumulate reward. Optimal policies maximize reward accumulation over the time spent foraging in that environment. In a forgo decision, an agent is faced with the decision to take, or to *forgo*, pursuit opportunities. We sought to determine the reward rate an agent would achieve were it to pursue rewards with magnitudes $r_1, r_2, \ldots . r_n$ each requiring an investment of time $t_1, t_2, \ldots . t_n$. At any particular time, the agent is either 1) investing time in a pursuit of a specific reward and time, or 2) available to encounter and take new pursuits from a pursuit to which it defaults. With the assumption that reward opportunities become randomly encountered by the agent at a frequency of $f_1, f_2, \ldots . f_n$ from the default pursuit, it becomes possible to calculate the total global reward rate of the environment, ρ_g, as in **Equation 1** (Ap 1)…

$$\boldsymbol{\rho_g = \frac{\sum_{i=1}^{n} f_i r_i + \rho_d}{\sum_{i=1}^{n} f_i t_i + 1}}$$

Equation 1. The global reward rate under multiple pursuits (Ap 1)

…where ρ_d is the rate of reward attained in the default pursuit per unit time in the default pursuit. Should rewards not occur while in the default pursuit, ρ_d, will be zero. **Equation 1** allows for the calculation of the global reward rate achieved by any policy accepting a particular set of pursuits from the environment. This derivation of global reward rate is akin to those similarly derived for prey selection models (see (Charnov and Orians, 1973) and (Stephens and Krebs, 1986)).

Parceling the world into the considered pursuit type ("in" pursuit) and everything else ("out" of pursuit)

In order to simplify representations of policies governing any given pursuit opportunity, we reformulate the above expression for global reward rate, ρ_g, from the perspective of a policy of accepting any given pursuit (**Equation 2**). The environment may be parcellated into the time spent and rewards achieved *inside* the considered pursuit on average, for every instance that time is spent and rewards achieved *outside* the considered pursuit, on average. We can pull out the inside reward (r_{in}) and inside time (t_{in}) from the equation above to isolate the inside and outside components of the equation.

$$\boldsymbol{\rho_g = \frac{r_{in} + \frac{\sum_{i \neq in}^{n} f_i r_i + \rho_d}{f_{in}}}{t_{in} + \frac{\sum_{i \neq in}^{n} f_i t + 1}{f_{in}}}}$$

Equation 2. The global reward rate expressed as "in" and "outside" considered pursuit type (Ap 2)

From there, we define t_{out} as the average time spent outside the considered pursuit for each instance that the considered pursuit is experienced (**Equation 3**).

$$t_{out} = \frac{\sum_{i \neq in} f_i t_i + 1}{f_{in}}$$

Equation 3. The average time spent outside of the considered pursuit type (Ap 2)

Similarly, the outside reward, r_{out}, encompasses the average amount of reward collected from all sources outside the considered pursuit (**Equation 4**).

$$r_{out} = \frac{\sum_{i \neq in} f_i r_i + \rho_d}{f_{in}}$$

Equation 4. The average reward collected outside of the considered pursuit type (Ap 3)

Parceling a pursuit world into a considered pursuit (all instances "inside" the considered pursuit type) and everything else (i.e., everything "outside" the considered pursuit type), then gives the generalized form for the reward rate of an environment under a given policy (**Equation 5**) as…

$$\rho_g = \frac{r_{in} + r_{out}}{t_{in} + t_{out}}$$

Equation 5. Global reward rate with respect to the considered pursuit type (Ap 3)

…which depends on the average reward earned and the average time spent between opportunities to make the decision, in addition to the average reward returned and average time spent in the considered pursuit **(Ap 3 & Figure 2).**

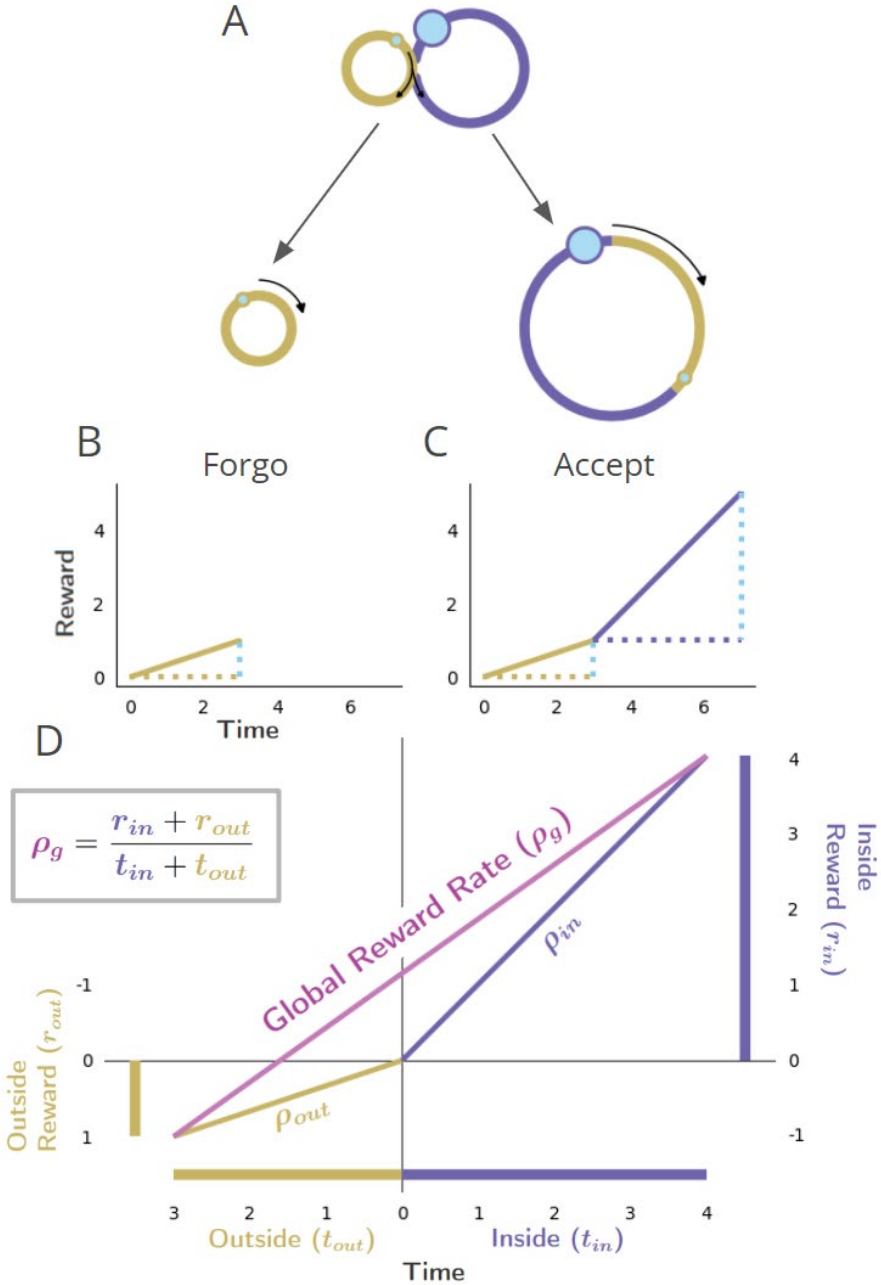


Figure 2. Global reward rate with respect to parceling the world into "in" and "outside" the considered pursuit. **A-C)** as in Figure 1 "Forgo". Conventions as in Figure 1. **D)** The world divided into "Inside" and "Outside" the purple pursuit type, as the agent decides whether to forgo or accept. The axes are centered on the position of the agent, just before the purple pursuit, where the upper right quadrant shows the inside (purple) pursuit's average reward, time, and reward rate (ρ_{in}), while the bottom left quadrant shows the outside pursuit (gold) average reward, time, and reward rate (ρ_{out}). The global reward rate (ρ_g) is shown in magenta, calculated from the boxed equation. The agent may determine the higher reward rate yielding policy by comparing the outside reward rate (ρ_{out}) with the resulting global reward rate (ρ_g) under a policy of accepting the considered pursuit.

Figure 2D depicts the global reward rate achieved with respect to the time and reward obtained from a considered pursuit ("Inside") and the time and reward obtained outside that considered pursuit type, i.e., that pursuit's ("Outside"). By so parsing the world into "in" and "outside" the considered pursuit, it can also be appreciated from **Figure 2D** that the fraction of time in the environment invested in the considered option, in, can be expressed as $w_{in} = \frac{t_{in}}{t_{in} + t_{out}}$, and the fraction of time spent outside the

considered pursuit as $1 - w_{in} = \frac{t_{out}}{t_{in} + t_{out}}$. A world can thus be understood in terms of its composing pursuits' reward rates and weights (their relative occupancy), with the global reward rate being a weighted average of the reward rate from the considered pursuit type, $\rho_{in} = \frac{r_{in}}{t_{in}}$, and the reward rate outside the considered pursuit type, $\rho_{out} = \frac{r_{out}}{t_{out}}$.

$$\boldsymbol{\rho_g = w_{in} \cdot \rho_{in} + (1 - w_{in}) \cdot \rho_{out}}$$

Equation 6. Global reward rate is the sum of weighted inside and outside rates (Ap 4)

Therefore, the global reward rate is the sum of the local reward rates of the world's constituent pursuits under a given policy when weighted by their relative occupancy: the weighted average of the local reward rates of the pursuits constituting the world (**Equation 6**).

Reward-rate optimizing forgo policy: compare a pursuit's local reward rate to its outside reward rate

We can now compare two competing policies to identify the policy that maximizes reward rate, such that it is the maximum possible reward rate, $\rho_g{}^*$. A policy of taking or forgoing a given pursuit type may improve the reward rate reaped from the environment as a whole (**Figure 3).** Using **Equation 5**, the policy achieving the greatest global reward rate can be realized through an iterative process where pursuits with lower reward rates than the reward rate obtained from everything other than the considered pursuit type are sequentially removed from the policy. While **Figure 3** illustrates a simple topology for clarity, this process applies to topologies comprising any n-pursuits **(Supplemental 1),** which may occur at varying frequency **(Supplemental 2)**. The optimal policy for forgoing can therefore be calculated directly from the considered pursuit's reward rate, ρ_{in}, and the reward rate outside of that pursuit type, ρ_{out}. Global reward rate can be maximized by iteratively forgoing the considered pursuit if its reward rate is less than its outside reward rate, $\rho_{in} < \rho_{out}$, treating forgoing and taking a considered pursuit as equivalent when $\rho_{in} = \rho_{out}$, and taking the considered pursuit when $\rho_{in} > \rho_{out}$ (**Ap 5**).

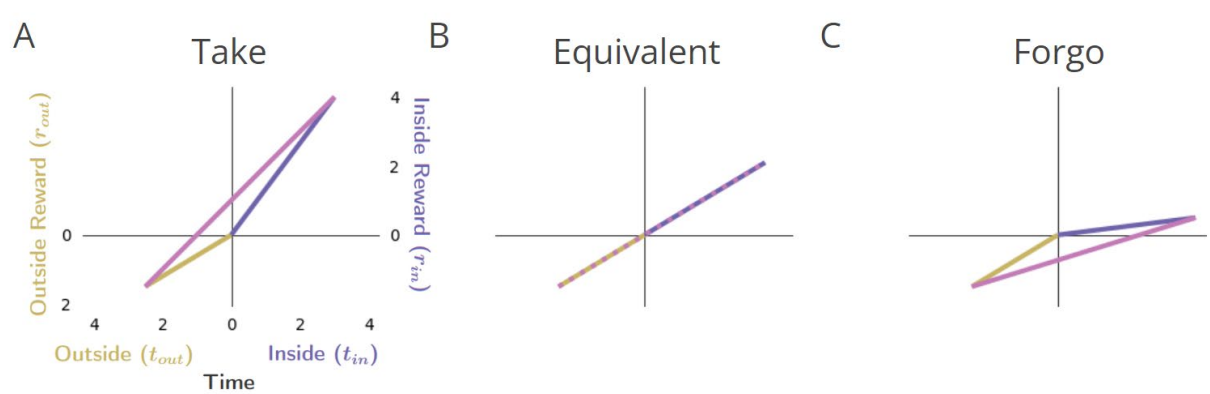


Figure 3. Forgo Decision-making. **A)** When the reward rate of the considered pursuit (slope of the purple line) exceeds that of its outside rate (slope of gold), the global reward rate (slope of magenta) will be greater than the outside, and therefore the agent should accept the considered pursuit. **B)** When the reward rates inside and outside the considered pursuit are equivalent, the global reward rate will be the same when accepting or forgoing: the policies are equivalent. **C)** When the reward rate of the considered pursuit is less than its outside rate, the resulting global reward rate if accepting the considered pursuit will be less than its outside reward rate and therefore should be forgone.

Following this policy would be equivalent to comparing the local reward rate of a pursuit to the global reward rate obtained under the reward-rate-optimal policy: forgo the pursuit when its local reward rate is less than the global reward under the reward-rate-optimal policy, $\rho_{in} < \rho_g{}^*$, take or forgo the pursuit when the reward rate of the pursuit is equal to the global reward rate under the optimal policy $\rho_{in} = \rho_g{}^*$, and take the pursuit when its local reward rate is more than the global reward rate under the reward-rate-

optimal policy, $\rho_{in} > \rho_g{}^*$ (**Ap 5**). The maximum reward rate reaped from the environment can thus be eventually obtained by comparing the local reward rate of a considered pursuit to its outside reward rate (i.e., the global reward rate of a policy of *not* accepting the considered pursuit type).

Equivalent immediate reward: the 'subjective value', *sv*, of a pursuit

Having recognized how a world can be decomposed into pursuits described by their rates and weights, and identifying optimal policies under forgo decisions, we may now ask anew, "What is the worth of a pursuit?" **Figure 2D** illustrates that the global reward rate obtained under a policy of taking a pursuit is not just a function of the time and return of the pursuit itself, but also the time spent and return gained outside of that pursuit type. Therefore, the worth of a pursuit relates to how much the pursuit would add (or detract) from the global reward rate realized in its acquisition.

Subjective Value of the considered pursuit with respect to the global reward rate.

This relationship between a considered pursuit type, its outside, and the global reward rate can be re-expressed in terms of an immediate reward magnitude requiring no time investment that yields the same global reward rate as that arising from a policy of taking the pursuit **(Figure 4)**. Thus, for any pursuit in a world, the amount of immediate reward that would be accepted in place of its initiation and attainment could serve as a metric of the pursuit's worth at the time of its initiation. Given the optimal policy above, an expression for this immediate reward magnitude can be derived (**Equation 7 & Ap 6**). This ***global reward-rate equivalent immediate reward*** **(see Figure 4)** is the *subjective value* of a pursuit, sv_{Pursuit} (or simply, *sv*, when the referenced pursuit can be inferred), as similarly expressed in prior foundational work (McNamara, 1982) and subsequent extensions (see (Fawcett et al., 2012)).

$$\boldsymbol{sv = r_{in} - \rho_g t_{in}}$$

Equation 7. The Subjective Value of a pursuit expressed in terms of the global reward rate achieved under a policy of accepting that pursuit (Ap 6)

The subjective value of a pursuit under the reward-rate-optimal policy will be denoted as sv^*_{Pursuit}.

$$sv^* = r_{in} - \rho_g^* t_{in}$$

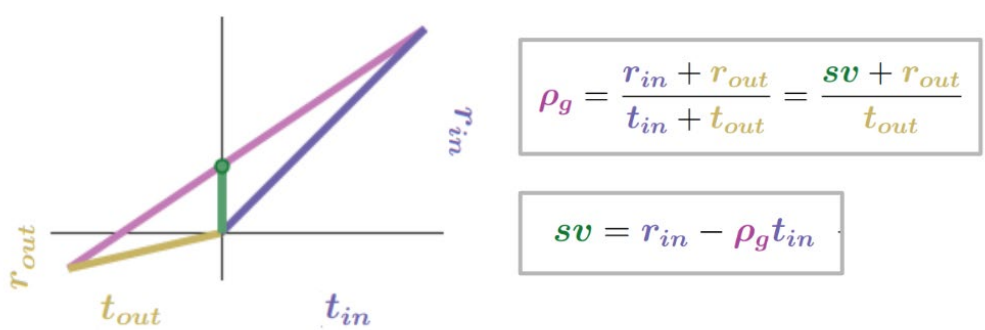


Figure 4. The Subjective Value (sv) of a pursuit is the global reward-rate-equivalent immediate reward magnitude. The subjective value (green bar) of a pursuit is that amount of reward requiring no investment of time that the agent would take as equivalent to a policy of accepting and acquiring the considered pursuit. For this amount to be equivalent, the immediate reward magnitude must result in the same global reward rate as that of accepting the pursuit. The global reward rate obtained under a policy of accepting the considered pursuit type is the slope of the line connecting the average times and rewards obtained in and outside the considered pursuit type. Therefore, the global reward rate-equivalent immediate reward (i.e., the subjective value of the pursuit) can be depicted graphically as the y-axis intercept (green dot) of the line representing the global reward rate achieved under a policy of accepting the considered pursuit. $r_{in} = 4$, $t_{in} = 4$, $r_{out} = 0.7$, $t_{out} = 3$

The calculation of the subjective value of a pursuit, sv, quantifies precisely the worth of a pursuit in terms of an immediate reward that would result in the same global reward rate as that pursuant to its attainment. Thus, choosing either an immediate reward of magnitude sv or choosing to pursue the considered pursuit, investing the required time and acquiring its reward, would produce an equivalent global reward rate. An agent pursuing an optimal policy would find immediate rewards of magnitude less than sv less preferred than the considered pursuit and immediate rewards of magnitude greater than sv more preferred than the pursuit.

The forgo decision can also be made from subjective value.

With this understanding, in the case that the considered pursuit's reward rate is greater than the optimal global reward rate, it will be greater than its outside reward rate, and therefore the subjective value under an optimal policy will be greater than zero **(Figure 3A)**.

$\rho_{in} > \rho_g{}^* > \rho_{out} \rightarrow sv^* > 0, sv > 0$, choose considered pursuit **(Ap 7)**

Should the considered pursuit's reward rate be equal to its outside reward rate, it will be equal to the optimal global reward rate, and the subjective value of the considered pursuit will be zero **(Figure 3B)**.

$\rho_{in} = \rho_g{}^* = \rho_{out} \rightarrow sv^* = 0, sv = 0$, forgoing and choosing are equivalent **(Ap 7)**

Finally, if the considered pursuit's reward rate is less than the outside reward rate, it must also be less than the global optimal reward rate; therefore, the subjective value of the considered pursuit under the optimal policy will be less than zero **(Figure 3C)**. A negative subjective value thus indicates that a policy of taking the considered pursuit would result in a global reward rate that is less than a policy of forgoing the considered pursuit. Equivalently, a negative subjective value can be considered the amount an agent ought be willing to pay to avoid having to take the considered pursuit.

$\rho_{in} < \rho_g{}^* = \rho_{out} \rightarrow sv^* < 0, sv < 0$, forgo considered pursuit **(Ap 7)**

While brains of humans and animals may not in fact calculate subjective value, converting to the equivalent immediate reward, sv, 1) makes connection to temporal decision-making experiments where such equivalences between delayed and immediate rewards are assessed, 2) serves as a common scale of comparison irrespective of the underlying decision-making process, and 3) deepens an understanding of how the worth of a pursuit is affected by the temporal structure of the environment's reward-time landscape.

Subjective value with respect to the pursuit's outside: insights into the cost of time

To the latter point, **Equation 7** has a (deceptively) simple appeal: the worth of a pursuit ought be its reward magnitude less its cost of time **(Figure 5 Left Column)**. But what is the cost of time? The **cost of time** of a considered pursuit is the global reward rate of the world under a policy of accepting the pursuit, times the time that the pursuit would take, $\rho_g t_{in}$ **(Figure 5B).** Therefore, the equivalent immediate reward of a pursuit, its ***subjective value***, corresponds to the subtraction of the cost of time from the pursuit's reward. The subjective value of a pursuit is how much *extra* reward is earned from the pursuit than would on average be earned by investing that amount of time, in that world, under a policy of accepting the considered pursuit.

While appealing in its simplicity, the terms on the right-hand side of **Equation 7**, r_{in} and $\rho_g t_{in}$, lack independence from one another—the reward of the considered pursuit type contributes to the global reward rate, ρ_g. Subjective value can alternatively and more deeply be understood by re-expressing subjective value in terms that are independent of one another. Rather than expressing the worth of a pursuit in terms of the global reward rate obtained when accepting it, as in **Equation 7** (**Figure 5 left column)**, the worth of a pursuit can be expressed in terms of the reward obtained and time spent outside the considered pursuit type (**Figure 5 right column)**, as in **Equation 8** (and see **Ap 8**).

$$\boldsymbol{sv_{in} = (r_{in} - \frac{r_{out}}{t_{out}} t_{in})(\frac{t_{out}}{t_{in} + t_{out}})}$$

Equation 8. Subjective value of a pursuit from the perspective of the considered pursuit and its outside (Ap 8)

These expressions are equivalent to one another **(Ap 8 & Figure 5, compare left and right sides).**

$$\boldsymbol{sv = r_{in} - \rho_g t_{in} = (r_{in} - \frac{r_{out}}{t_{out}} t_{in})(\frac{t_{out}}{t_{in}+t_{out}})}$$

For an interactive exploration of these effects of changing the outside and inside reward and time on subjective value, see Supplemental GUI.

Time's cost: opportunity & apportionment costs determine a pursuit's subjective value

By decomposing the global reward rate into 'inside' and 'outside' the considered pursuit type, the cost of time is revealed as being determined by 1) an opportunity cost, *and* 2) an apportionment cost **(Figure 5)**. We show in **Figure 5** that the left and right parenthetical terms of **Equation 8** correspond to subtracting the *opportunity* cost of time (r_{in}- $r_{out}/t_{out} * t_{in}$) and then scaling by the *apportionment* of time spent 'outside' the considered pursuit, ($t_{out}/(t_{out} + t_{in})$, i.e. the weight of the outside. The ***opportunity cost*** associated with a considered pursuit, $\rho_{out} t_{in}$, is the reward rate outside the considered pursuit type, ρ_{out}, times the time of the considered pursuit, t_{in} **(Figure 5C**). This opportunity cost is subtracted from the reward obtained from accepting the considered pursuit to yield the opportunity cost-subtracted reward (**Figure 5D**). **Figure 5E** shows that subjective value is to the opportunity cost-subtracted reward as the outside time is to the total time to traverse the world. Therefore, the right parenthetical term, $t_{out}/(t_{out} + t_{in})$, uses time's apportionment in and outside the pursuit to scale the opportunity cost-subtracted reward to the equivalent reward magnitude requiring no time investment, ie, the subjective value. This "apportionment scaling" term thus relates the agent's time allocation in the world: the time spent outside a pursuit type relative to the time spent to traverse the world. In so downscaling, the subjective value of a considered pursuit (green) is to the time it would take to traverse the world were the pursuit *not taken*, t_{out}, as its opportunity cost subtracted reward (cyan) is to the time to traverse the world *were it to be taken* (t_{in}+ t_{out}) **(Figure 5E)**.

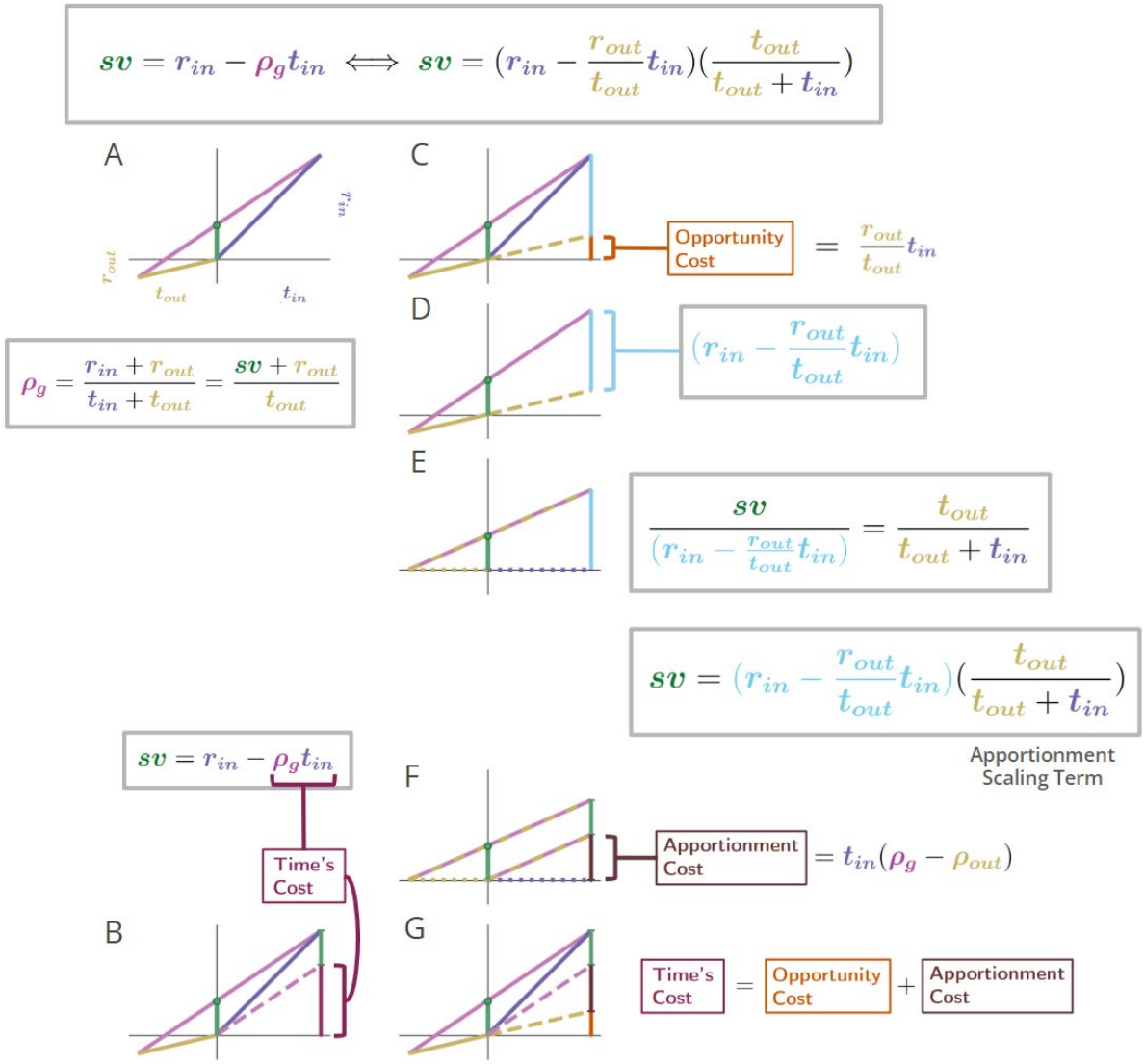


Figure 5. Equivalent expressions for subjective value reveal time's cost comprises an apportionment as well as opportunity cost.

A. The subjective value of a pursuit can be expressed in terms of the global reward rate obtained under a policy of accepting the pursuit. $r_{in} = 4$, $t_{in} = 4$, $r_{out} = .7$, $t_{out} = 3$. **B.** The cost of time of a pursuit is the amount of reward earned on average in an environment over the time needed for its obtainment under a policy of accepting the pursuit. The reward rate earned on average is the global reward rate (slope of magenta line). Projecting that global reward over the time of the considered pursuit (dashed magenta line) provides the cost of time for the pursuit (vertical maroon bar). Therefore, the subjective value of a pursuit is equivalent to its reward magnitude less the cost of time of the pursuit (box equation above B). **C.** Expressing subjective value with respect to the outside reward and time rather than the global reward rate reveals that a portion of a pursuit's time cost arises from an opportunity cost (orange bar). The opportunity cost of a pursuit is the amount of reward earned over the considered pursuit's time on average under a policy of *not* taking the considered pursuit (i.e., the outside reward rate, slope of gold line). Projecting the slope of the gold line over the time of the considered pursuit (dashed gold line) provides the opportunity cost of the pursuit (vertical orange bar). **D.** The opportunity cost-subtracted reward (cyan bar). The triangle with sides cyan, magenta, and gold (solid and dashed) is congruent with the triangle with sides green, magenta, and gold (solid). Therefore, **E.** *sv* is to the opportunity cost-subtracted reward, as t_{out} is to $t_{out} + t_{in}$. The opportunity cost-subtracted reward can thus be scaled by $t_{out}/(t_{out} + t_{out})$ to yield the subjective value of the pursuit. This scaling term, which we coin 'apportionment scaling', is the proportion that the outside time is to the total time to traverse the world. The slope of the dashed magenta and cyan line is the global reward rate *after accounting for the opportunity cost*. **F.** The global reward rate after accounting for the opportunity cost (dashed magenta and gold), projected from the origin over the time of the pursuit is the apportionment cost of time (brown vertical line). It is the amount of the opportunity cost-

subtracted reward that would occur on average over the pursuit's time under a policy of accepting the pursuit. **G**. Time's cost is apportionment cost plus the opportunity cost. Whether expressed in terms of the global reward rate achieved under a policy of accepting the considered pursuit (A-B) or from the perspective of the outside time and reward (C-G), the subjective value expressions are equivalent.

What, then, is the amount of reward by which the opportunity cost-subtracted reward is scaled down to equal the *sv* of the pursuit? This amount is the *apportionment cost of time*. The apportionment cost of time **(height of the brown vertical bar, Figure 5F**) is the global reward rate *after taking into account the opportunity cost* (slope of the magenta-gold dashed line in **Figure 5F**) times the time of the considered pursuit. Equally, the difference between the inside and outside reward rates, times the time of the pursuit, is the apportionment cost when scaled by the pursuit's weight, i.e., the fraction that the considered pursuit is to the total time to traverse the world (**Equation 9, right hand side)**. From the perspective of decision-making policies*, apportionment cost* is the difference in reward that can be expected, on average, between a policy of *taking* versus a policy of *not taking* the considered pursuit, over a time equal to its duration **(Equation 9 center, Figure 5F).**

$$\boldsymbol{Apportionment\ Cost} = \boldsymbol{t_{in}}\left(\boldsymbol{\rho_g} - \boldsymbol{\rho_{out}}\right) = \boldsymbol{t_{in}}\left(\frac{\boldsymbol{r_{in}}}{\boldsymbol{t_{in}}} - \frac{\boldsymbol{r_{out}}}{\boldsymbol{t_{out}}}\right)\left(\frac{\boldsymbol{t_{in}}}{\boldsymbol{t_{in}} + \boldsymbol{t_{out}}}\right)$$

Equation 9. Apportionment Cost.

While this difference is the apportionment cost of time, the opportunity cost of time is the amount that would be expected from a policy of *not* taking the considered pursuit over a time equal to the considered pursuit's duration. Together, they sum to Time's Cost (**Figure 5G**). Expressing a pursuit's worth in terms of the global reward rate obtained under a policy of accepting the pursuit type (**Figure 5 left column**), or from the perspective of the outside reward and time (**Figure 5 right column)**, are equivalent. However, the latter expresses *sv* in terms that are independent of one another, conveys the constituents giving rise to global reward rate, and provides the added insight that time's cost comprises an apportionment as well as an opportunity cost.

The effect of increasing the outside reward on the subjective value of a pursuit

Let us now consider the impact that changing the outside reward and/or outside time has on these two determinants of time's cost—opportunity and apportionment cost—to further our understanding of the subjective value of a pursuit. **Figure 6** illustrates the impact of changing the reward reaped from outside the pursuit on the pursuit's subjective value. By holding the time spent outside the considered pursuit constant, changing the outside reward thus changes the outside reward rate. When the considered pursuit's reward rate is greater than its outside reward rate, the subjective value is positive **(Figure 6A, green dot).** The subjective value diminishes linearly (**Figure 6B, green dots**) to zero as the outside reward rate increases to match the pursuit's reward rate, and turns negative as the outside reward rate exceeds the pursuit's reward rate, indicating that a policy of accepting the considered pursuit would result in a lower attained global reward rate than that garnered under a policy of forgoing the pursuit. Under these conditions, the subjective value is shown to decrease linearly **(Figure 6C, green dotted line)** as the outside reward increases because the cost of time increases linearly **(Figure 6C, maroon dotted region)**.

Time's cost is the sum of the opportunity cost **(Figure 6D)** and apportionment cost **(Figure 6E)** of time. **Figure 6F** overlays the opportunity cost and apportionment cost of time to demonstrate how their sum constitutes time's cost, and thus accounts for subjective value **(green dotted line)**. When the outside reward is zero, there is zero opportunity cost of time **(Figure 6D,F orange dots)**, with time's cost being entirely constituted by the apportionment cost of time **(Figure 6E,F brown annuli)**. Apportionment cost **(Figure 6E,F)** decreases as outside reward increases because the difference between the inside and outside reward rate diminishes, thus making how time is apportioned in and outside the pursuit less relevant. At the same time, as outside reward increases, the opportunity cost of time increases **(Figure 6D)**. When inside and outside rates are the same, how the agent apportions its time in or outside the pursuit does not impact the global rate of reward. At this point, the apportionment cost of time has fallen to zero, while the

opportunity cost of the pursuit has now come to entirely constitute time's cost. Further increases in the outside reward now result in the outside rate being increasingly greater than the inside rate making the apportionment of time in/outside the pursuit increasingly relevant. Now, however, though the opportunity cost of time continues to grow positively, the apportionment cost of time grows increasingly *negative* (which is to say the pursuit has an apportionment *gain*). Summing the opportunity cost of the pursuit and the now *negative* apportionment cost (i.e., the apportionment gain) continues to yield time's cost; therefore, subtracting time's cost from the pursuit's reward continues to yield the subjective value of the pursuit (green dotted line).

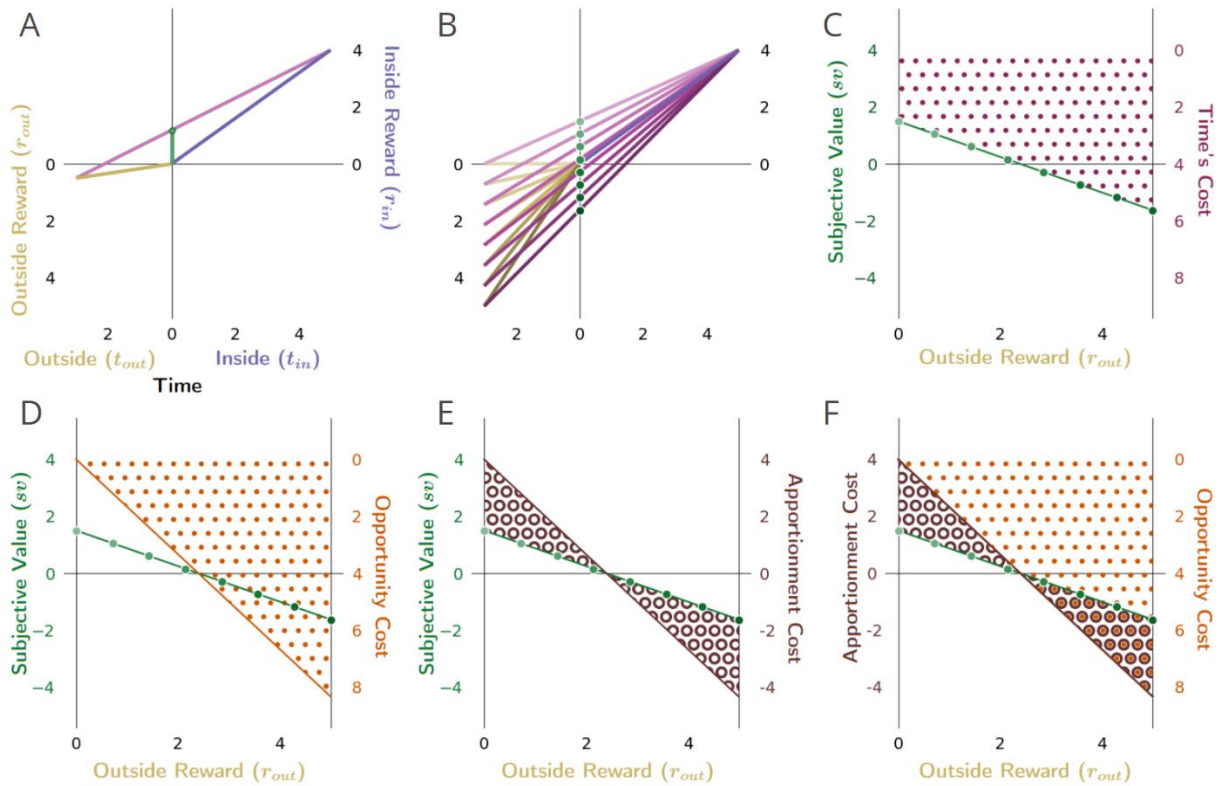


Figure 6. The impact of outside reward on the subjective value of a pursuit. A) The subjective value (green dot) of a considered pursuit type (purple) in the context of its 'outside' (gold) is the resulting global reward rate vector's (magenta) intersection of the y-axis in the present (t = 0). **B)** A family of vectors showing that increasing the outside reward while holding the outside time constant increases the outside reward rate (slope of gold lines), resulting in increasing the global reward rate (slope of the purple lines), and decreasing the subjective value (green dots) of the pursuit. As the reward rate of the environment outside the considered pursuit type increases from lower than to higher than that of the considered pursuit, the subjective value of the pursuit decreases, becomes zero when the in/outside rates are equivalent, and goes negative when ρ_{out} exceeds ρ_{in}. **C)** Plotting the subjective value of the pursuit (green dots in B) as a function of increasing the outside reward (green dotted line in C through F) reveals that the subjective value of the pursuit decreases linearly. This linear decrease is due to the linear increase in the cost of time of the pursuit (maroon dotted region) as outside reward, and thus global reward rate, increases. Subtracting time's cost from the pursuit's reward yields the pursuit's subjective value as outside reward increases (green-dotted line, C through F).

Time's cost (the area, as in C, between the pursuit's reward magnitude and its subjective value) is the sum of D) the opportunity cost of time (orange dotted region), and E) the apportionment cost of time (brown annuli region), as shown in F. **D)** As the outside reward increases, the opportunity cost of time increases linearly. **E)** As the outside reward increases, the apportionment cost of time decreases linearly, becomes zero when the inside and outside reward rates are equal, and then becomes negative (becomes an apportionment gain). **F)** When the outside reward rate is zero, time's cost is composed entirely of an apportionment cost. As the outside reward increases, opportunity cost increases linearly as apportionment cost decreases linearly, summing to time's cost. This continues until the reward rates in and outside the pursuit become equivalent, at which point the subjective value of the pursuit is zero. When subjective value is zero, the cost of time is entirely composed of opportunity cost. As the outside rate exceeds the inside rate, opportunity cost continues to increase while the apportionment cost becomes increasingly negative (which is to say, the apportionment cost of time becomes an apportionment gain of time). Adding the

positive opportunity cost and the now negative apportionment cost (the region of overlap between brown annuli & orange dots) continues to sum to yield time's cost.

The effect of changing the outside time on the subjective value of the considered pursuit

Figure 7 examines the effect of changing the outside time on the subjective value of a pursuit, while holding the outside reward constant at a value of zero **(Figure 7A,B)**. Doing so affords a means to examine the apportionment cost of time in isolation from the opportunity cost of time. Despite there being no opportunity cost, there *is* yet a cost of time **(Figure 7C)** composed entirely of the apportionment cost **(Figure 7D)**. When the portion of time spent outside dominates, time's apportionment cost of the pursuit is small. As the portion of time spent outside the pursuit decreases and the relative apportionment of time spent in the pursuit increases, the apportionment cost of the pursuit increases purely hyperbolically, resulting in the subjective value of the pursuit decreasing purely hyperbolically **(Figure 7D)**. As time spent outside the considered pursuit becomes diminishingly small, the pursuit comprises more and more of the world, until the apportionment of time is entirely devoted to the pursuit, at which point the apportionment cost of time equals the pursuit's reward rate (ρ_{in}) times the pursuit's time (t_{in}) (i.e., the pursuit's reward magnitude, r_{in}).

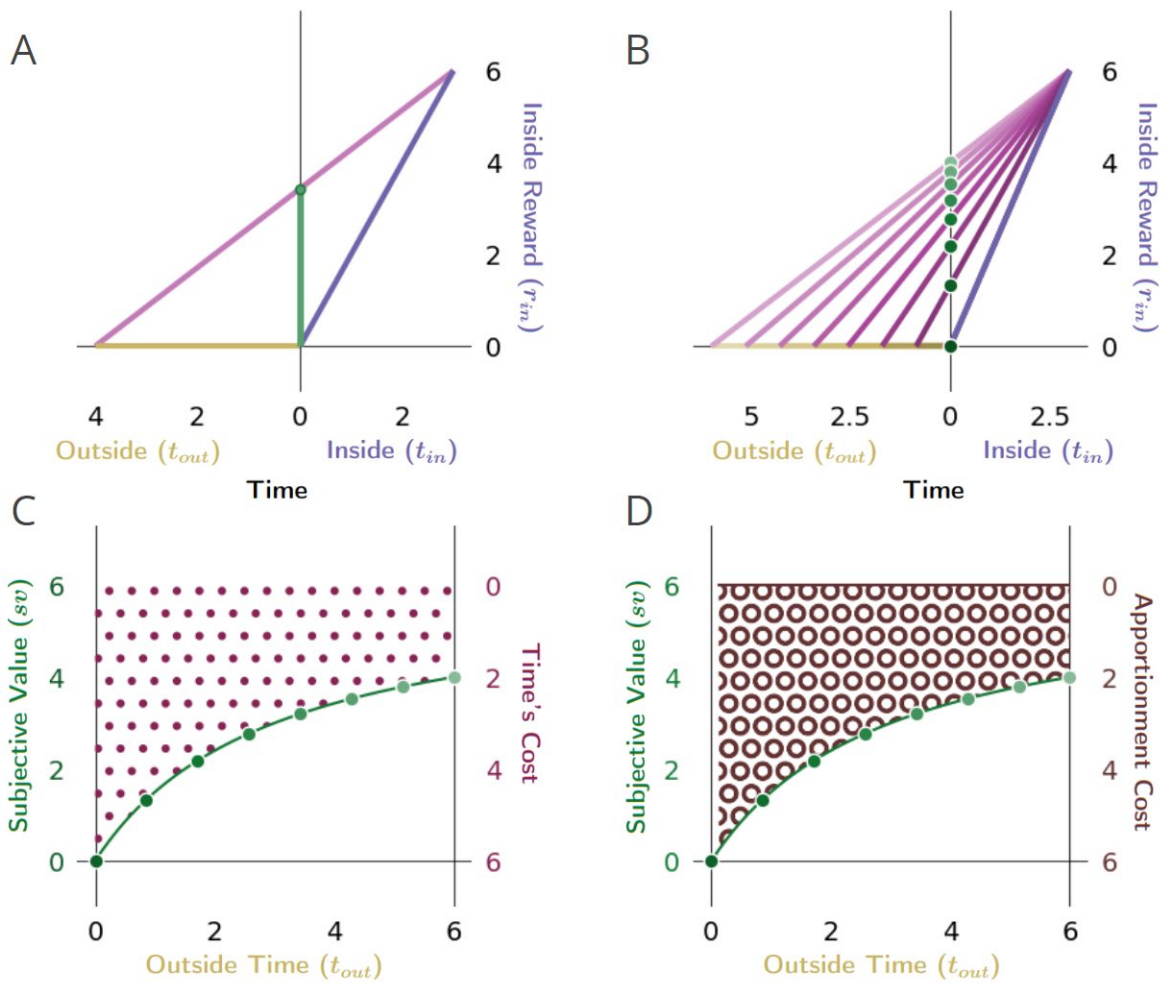


Figure 7. The impact of the apportionment cost of time on the subjective value of a pursuit. **A)** The apportionment cost of time can best be illustrated dissociated from the contribution of the opportunity cost of time by considering the special instance in which the outside has no net reward, and therefore has a reward rate of zero. **B)** In such instances, the considered pursuit type (purple vector) still has a cost of time, however, which is entirely composed of apportionment cost. Time's apportionment cost

will decrease as the time in the outside (the family of outside time golden vectors) increases, decreasing the slope of corresponding global reward rate vectors (family of magenta vectors) and, thus, increasing the subjective value (green dots) of the considered pursuit. **C)** Here, the cost of time is entirely composed of apportionment cost, which arises from the fact that the considered pursuit is contributing its proportion to the global reward rate. How much the pursuit's time cost is (maroon dots) is thus determined by the ratio of the time spent in the pursuit versus outside the pursuit; the more time is spent outside the pursuit, the less the apportionment cost of time of the pursuit, and therefore, the greater the subjective value of the pursuit (green dotted line). **D)** When apportionment cost (brown annuli) solely composes the cost of time, the cost of time decreases hyperbolically as the outside time increases, resulting in the subjective value of a pursuit increasing hyperbolically (green dotted line). Conventions as in **Figure 6**.

The effect of changing the outside time and outside reward rate on the subjective value of a pursuit

In having examined the effect of varying outside reward and thus outside reward rate **(Figure 6),** and outside time while holding outside reward rate constant **(Figure 7)**, let us now consider the impact of varying the outside time and the outside reward rate **(Figure 8)**. By changing the outside time while holding the outside reward constant **(Figure 8B)**, the reward rate obtained in the outside will be varied while the apportionment of time in & outside the pursuit changes, thus impacting the opportunity and apportionment cost of time. Plotting the subjective value-by-outside time function (**green dotted line**) then reveals that subjective value increases hyperbolically under these conditions as outside time increases, which is to say, time's cost decreases hyperbolically **(Figure 8C)**. Decomposing time's cost into its constituent opportunity **(Figure 8C)** and apportionment costs **(Figure 8D)** illustrates how these components vary when varying outside time. Opportunity cost **(orange dots, Figure 8C)** decreases hyperbolically as the outside time increases. Apportionment cost varies as the difference of two hyperbolas **(brown annuli area, Figure 8D)**, its sign being initially negative (an apportionment gain), decreases to zero as the outside and inside rates become equal, and then becomes positive (an apportionment cost). Taken together **(Figure 8F)**, their sum (opportunity and apportionment costs) decreases hyperbolically as outside time increases, resulting in subjective values that hyperbolically increase, spanning from the negative of the outside reward magnitude to the inside reward magnitude.

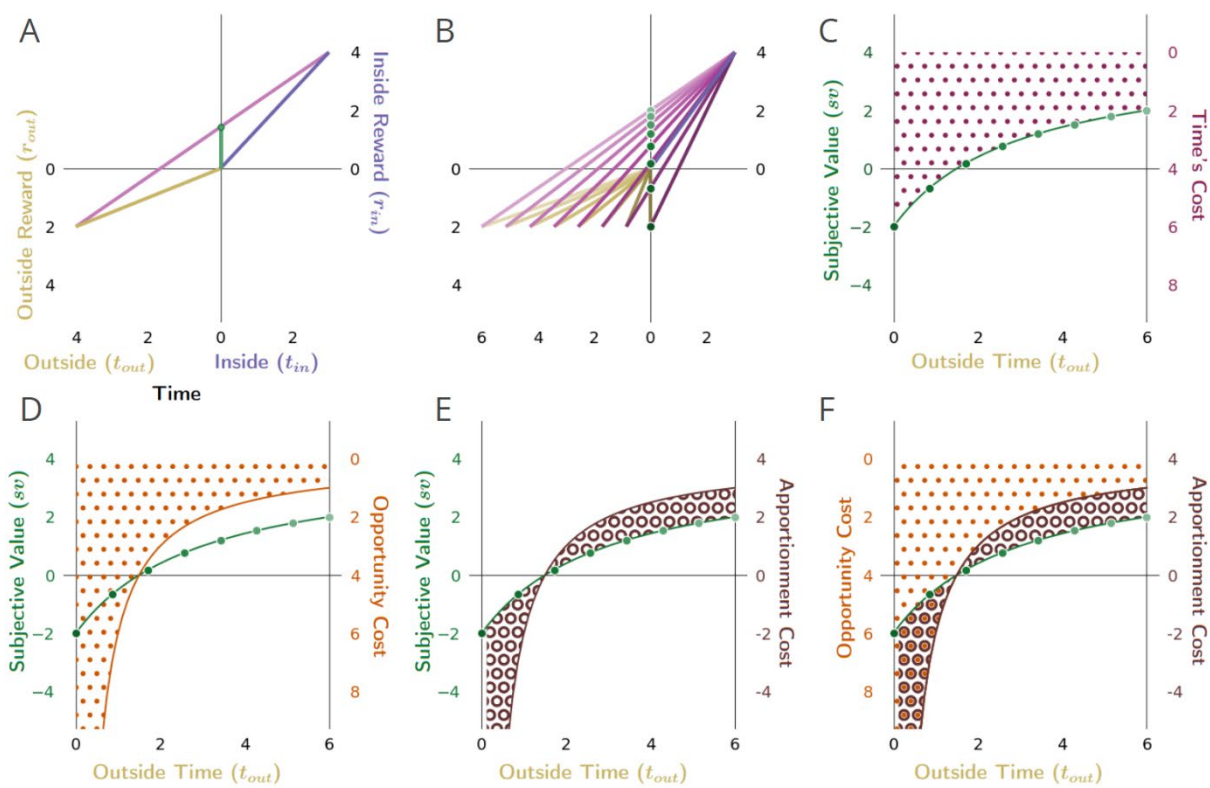


Figure 8. The effect of changing the outside time and thus the outside reward rate, on the subjective value of a pursuit. **A)** The subjective value (green dot) of the considered pursuit, when **B)** changing the outside time, and thus, outside reward rate (green

dots). **C)** As outside time increases under these conditions (holding outside reward constant), the subjective value of the pursuit increases hyperbolically (green dotted line), from the negative of the outside reward magnitude to, in the limit, the inside reward magnitude. Conversely, time's cost (right hand axis, maroon dots) decreases hyperbolically. **D)** Opportunity cost (orange stippling, right hand y-axis) decreases hyperbolically as outside time increases. **E)** Apportionment cost (right hand y-axis) is the difference of two hyperbolas (brown annuli area). It is initially negative, meaning that apportionment cost is a *gain*, as the outside reward rate is greater than the inside reward rate. This apportionment gain decreases to zero as outside time increases to make the outside and inside rates equal, and then becomes an apportionment cost (positive). **F)** Summing opportunity cost and apportionment cost yields time's cost; subtracting time's cost from the pursuit's reward magnitude yields the subjective value (green dotted curve). Conventions as in **Figure 6**.

The value of initiating pursuits in choice decision-making

Above, we determined how a reward-rate-maximizing agent would evaluate the worth of a pursuit, thereby identifying the impact of a policy of taking (or forgoing) that pursuit on the realized global reward rate and expressing that pursuit's worth as subjective value. We did so by opposing a pursuit with its equivalent offer requiring no time investment—a special and instructive case. In this section we consider what decision should be made when an agent is simultaneously presented with a choice of more than one pursuit of any potential magnitude and time investment. Using the subjective value under these choice decisions, we more thoroughly examine how the duration and magnitude of a pursuit, and the context in which it is embedded (its 'outside'), impacts reward-rate-optimal valuation. We then re-express subjective value as a temporal discounting function, revealing the nature of the *apparent* temporal discounting function of a reward-rate-maximizing agent as one determined wholly by the temporal structure and magnitude of rewards in the environment. We then assess whether hyperbolic discounting, the "Delay" effect, and the "Magnitude" and "Sign" effect—purported signs of suboptimal decision-making (McDiarmid and Rilling, 1965; Rachlin et al., 1972; Ainslie, 1974; Snyderman, 1983; Green et al., 1994; Kirby and Herrnstein, 1995; Myerson and Green, 1995; Bateson and Kacelnik, 1996; Ostaszewski, 1996; Stephens and Anderson, 2001; Cheng et al., 2002; Frederick et al., 2002; Hayden and Platt, 2007; Hayden et al., 2007; McClure et al., 2007; Beran and Evans, 2009; Peters and Büchel, 2011; Stevens and Mühlhoff, 2012; Carter et al., 2015; Carter and Redish, 2016)—are in fact consistent with optimal decision-making.

Choice decision-making

Consider a temporal decision in which two or more mutually exclusive options are simultaneously presented following a period that is common to policies of choosing one or another of the considered options **(Figure 9)**. In such scenarios, subjects choose between outcomes differing in magnitude and the time at which they will be delivered. Of particular interest are choices between a smaller, sooner reward pursuit ("SS" pursuit) and a larger, later reward pursuit ("LL" pursuit) (Myerson and Green, 1995; Frederick et al., 2002; Madden and Bickel, 2010; Peters and Büchel, 2011). Such intertemporal decision-making is commonplace in the laboratory setting (McDiarmid and Rilling, 1965; Rachlin et al., 1972; Ainslie, 1974; Snyderman, 1983; Myerson and Green, 1995; Bateson and Kacelnik, 1996; Ostaszewski, 1996; Stephens and Anderson, 2001; Cheng et al., 2002; Frederick et al., 2002; Hayden et al., 2007; Hayden and Platt, 2007; McClure et al., 2007; Beran and Evans, 2009; Peters and Büchel, 2011; Stevens and Mühlhoff, 2012; Carter et al., 2015; Carter and Redish, 2016).

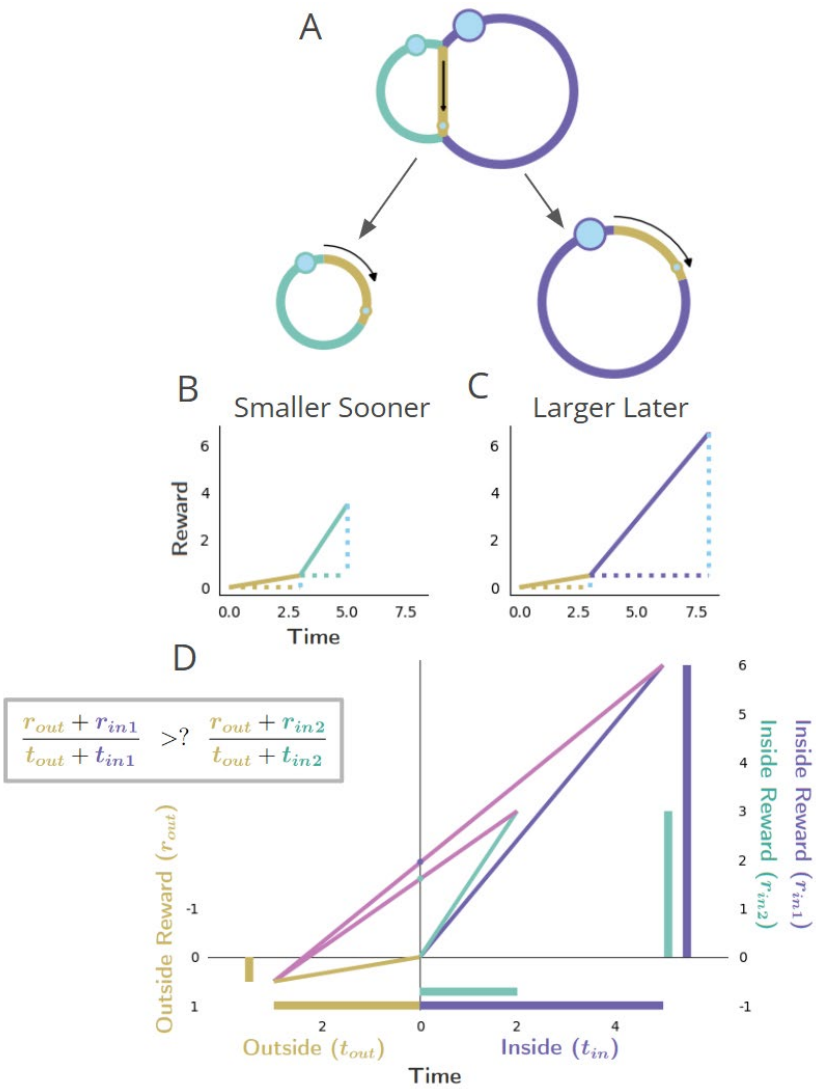


Figure 9. Policy options considered during the initiation of pursuits in worlds with a "Choice" topology. A-C) Choice topology, and policies of choosing the smaller-sooner or larger-later pursuit, as in **Figure 1** "Choice". **D)** The world divided into "Inside" and "Outside" the selected pursuit type, as the agent decides whether to accept SS (aqua) or LL (purple) pursuit. The global reward rate (ρ_g) under a policy of choosing the SS or LL (slopes of the magenta lines) is calculated from the equation in the box to the right.

Global reward rate equation and Optimal Choice Policy

With the global reward rate equation previously derived, which choice policy (i.e., choosing SS, or LL) would maximize global reward rate can be identified. The optimal choice between the SS and the LL pursuit is as follows…

$\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} < \rho_g{}^*$, choose Smaller-Sooner, SS, pursuit (Ap 9)

$\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} = \rho_g{}^*$, both SS and LL pursuits are equivalent (Ap 9)

$\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} > \rho_g{}^*$, choose Larger-Later, LL, pursuit (Ap 9)

These policies' optimality is intuitive. By choosing option LL, the subject earns $r_{LL} - r_{SS}$ more reward than when choosing SS but spends $t_{LL} - t_{SS}$ more time. If the reward rate from that extra time spent exceeds the reward rate of the environment generally, it would be optimal to spend the extra time on the larger-later option. In other words, if the agent were to choose pursuit SS, $t_{LL} - t_{SS}$ time would be spent earning reward at a global reward rate under that policy, $\rho_{g,choose\ SS}$, with the magnitude $\rho_g(t_{LL} - t_{SS})$. If $\rho_g(t_{LL} - t_{SS})$ exceeds the extra reward $r_{LL} - r_{SS}$ that could be earned with that extra time by investing the LL pursuit, more reward would be earned in the same amount of time by choosing the SS Pursuit.

Optimal Choice Policies based on Subjective Value

As under forgo decision-making, we can now also identify the global reward rate optimizing choice policies based on subjective value **(Figure 9).** The following policies would optimize reward rate when choosing between two options of different magnitude that require different amounts of time invested:

$sv_{LL} < sv_{SS}$, take pursuit SS (Ap 10)

$sv_{LL} = sv_{SS}$, SS and LL pursuits are equivalent (Ap 10)

$sv_{LL} > sv_{SS}$, take pursuit LL (Ap 10)

The impact of opportunity & apportionment costs on choice decision-making

With optimal policies for choice expressed in terms of subjective value, the impact of time's opportunity and apportionment costs on choice decision-making can now be more deeply appreciated. Keeping the outside time constant, the opportunity cost of time increases as the outside reward (and thus the outside reward rate) increases. As increasing the outside reward rate will also impact the apportionment cost of time linearly, (see **Figure 6**), adding these costs results in a net linear decrease in the subjective value of the considered pursuits **(Figure 10)**, with the decrease being driven steadily more by the opportunity cost. When the outside reward rate is relatively small, reward-rate maximization may dictate that the agent choose the LL pursuit (as in **Figure 10**, upper left). However, as the opportunity cost of the LL pursuit is greater than that of the SS due to its greater reward and time requirement, the subjective value-by-outside time slope will be more negative for the LL than the SS pursuit, resulting in their subjective values crossing at a critical outside reward threshold (see **Figure 10**, upper middle). This critical threshold occurs when the global reward rates under a policy of accepting the LL or of accepting the SS pursuits are equal. A switch in preference from the LL pursuit to that of the SS pursuit will then be observed as outside reward values become larger, past this critical threshold (see **Figure 10**, upper right).

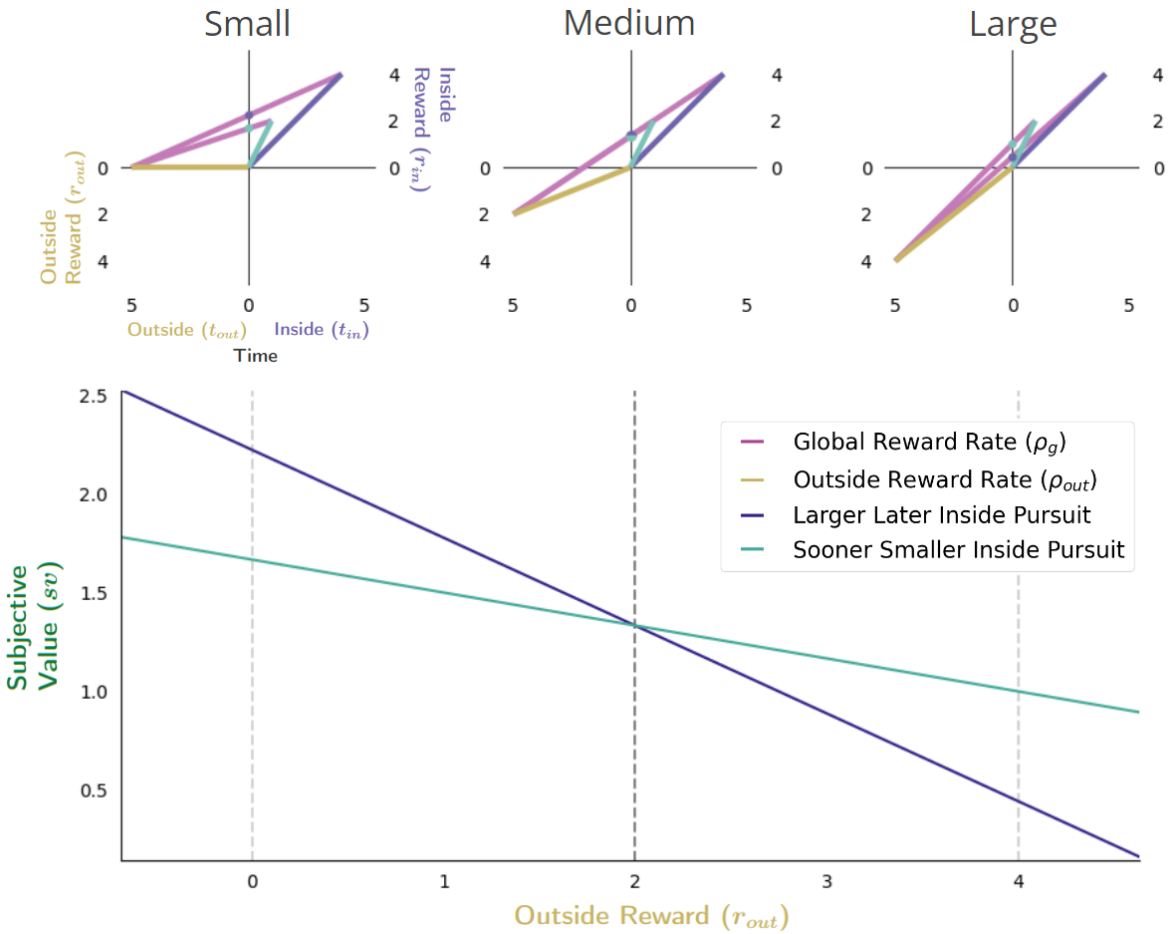


Figure 10. The effect of increasing outside reward on subjective value in choice decision-making. The effect of increasing the outside reward while holding the outside time constant is to linearly increase the cost of time, thus decreasing the subjective value of pursuits considered in choice decision-making. When the outside reward is sufficiently small **(top left)**, the subjective value of the LL pursuit can exceed the SS pursuit, indicating that selection of the LL pursuit would maximize the global reward rate. As outside reward increases, however, the subjective value of pursuits will decrease linearly as the opportunity cost of time increases. Since a policy of choosing the LL pursuit will have the greater opportunity cost, the slope of its function relating subjective value to outside reward will be greater than that of a policy of choosing the SS pursuit (**bottom**). Thus, outside reward can be increased sufficiently such that the subjective value of the LL and SS pursuits will become equal **(top middle)**, past which the agent will switch to choosing the SS pursuit **(top right)**. Vertical dashed lines (bottom) correspond to instances along top.

A switch in preference between the SS and LL pursuits will also occur when the time spent outside the considered pursuit increases, even if the outside reward rate earned remains constant **(Figure 11).** In this case, the opportunity cost of time will remain unchanged, with only the apportionment cost of time driving changes in the subjective value (the aqua (SS) and purple (LL) dots on the y-axis, **Figure 11**) of pursuits. As any inside time will constitute a greater fraction of the total time under a LL versus a SS pursuit policy, the apportionment cost of the LL pursuit will be greater. This can result in the subjective value of the SS pursuit being greater initially (when outside time is short, **Figure 11 upper left**), than the LL pursuit. As the outside time increases, however, the ordering of subjective value will switch as apportionment costs becoming diminishingly small.

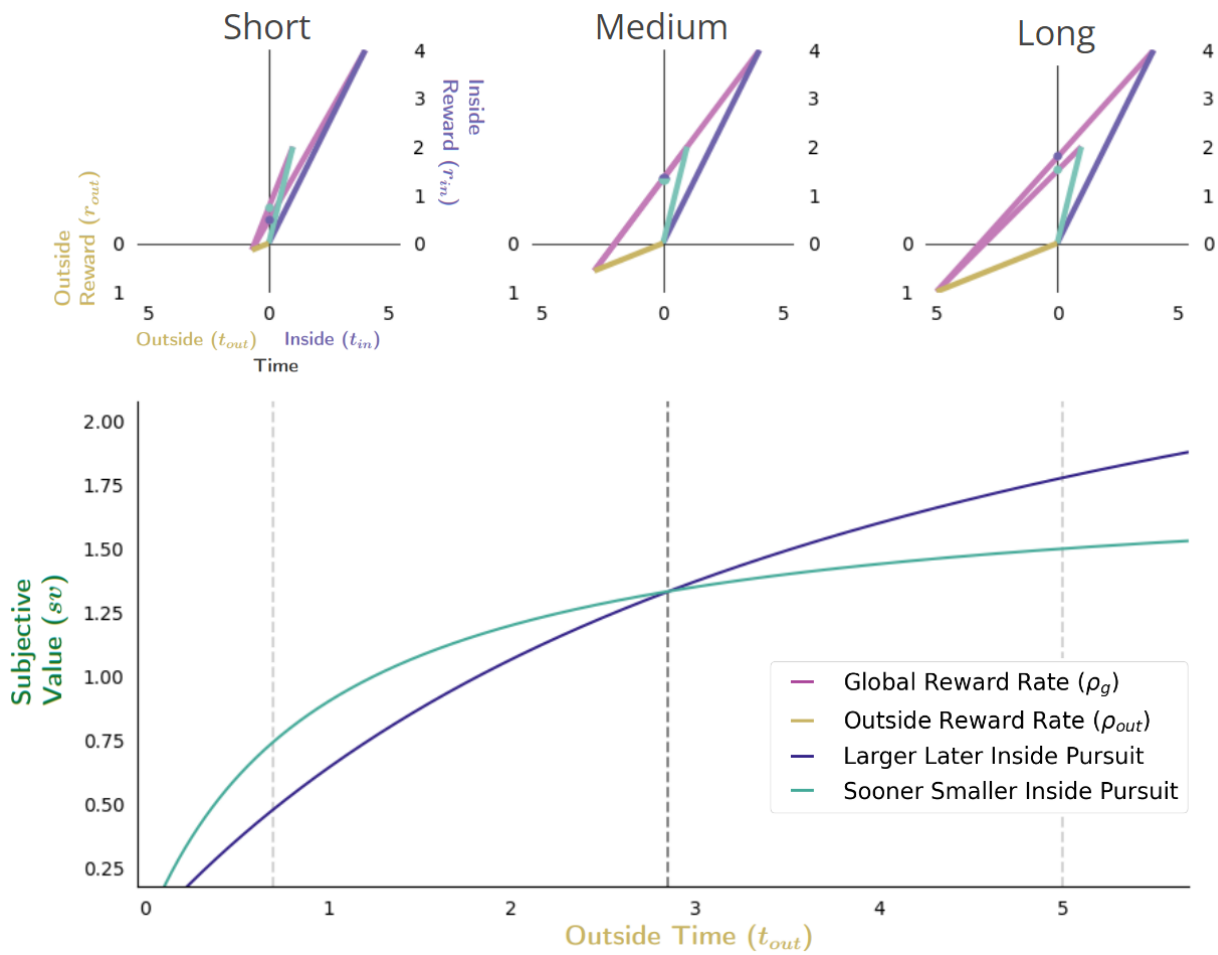


Figure 11. Effect of apportionment cost on subjective value in choice decision-making. The effect of apportionment cost can be isolated from the effect of opportunity cost by increasing the outside time while holding outside rate constant. Doing so results in decreasing the apportionment cost of the considered pursuit, thus increasing its subjective value **(bottom)**. When the outside time is sufficiently small **(top left)**, the apportionment cost for LL and SS pursuits will be relatively large, but can be greater still for the LL pursuit given its proportionally longer duration to the outside time and greater reward magnitude. As outside reward time increases, however, the subjective value of pursuits increase as the apportionment cost of time of the considered pursuit decreases. As apportionment costs diminish and the magnitudes of pursuits' rewards become more fully realized, the subjective value of the LL and SS pursuits will become equal **(top middle)**, with the LL pursuit eventually exceeding that of the SS pursuit at sufficiently long outside times **(top right)**. Vertical dashed lines (bottom) correspond to instances along top.

Finally, the effect of varying the outside time and the outside reward rate on subjective value in Choice behavior is considered (**Figure 12**). Opportunity and apportionment costs can simultaneously be

varied, for instance, by maintaining outside reward but increasing outside time. Recall that doing so decreases the opportunity cost of time hyperbolically while the apportionment cost of time varies as the difference of two hyperbolas; their sum (time's cost) decreases hyperbolically (**see Figure 8).** Therefore, the subjective value-by-outside time functions of the LL and SS pursuits *increase* hyperbolically (**Figure 12, bottom**), but since they span different ranges (from the negative of the outside reward magnitude to their respective inside reward magnitudes) these functions will cross, thus switching from the SS to the LL pursuit (**Figure 12**) at a critical outside time.

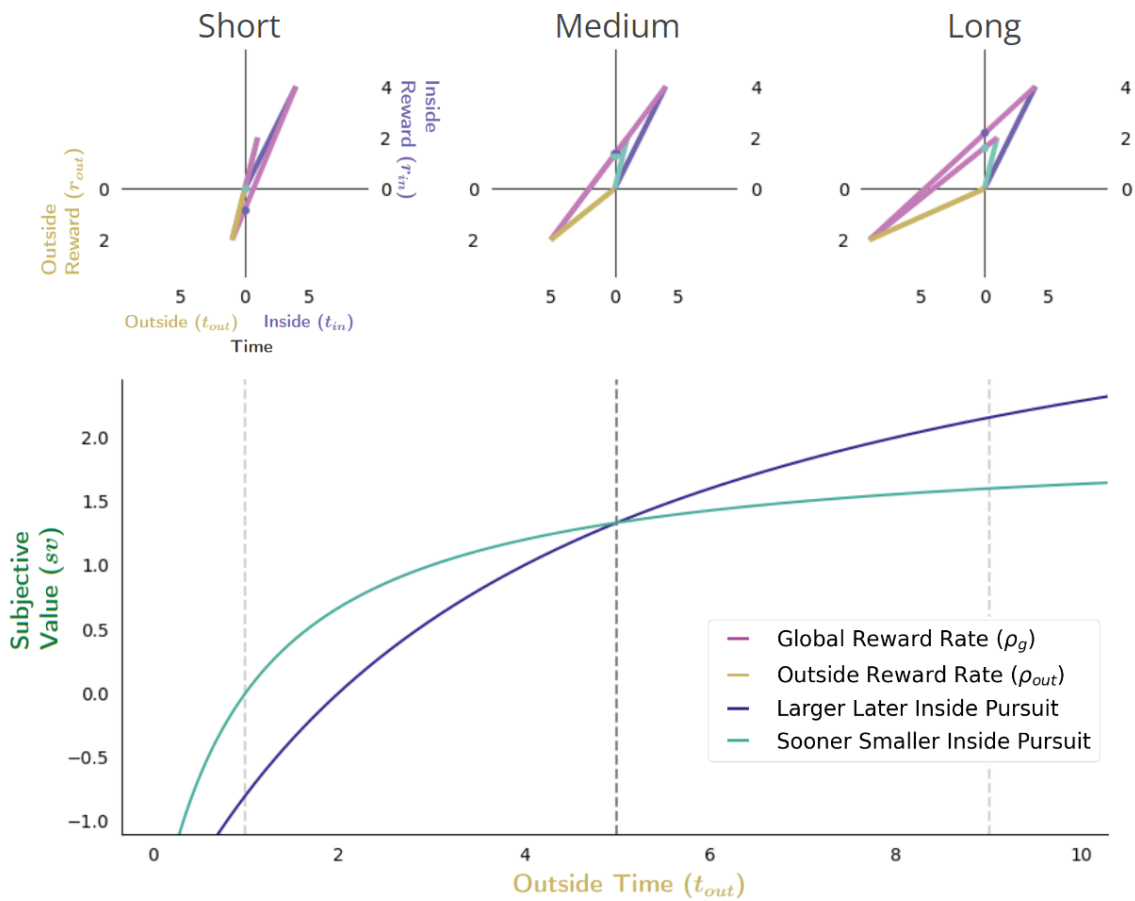


Figure 12. Effect of varying outside time and outside reward rate. The effect of increasing the outside time while maintaining outside reward is to decrease the apportionment as well as the opportunity cost of time, thus increasing pursuit's subjective value. Increasing outside time, which in turn, also decreases outside reward rate, results in the agent appearing as if to become more patient, being willing to switch from a policy of selecting the SS pursuit **(upper left)** to a policy of treating them equally **(upper middle)** at some critical threshold **(vertical dashed black line, bottom)**, to a policy of selecting the LL pursuit **(upper right)**. Vertical dashed lines (bottom) correspond to instances along top.

A reward-rate-optimal agent will thus appear as if more patient 1) the longer the time spent outside a considered pursuit, 2) the lower the outside reward rate, or 3) both, switching from a policy of choosing the SS to choosing the LL option at some critical outside reward rate and/or time. Having analyzed the impact of time spent and reward obtained outside a pursuit on a pursuit's valuation, we now examine the impact time spent within a pursuit has on its valuation.

The Discounting Function of a reward-rate-optimal agent

How does the value of a pursuit change as the time required for its obtainment grows? Intertemporal decision-making between pursuits requiring differing time investments resulting in different reward

magnitudes has typically been examined using a 'temporal discounting function' to describe how delays in reward influence their valuation. This question has been investigated experimentally by pitting SS against LL options to experimentally determine the *subjective value* of the delayed reward (Mischel, Grusec, & Masters, 1969), with the best fit to many such observations across delays determining the subjective value-time function. After normalizing by the magnitude of reward, the curve of subjective values as a function of delay is the "temporal discounting function" ((Green et al., 1994; Kirby and Herrnstein, 1995), and for review see (Frederick et al., 2002)). While the temporal discounting function has historically been used in many fields, including economics, psychology, ethology, and neuroscience to describe how delays influence rewards' subjective value, its origins—from a normative perspective—remain unclear (Hayden, 2015). What, then, is the temporal discounting function of a reward-rate-optimal agent? And would its determination provide insight into why experimentally derived discounting functions present in the way they do, with their varied forms and curious sensitivity to the context, delay, magnitude, and sign of pursuit outcomes?

The Discounting Function of an Optimal Agent is a Hyperbolic Function

To examine the temporal discounting function of a reward-rate-optimal agent, we begin with the subjective value-time function introduced in **Equation 8**, rearranging the apportionment scaling term to resemble the standard temporal discounting function form…

$$\boldsymbol{sv(r,t) = (r_{in} - \rho_{out} t_{in})\left(\frac{t_{out}}{t_{out} + t_{in}}\right) = \frac{r_{in} - \rho_{out} t_{in}}{1 + \frac{1}{t_{out}} t_{in}}}$$

Equation 10. Subjective value-time function rearranged to resemble the standard temporal discounting function form

The temporal discounting function of an optimal agent can then be expressed by normalizing this subjective value-time function by the considered pursuit's magnitude (**Equation 11**).

$$\boldsymbol{Discounting\ Function = \frac{sv(r,t)}{r} = \frac{r_{in} - \rho_{out} t_{in}}{1 + \frac{1}{t_{out}} t_{in}} * \left(\frac{1}{r_{in}}\right) = \frac{1 - \rho_{out} * \frac{t_{in}}{r_{in}}}{1 + \frac{1}{t_{out}} t_{in}}}$$

Equation 11. The Discounting Function of a global reward-rate-optimal agent

To illustrate the discounting function of a reward-rate maximal agent, **Figure 13** depicts how the worth of a pursuit's reward would change as its context—the 'outside' world in which it is embedded—changes. We do so by examining the apparent discounting function in three different world contexts: a world in which there is, A) zero outside reward rate & large outside time, B) zero outside reward rate & small outside time, and, C) positive outside reward rate & small outside time. **Figure 13** first graphically depicts the subjective values of the pursuit's reward at increasing temporal delays (the y-intercepts of the lines depicting the resulting global reward rates, green dots) in each of these world contexts (**A-C**). Then, by replotting these subjective values at their corresponding delays, the subjective value-time function is created for this increasingly delayed reward in each of these worlds (**D-F**). By normalizing by the reward magnitude, these subjective value-time functions are then converted to their corresponding discounting functions (color coded) and overlaid so that their shapes may be compared (**G**).

Doing so illustrates how the mathematical form of the temporal discount function—as it appears for the optimal agent—is a hyperbolic function. This function's form depends wholly on the temporal reward structure of the environment and is composed of hyperbolic and linear components which relate to the apportionment and to the opportunity cost of time. To best appreciate the contributions of opportunity and apportionment costs to the discounting function of a reward rate-optimal agent, consider the following instances exemplified in **Figure 13**. First, in worlds in which no net reward is received outside a considered pursuit, the apparent discounting function is *purely* hyperbolic **(Figure 13A).** Purely hyperbolic discounting is therefore optimal when the subjective value function follows the equation…

$$sv = \frac{r_{in}}{1 + \frac{t_{in}}{t_{out}}}$$

Therefore, a purely hyperbolic discounting function would arise from a reward-rate maximizing agent in a world common to many experimental designs where, for instance, t_{out} may represent the intertrial interval with no reward. Second, as less time is apportioned outside the considered pursuit type **(Figure 13B),** this hyperbolic curve becomes more curved as the pursuit's time apportionment cost increases. The curvature of the hyperbolic component is thus controlled by how much time the agent spends 'in' versus 'outside' the considered pursuit: with the more time spent outside the pursuit, the gentler the curvature of apparent hyperbolic discounting and the more patient the agent appears to become for the considered pursuit. Third, in worlds in which reward is received outside a considered pursuit (compare **B to C**), the apparent discounting function will become steeper the more outside reward is obtained, as the linear component relating the opportunity cost of time increases.

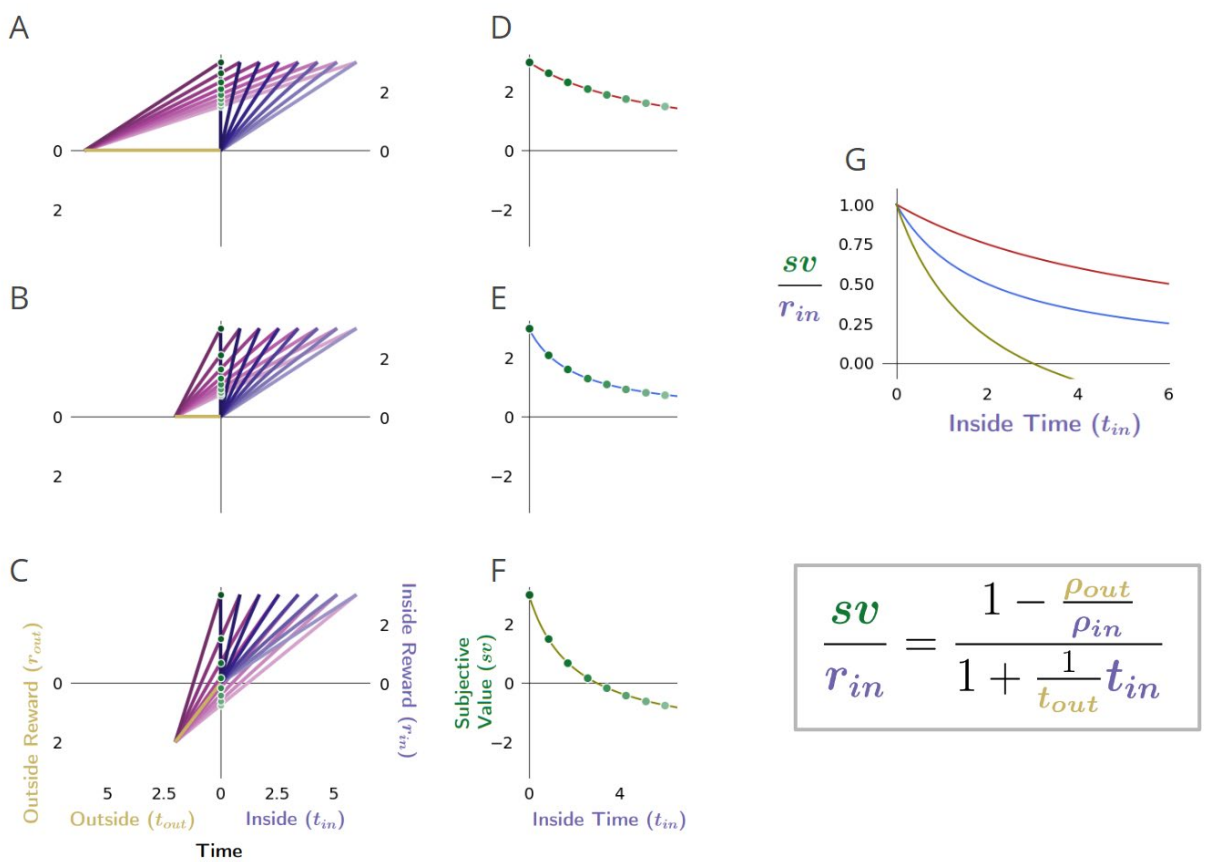


Figure 13. The temporal discounting function of a global reward-rate-optimal agent is a hyperbolic function relating the apportionment and opportunity cost of time. A-C) The effect, as exemplified in three different worlds, of varying the outside time and reward on the subjective value of a pursuit as its reward is displaced into the future. The subjective value, sv, of this pursuit as its temporal displacement into the future increases is indicated as the green dots along the y-intercept in these three different contexts: a world in which there is A) zero outside reward rate & large outside time, B) zero outside reward rate & small outside time, and C) positive outside reward rate & the small outside time as in B. **D-F)** Replotting these subjective values at their corresponding temporal displacement yields the subjective value-time function of the offered reward in each of these contexts. **G:** Normalizing these subjective value functions by the reward magnitude and superimposing the resulting temporal discounting functions reveals how the steepness and curvature of the apparent discounting function of a reward-rate-maximizing agent changes with respect to the average reward and time spent outside the considered pursuit. When the time spent outside is increased (compare

B to A)—thus decreasing the apportionment and opportunity cost of time—the temporal discounting function becomes less curved, making the agent appear as if more patient. When the outside reward is increased (compare B to C)—thus increasing the opportunity and apportionment cost of time—the temporal discounting function becomes steeper, making the agent appear as if less patient.

Thus, by expressing the worth of a pursuit as would be evaluated by a reward-rate-optimal agent in terms of its discounting function, we find that its form is consonant with what is commonly reported experimentally in humans and animals and will exhibit apparent changes in curvature and steepness that relate directly to the reward acquired and time spent outside the considered pursuit for every time spent within it.

The Delay Effect, the Magnitude Effect, and the Sign Effect

With this insight into how opportunity and apportionment costs impact the cost of time, and therefore the subjective value of pursuits in Choice decision-making, reward-rate-optimal agents are now understood to exhibit a hyperbolic form of discounting, as commonly exhibited by humans and animals (Rachlin et al., 1972, 2000; Ainslie, 1975; Thaler, 1981; Mazur, 1987; Benzion et al., 1989; Green et al., 1994; Kobayashi and Schultz, 2008; Calvert et al., 2010; Fedus et al., 2019). As hyperbolic discounting is not a sign of suboptimal decision-making, as is widely asserted, are other purported signs of suboptimal decision-making, namely the "Delay Effect", the "Magnitude Effect" and "Sign effect" also consistent with optimal temporal decisions?

The Delay Effect

The Delay Effect (also known as preference reversal) refers to the experimental observation made in Choice decisions where preference switches from the SS to the LL option as the delay to the options increase (Green et al., 1994; Kirby and Herrnstein, 1995). The switch in preference, despite the magnitudes and the time between the options remaining unchanged, is taken as a sign of anomalous decision making (Cruz Rambaud et al., 2023). **Figure 14** illustrates why preference reversal under increasing delay isn't a sign of suboptimal decision-making but rather is consistent with reward-rate maximization. When the delay to the options is sufficiently short, reward-rate maximization can dictate selecting the SS pursuit (**Figure 14 top left)** in a choice between SS and LL. As the delay increases, the global reward-rate maximizing policy under choose LL and choose SS policies would drop until they become equivalent, at which time their subjective values would be the same (**Figure 14 top middle, and dashed vertical black line Figure 14 bottom).** At delays greater than this critical value, the agent would switch to choosing the LL pursuit in Choice tasks (**Figure 14 top right).** Additionally, we note that further delay can result in the subjective value of the SS pursuit turning negative. At this point, the agent would choose the LL pursuit in a Choice context and would forgo the SS were it to be offered in a Forgo context. At even greater delay, the LL pursuit's subjective value would turn negative. This would result in the agent choosing LL under *forced* choice conditions but forgoing either pursuit in a Forgo context.

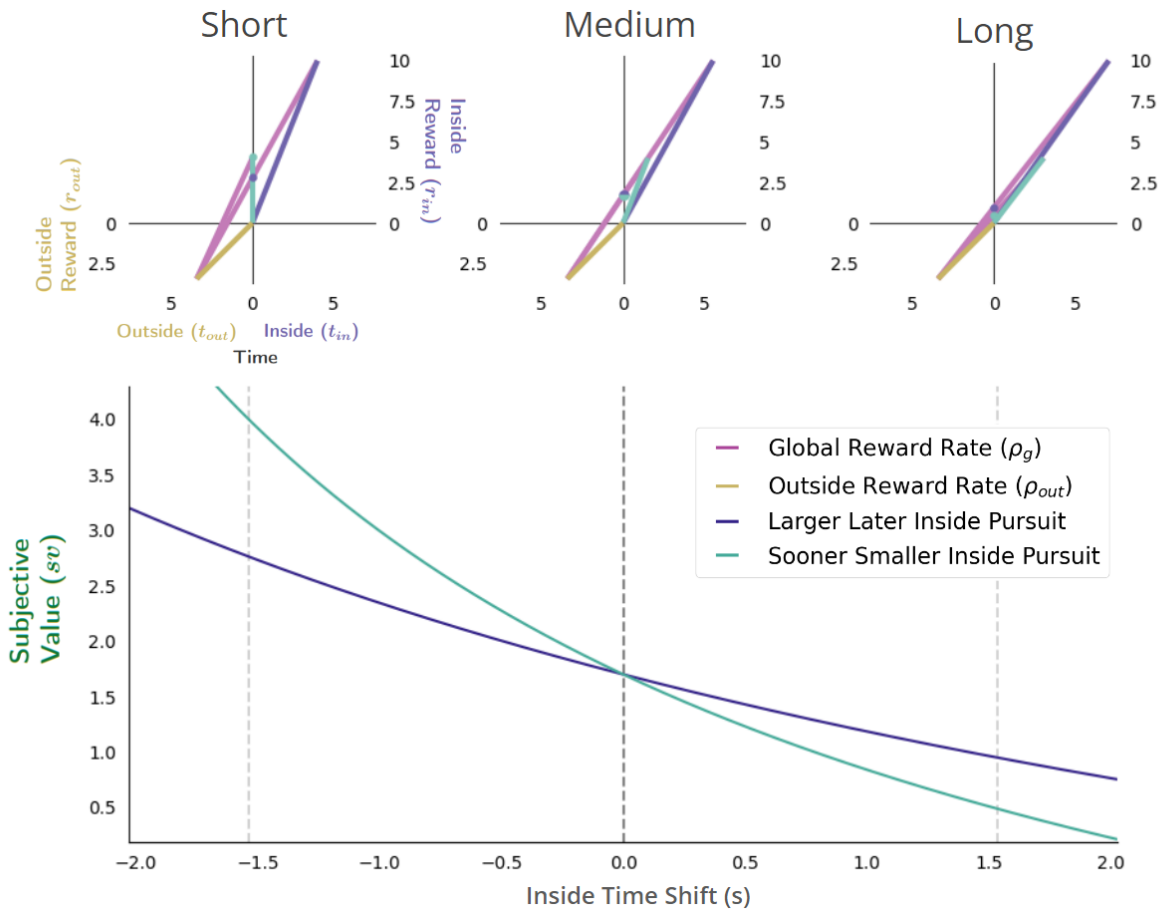


Figure 14. Reward-rate-maximizing agents would exhibit the "Delay Effect". A "switch" in preference from a SS—when the delay to the pursuits is relatively short (upper left)—to a LL pursuit, when the delay to the pursuits is relatively long (upper right), would occur as a consequence of global reward-rate maximization. **Bottom**: The subjective value-time functions varying the delay to the SS (cyan) and LL (purple) pursuit with respect to the delay in which selecting either pursuit results in the same global reward-rate. Vertical dashed lines (Bottom) correspond to instances along top.

The Magnitude Effect

The Magnitude Effect refers to the observation that the temporal discounting function, as experimentally determined, is observed to become less steep the larger the offered reward. If brains apply a discounting function to account for the delay to reward, why, as it is posed, do different magnitudes of reward appear as if discounted with different temporal discounting functions? **Figure 15** considers how a reward-rate-maximizing agent would appear to discount rewards of two magnitudes (large - top row; small - bottom row), first by determining the subjective value (green dots) of differently sized rewards **(Figure 15 A & B)** across a range of delays, and second by replotting the sv's at their corresponding delays (**Figure C & D)**, to form their subjective value functions (blue and red curves, respectively). After normalizing these subjective value functions by their corresponding reward magnitudes, the resulting temporal discounting functions that would be fit for a reward-rate-maximizing agent are then shown in **(Figure 15E).** The pursuit with the larger reward outcome (blue) thus would appear as if discounted by a less steep discounting function than the smaller pursuit (red), under what are otherwise the same circumstances. Therefore, the 'Magnitude Effect', as observed in humans and animals, would also be exhibited by a reward-rate-maximizing agent.

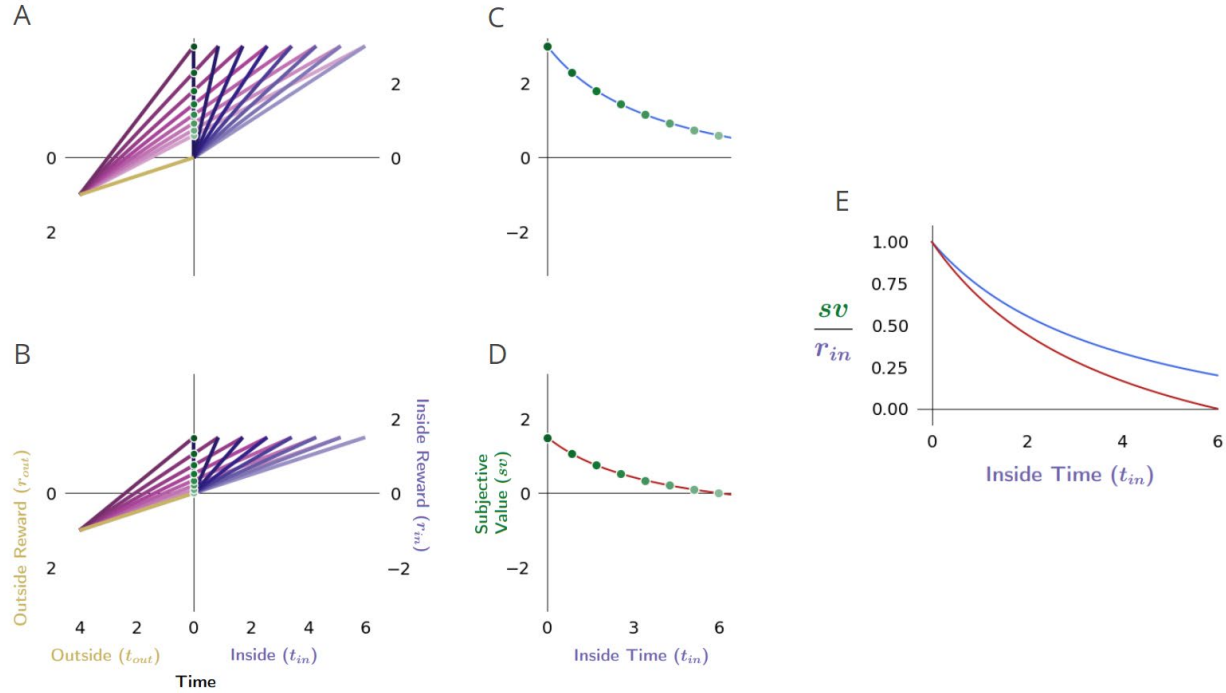


Figure 15. Reward-rate-maximizing agents would exhibit the "Magnitude effect". A&B) The global reward rate (the slope of magenta vectors) that would be obtained when acquiring a considered pursuit's reward of a given size (either relatively large as in A or small as in B) but at varying temporal removes, depicts how a considered pursuit's subjective value (green dots, y-intercept) would decrease as the time needed for its obtainment increases in environments that are otherwise the same. **C&D)** Replotting the subjective values of the considered pursuit to correspond to their required delay forms the subjective value-time function for the "large" reward case (C), and the "small" reward case (D). **E)** Normalizing the subjective value-time functions by their reward magnitude transforms these functions into their corresponding discounting functions (blue: large reward DF; red: small reward DF), and reveals that a reward-rate-maximizing agent would exhibit the "Magnitude Effect" as the steepness of the apparent discounting function would change with the size of the pursuit, and manifest as being less steep the greater the magnitude of reward.

The Sign Effect

The Sign Effect refers to the observation that the discounting functions for outcomes of the same magnitude but opposite valence (rewards and punishments) appear to discount at different rates, with punishments discounting less steeply than rewards. Should the brain apply a discounting function to outcomes to account for their temporal delays, why does it seemingly use different discount functions for rewards and punishments of the same magnitude? **Figure 16** considers how a reward-rate-maximizing agent would appear to differently discount outcomes of the same magnitude but opposite valence **(rewards-Figure 16A v punishments-Figure 16B)** when the time spent outside the considered pursuit type generates a non-zero reward rate. By determining the subjective value of these oppositely signed rewarding (**Figure 16C**) and punishing (**Figure 16D)** outcomes across a range of delays and plotting their normalized subjective values at their corresponding delay (**Figure 16E**), the apparent discounting function for reward and punishment, as expressed by a reward-rate-maximizing agent, exhibits the "Sign effect", as observed in humans and animals. In addition, we note that the difference in the discounting function slopes between rewards and punishments of equal magnitude would diminish as the outside reward rate approached zero. With an outside reward rate equal to zero, the discounting functions would be identical. If the outside reward rate became increasingly negative, the effect would invert and rewards would increasingly appear to discount less steeply than punishments.

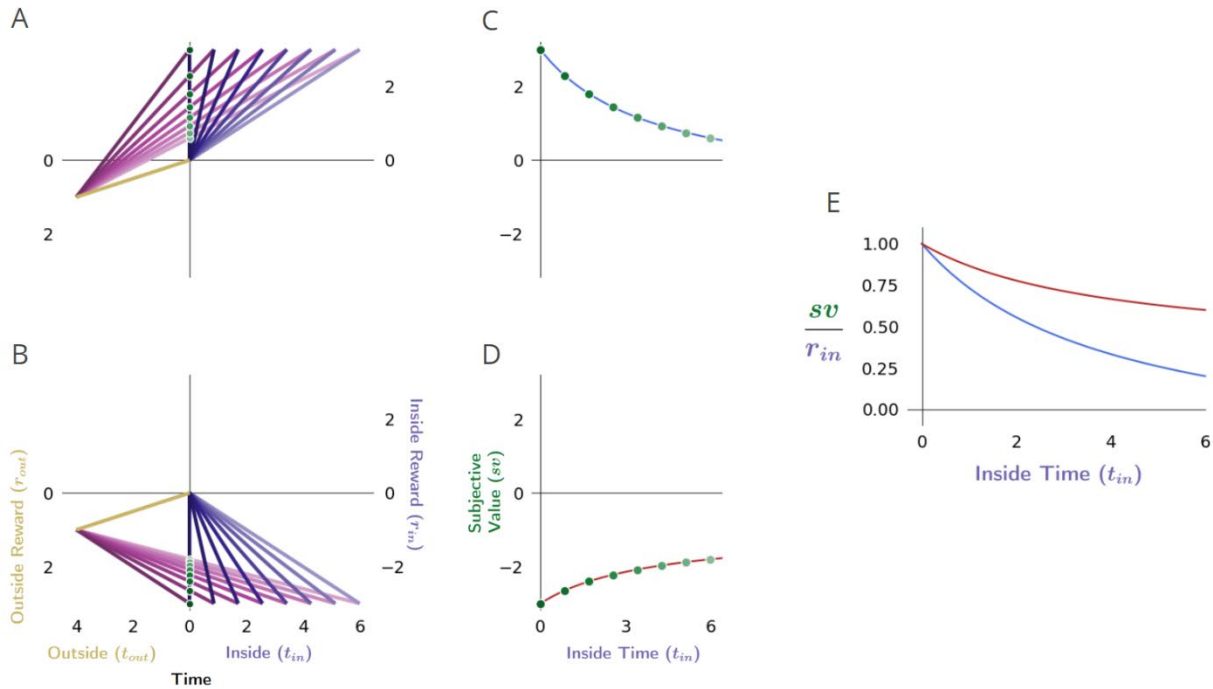


Figure 16. Reward-rate-maximizing agents would exhibit the "Sign effect". A&B) The global reward rate (the slope of magenta lines) that would be obtained when acquiring a considered pursuit's outcome of a given magnitude but differing in sign (either rewarding as in A, or punishing as in B), depicts how the subjective value (green dots, y-intercept) would decrease as the time of its obtainment increases in environments that are otherwise the same (one in which the agent spends the same amount of time and receives the same amount of reward outside the considered pursuit for every instance within it). **C&D)** Replotting the subjective values of the considered pursuit to correspond to their required delay forms the subjective value-time function for the reward (C) and for the punishment (D). **E)** Normalizing the subjective value-time functions by their outcome transforms these functions into their corresponding discounting functions (blue: reward DF; red: punishment DF). This reveals that a reward-rate-maximizing agent would exhibit the "Sign Effect", as the steepness of the apparent discounting function would change with the sign of the pursuit, manifesting as being less steep for punishing than for rewarding outcomes of the same magnitude.

Summary

In the above sections, we provide a richer understanding of the origins of time's cost in evaluating the worth of initiating a pursuit. We demonstrate that the intuitive, if deceptively simple, equation for subjective value (**Equation 7**), where time's cost is subtracted from the reward magnitude, is equivalent to subtracting an opportunity cost *and* an apportionment cost of time from the reward magnitude (**Equation 8**). Where time's cost in the simple equation is calculated from the global reward rate under a policy of accepting the considered pursuit (**Equation 7**), parceling the world into contributions from time spent 'inside' and 'outside' the considered pursuit type (**Equation 8**) reveals that the opportunity cost of time arises from the global reward rate achieved under a policy of *not* accepting the considered pursuit (its outside reward rate), and that the apportionment cost of time arises from the allocation of time spent in, versus outside, the considered pursuit. These equivalent expressions for the normatively-defined (reward-rate maximizing) subjective value of a pursuit give rise to an apparent discounting function that is 1) a hyperbolic function of time, 2) whose curvature is determined by the apportionment cost of time, and 3) whose scaling is linearly determined by the opportunity cost of time. By re-expressing reward-rate maximization as its apparent temporal discounting function, we demonstrate how fits of hyperbolic discounting, as well as observations of the Delay, Magnitude, and Sign effect—commonly taken as signs of suboptimal decision-making—are in fact consistent with optimal temporal decision-making.

Sources of error and their consequences

While these added insights enrich our understanding of time's cost and reveal how purported signs of irrationality can in fact be consistent with a reward-rate-maximizing agent, it nonetheless remains true that animals and humans *are* suboptimal temporal decision makers—exhibiting an "impatience" by selecting smaller, sooner (SS) options in cases where selecting larger, later (LL) options would maximize global reward rate. However, when decisions to accept or reject pursuits are presented in Forgo situations, they are observed to be optimal. As the equivalent immediate reward equations enabling global reward rate optimization may potentially be instantiated by neural representations of their underlying variables, we conjecture that misrepresentation of one or another variable may best explain the particular ways in which observed behavior deviates, *as well as accords*, with optimality. Therefore, we now ask what errors in temporal decision-making behavior would result from misestimating these variables, with the aim of identifying the nature of misestimation that best accounts for the pattern actually observed in animals and humans regarding whether to initiate a given pursuit.

To understand how systematic error in an agent's estimation of different time and/or reward variables would affect its behavior, we examine the agent's pattern of behavior in both Choice and Forgo decisions across different outside reward rates. First, we ask whether the agent would choose a SS or LL pursuit as in a choice task. Then we ask whether the agent would take or forgo the same LL and SS pursuits when either are presented alone in a forgo task. The actions taken by the agent can therefore be described as a triplet of policies referring to the two pursuits (e.g., **choose SS, forgo LL, forgo SS**).

Let us first consider how a reward-rate-optimal agent would transition from one to another pattern of decision-making as outside reward rate increases for the situation of fundamental interest: where the reward rate of the SS pursuit is greater than that of the LL pursuit **(Figure 17).** When the outside reward rate (slope of golden line) is sufficiently low **(Figure 17A),** the agent should prefer LL in Choice, be willing to take the LL pursuit in Forgo, and be willing to take the SS pursuit in Forgo **(**choose LL, take LL, take SS**)**. Here, a "sufficiently low" outside rate is one such that the resulting global reward rate (slope of magenta line) is less than the difference in the reward rates of the SS and LL pursuits. When the outside reward rate increases to greater than this difference in the pursuits' reward rates but is less than the reward rate of the LL option, the agent should choose SS in Choice and be willing to take either in Forgo (choose SS, take LL, take SS**) (Figure 17B).** Further increases in outside rate up to that equaling the reward rate of the SS results in the agent selecting the SS in Choice, forgoing LL in Forgo, and taking SS in Forgo (choose SS, forgo LL, take SS**) (Figure 17C).** Finally, any additional increase in outside rate would result in choosing the SS pursuit under Choice, and forgoing both pursuits in Forgo (choose SS, forgo LL, forgo SS**) (Figure 17D).** Colored regions thus describe the pattern of decision-making behavior exhibited by a reward-rate-optimal agent under any combination of outside reward and time.

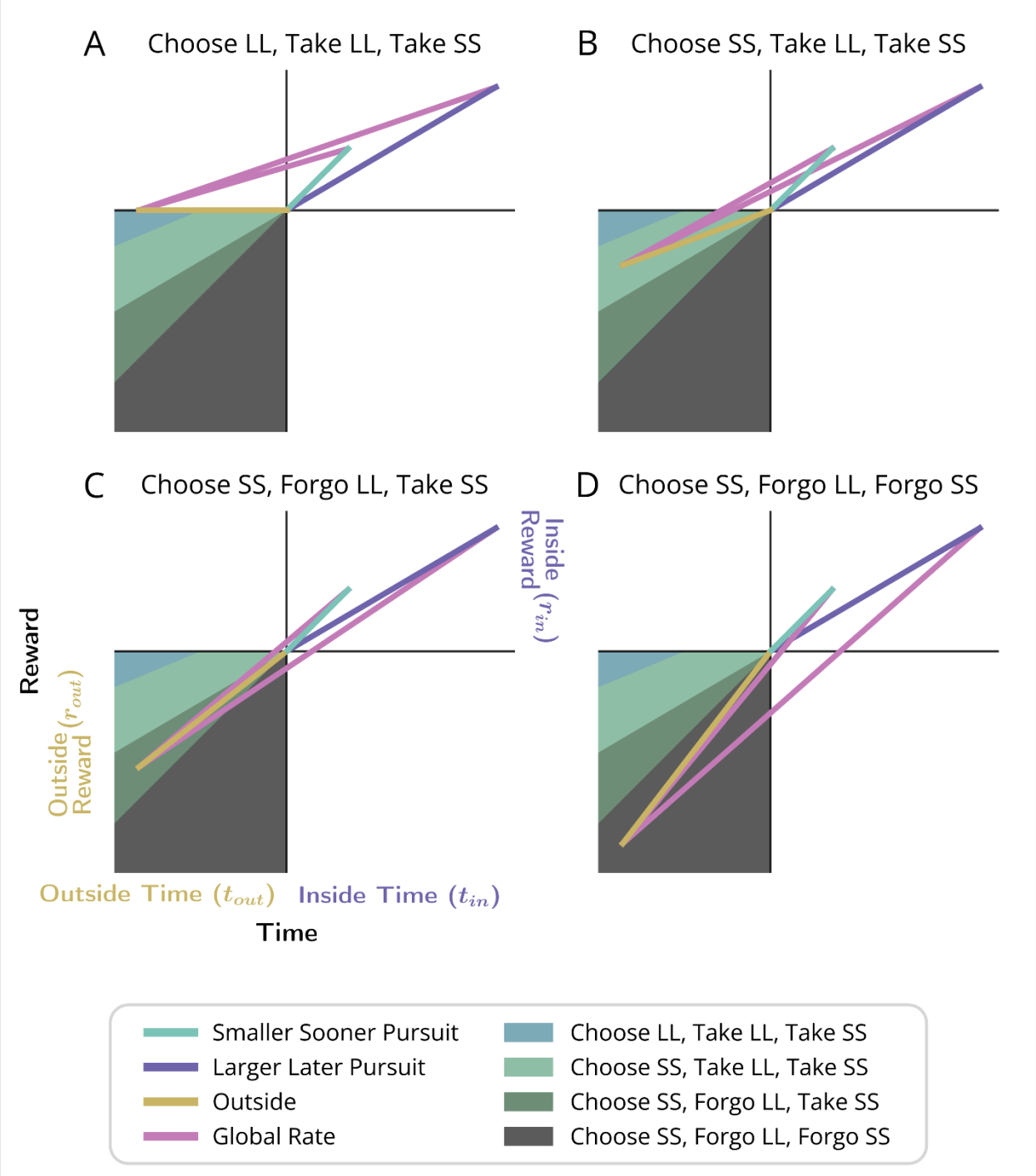


Figure 17. Relationship between outside time and reward with optimal temporal decision-making behavioral transitions. An agent may be presented with three decisions: the decision to take or forgo a smaller, sooner reward of 2.5 units after 2.5 seconds (SS pursuit), the decision to take or forgo a larger, later reward of 5 units after 8.5 seconds (LL pursuit), and the decision to choose between the SS and LL pursuits. The slope of the purple line indicates the global reward rate (ρ_g) resulting from a Choice or Take policy, while the slope of "outside" the pursuit (golden line) indicates the outside reward rate (i.e., global reward rate resulting from a Forgo policy). In each panel (A-D), an example outside reward rate is plotted, illustrating the relative ordering of ρ_g slopes for each policy. Location in the lower left quadrant is thereby shaded according to the combination of global rate-maximizing policies for each of the three decision types. r_{ss} = 2.5, t_{ss} = 2.5, r_{ll} = 5, t_{ll} = 8.5, t_{out} = 6, r_{out} = 0, 0.4, 0.8, 1.3 in A, B, C, & D, respectively.

With this understanding of the optimal thresholds between behavior policies, we can now examine the impact on decision-making behavior of different types of error in the agent's understanding of the world **(Figure 18).** We introduce an error term, ω, such that different parameters impacting the global reward rate of each considered policy are underestimated (ω<1) or overestimated (ω>1) (**Figure 18** column 1, see **Table 1** for formal definitions). Resulting global reward rate mis-estimations are equivalent to introducing error in the considered pursuit's subjective value, which will result in various deviations from reward-rate maximization **(Figure 18)**. Conditions wherein overestimation of global reward rate would lead to suboptimal choice behavior are identified formally in **Ap 11**.

Misestimated Variable	**Misestimated Global Reward Rate**
True (No Misestimation)	$\rho_g = \frac{r_{in} + \rho_{out} t_{out}}{t_{in} + t_{out}}$
Outside Time	$\hat{\rho}_g = \frac{r_{in} + \rho_{out} t_{out}}{t_{in} + \omega t_{out}}$
Outside Reward	$\hat{\rho}_g = \frac{r_{in} + \omega \rho_{out} t_{out}}{t_{in} + t_{out}}$
Outside Time and Reward (maintaining ρ_{out})	$\hat{\rho}_g = \frac{r_{in} + \omega \rho_{out} t_{out}}{t_{in} + \omega t_{out}}$
Inside Time	$\hat{\rho}_g = \frac{r_{in} + \rho_{out} t_{out}}{\omega t_{in} + t_{out}}$
Inside Reward	$\hat{\rho}_g = \frac{\omega r_{in} + \rho_{out} t_{out}}{t_{in} + t_{out}}$
Inside Reward and Time (maintaining ρ_{in})	$\hat{\rho}_g = \frac{\omega r_{in} + \rho_{out} t_{out}}{\omega t_{in} + t_{out}}$

Table 1. Definitions for misestimating global reward rate-enabling parameters. Each misestimated variable (column 1) is multiplied by an error term, ω, to give $\hat{\rho}_g$, the misestimated global reward rate (column 2). When $\omega = (0,1)$ the variable is underestimated, when $\omega = (1,2)$ the variable is overestimated, and when $\omega = 1$ the variable is correctly estimated and $\hat{\rho}_g = \rho_g$.

The sources of error considered are mis-estimations of the reward obtained and/or time spent "outside" (rows B-D) and "inside" (rows E-G) the considered pursuit. When both reward and time are misestimated, we examine the case in which the reward rate of that portion of the world is maintained (rows D & G). The agent's resulting policies in Choice (second column) and both Forgo situations (third and fourth columns) are determined across a range of outside reward rates (x-axes) and degrees of parameter misestimation (y-axes) and color-coded, with the boundary between the colored regions indicating the outside reward rate threshold for transitions in the agent's behavior. These individual policies are collapsed into the triplet of behavior expressed across the decision types (fifth column). In this way, characterization of the nature of suboptimality is aided by the use of the outside reward rate as the independent variable influencing decision-making, with the outside reward rate thresholds for optimal behavior being compared to the outside reward rate thresholds under any given parameter misestimation (comparing top "optimal" row A, against any subsequent row B-G). Any deviations in this pattern of behavior from that of the optimal agent (row A) are suboptimal, resulting in a failure to maximize reward rate in the environment.

While misestimation of any of these parameters will lead to suboptimal behavior, only specific sources and directions of error may result in behavior that qualitatively matches human and animal behavior observed experimentally. Misestimation of outside time (B), outside reward (C), inside time (E), and inside reward (F) all display Choice behavior that is qualitatively similar to experimentally observed behavior, either via underestimation or overestimation of the key variable. For example, underestimation of the outside time (B, $\omega<1$) leads to selection of the SS pursuit at sub-optimally low outside reward rates. However, agents with these types of error never display optimal Forgo behavior. By contrast, misestimation of either outside time and reward (D) or inside time and reward (G) display suboptimal Choice while maintaining optimal Forgo. Specifically, underestimation of outside time and reward (D, $\omega<1$) and overestimation of inside time and reward (G, $\omega>1$) both result in suboptimal preference for SS at low outside rates. Therefore, and critically, if the rates of both inside and outside are maintained despite misestimating reward and time magnitudes, the resulting errors allow for optimal Forgo behavior while displaying suboptimal "impatience" in Choice, and thus match experimentally observed behavior.

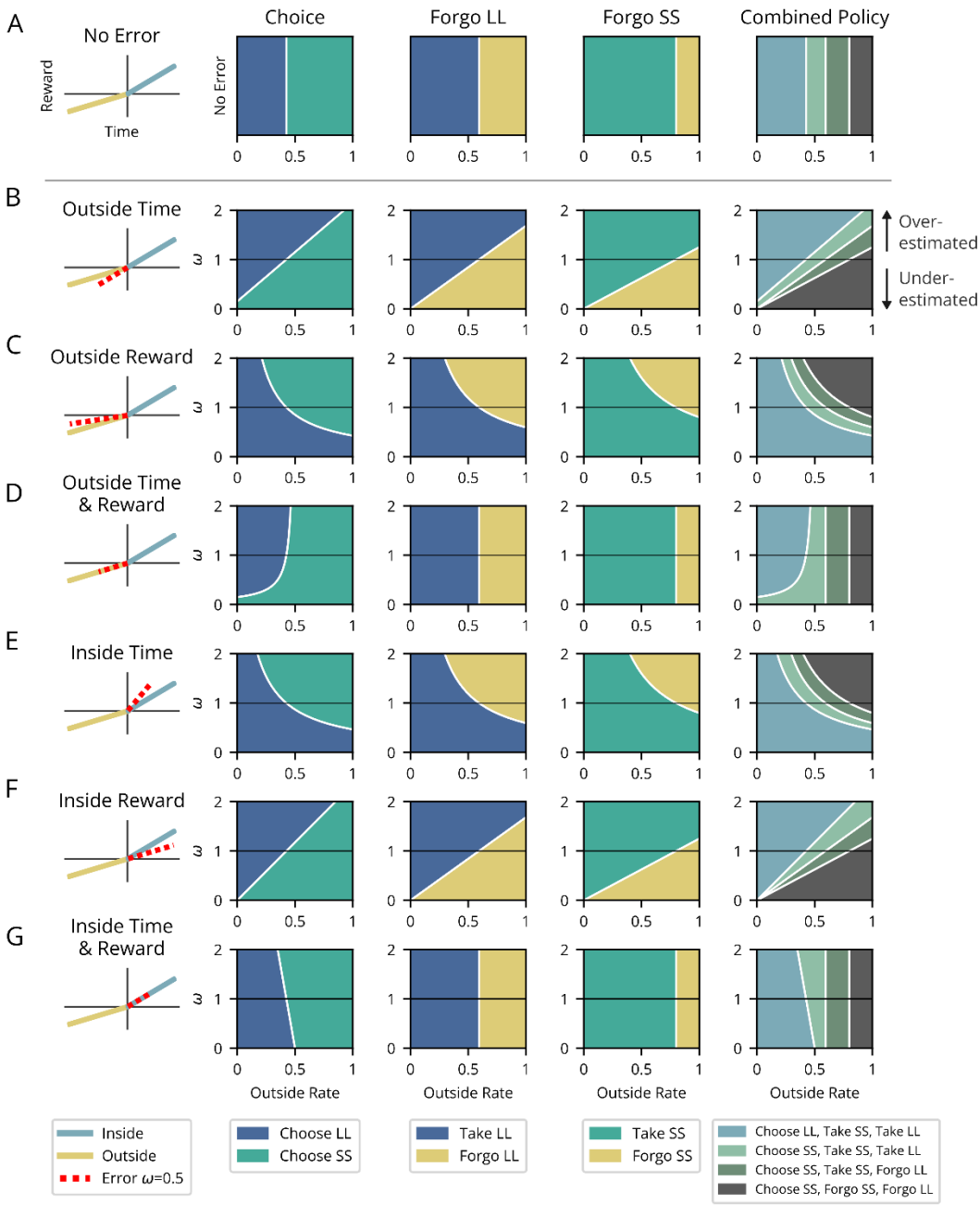


Figure 18. Patterns of suboptimal temporal decision-making behavior resulting from time and/or reward misestimation. Patterns of temporal decision-making in Choice and Forgo situations deviate from optimal (top row) under various parameter misestimations (subsequent rows). Characterization of the nature of suboptimality is aided by the use of the outside reward rate as the independent variable influencing decision-making (x-axis), plotted against the degree of error (y-axis) of a given parameter (ω<1 underestimation, ω=1 actual, ω>1 overestimation). The leftmost column provides a schematic exemplifying true outside (gold) and inside (blue) pursuit parameters and the nature of parameter error (dashed red) investigated per row (all showing an instance of underestimation). For each error case, the agent's resulting choice between SS and LL pursuits (2nd column), decision to take or forgo the LL pursuit (3rd column), and decision to take or forgo the SS pursuit (4th column) are indicated by the shaded color (legend, bottom of columns) for a range of outside rates and degrees of error. The rightmost column depicts the triplet of behavior observed, combined across tasks. Rows: **A)** "No error" - Optimal choice and forgo behavior. Vertical white lines show outside reward rate thresholds for optimal forgo behavior. **B-G)** Suboptimal behavior resulting from parameter misestimation. **B-D)** The impact of outside pursuit parameter misestimation. **B)** "Outside Time"- The impact of misestimating outside time (and thus

misestimating outside reward rate). **C)** "Outside Reward"- The impact of misestimating outside reward (and thus misestimating outside reward rate). **D)** "Outside Time & Reward"- The impact of misestimating outside time and reward, but maintaining outside reward rate. E-G) The impact of inside pursuit parameter misestimation. **E)** "Pursuit Time"- The impact of misestimating inside pursuit time (and thus misestimating inside pursuit reward rate. **F)** "Pursuit Reward" - The impact of misestimating the pursuit reward (and thus misestimating the pursuit reward rate). **G)** "Pursuit Time and Reward" - The impact of misestimating the pursuit reward and time, but maintaining the pursuit's reward rate. For this illustration, we determined the policies for a SS pursuit of 2 reward units after 2.5 seconds, a LL pursuit of 4.75 reward units after 8 seconds, and an outside time of 10 seconds. The qualitative shape of each region and resulting conclusions are general for all situations where the SS pursuit has a higher rate than the LL pursuit (and where a region exists where the optimal agent would choose LL at low outside rates).

Discussion

In order to understand why humans and animals factor time the way they do in temporal decision-making, our initial step has been to understand how a reward-rate-maximizing agent would evaluate the worth of initiating a pursuit within a temporal decision-making world. We did so in order to identify what are and are not signs of suboptimality and to gain insight into how animals' and humans' valuation of pursuits actually deviate from optimality. By analyzing fundamental temporal decisions, we identified equations enabling reward-rate maximization that evaluate the worth of initiating a pursuit. We first considered **Forgo** decisions to appreciate that a world can be parcellated into its constituent pursuits, revealing how pursuits' rates and relative occupancies (their 'weights'), along with the decision policy, determine the global reward rate. In doing so, we derived an expression for the worth of a pursuit in terms of the resulting global reward rate. From it, we re-expressed the pursuit's worth in terms of its global reward rate-equivalent immediate reward, i.e., its 'subjective value', reprising McNamara's foundational expression (McNamara, 1982). We then show that subjective value, rather than being calculated from the global reward rate under a policy of accepting the considered pursuit, can equally be calculated in terms of the outside reward rate (a policy of *not* accepting the considered pursuit type) and the proportion of time spent outside the pursuit. Expressing subjective value in terms of a pursuit's outside reward rate and time reveals that time's cost is constituted by an apportionment cost, as well as an opportunity cost. By then examining **Choice** decisions, we provide a deeper understanding of the nature of apparent temporal discounting in reward-rate-maximizing agents and establish that hyperbolic discounting, the Delay Effect, the Magnitude Effect, and the Sign Effect, are *not* signs of suboptimal decision-making, but rather are consistent with reward-rate maximization. While these purported signatures of suboptimality would in fact arise from reward-rate maximization, humans and animals are, nonetheless, suboptimal temporal decision makers, exhibiting apparent discounting functions that are too steep. By examining misestimation of the parameters that enable reward-rate maximization identified here, we implicate underestimation of the relative time spent outside versus in the considered pursuit type as the likely source of error committed by animals and humans in temporal decision-making that underlies their suboptimal pursuit valuation. We term this "The Malapportionment Hypothesis".

Temporal decision-making theories and frameworks

Two theories have predominated over the course of theorizing about how animals should invest time when pursuing rewards of a diversity of magnitudes and delays: a theory of exponential discounting (Samuelson, 1937; Frederick et al., 2002; Kalenscher and Pennartz, 2008) and a theory of optimal foraging (Charnov, 1976b; Pyke et al., 1977; Stephens and Krebs, 1986; Stephens, 2008). According to the former, exhibiting a permanent preference for one option over another through time was argued to be rational (Montague and Berns, 2002; Mazur, 2006; Nakahara and Kaveri, 2010), as in Discounted Utility Theory (DUT) (Samuelson, 1938). Discounting functions operating under this principle would then be exponential, with the best fit exponent controlling and embodying the agent's appreciation of the cost of time. In contrast, Optimal Foraging Theory (OFT) invoked reward-rate maximization as the normative principle. Referenced by a wide assortment of ethologists and ecologists (for review see (Pyke, 1984)), the specific formulation proponents of OFT generally use would result in an apparent discounting function that is hyperbolic. Indeed, in controlled laboratory experiments in which animals make decisions about how to spend time between rewarding options (Hariri et al., 2006; Hayden et al., 2011; Wikenheiser et al., 2013;

Blanchard and Hayden, 2014, 2015; Carter et al., 2015; Carter and Redish, 2016), experimental observations have demonstrated that hyperbolic functions are better fits to choice behavior in intertemporal choice tasks than exponential functions (Ainslie, 1975; Thaler and Shefrin, 1981; Frederick et al., 2002; Green and Myerson, 2004; Kim et al., 2008; Blanchard and Hayden, 2015). Nonetheless, and problematically for OFT, in most intertemporal choice tasks, animal behavior is far from optimal for maximizing reward rate (Reynolds and Schiffbauer, 2004; Hayden et al., 2011; Blanchard et al., 2013; Blanchard and Hayden, 2015).

Hyperbolic Temporal Discounting Functions

Indeed, with respect to global reward-rate maximization, animals and humans typically exhibit much too great a preference for smaller-sooner rewards (SS) in apparent discounting of delayed rewards (Chung and Herrnstein, 1967; Rachlin et al., 1972; Ainslie, 1974; Thaler, 1981; Ito and Asaki, 1982; Grossbard and Mazur, 1986; Mazur, 1988; Benzion et al., 1989; Loewenstein and Prelec, 1992; Green et al., 1994; Bateson and Kacelnik, 1996; Kacelnik and Bateson, 1996; Cardinal et al., 2001; Stephens and Anderson, 2001; Bennett, 2002; Frederick et al., 2002; Holt et al., 2003; Winstanley et al., 2004; Kalenscher et al., 2005; Roesch et al., 2007; Kobayashi and Schultz, 2008; Louie and Glimcher, 2010; Pearson et al., 2010). More precisely, we show here that what is meant by this suboptimal bias for SS is that the switch in preference from LL to SS occurs at an outside reward rate that is lower—and/or an outside time that is greater—than what an optimal agent would exhibit. It was Ainslie (Ainslie, 1975) who first understood that the empirically observed "preference reversals" between SS and LL pursuits could be explained if temporal discounting took on a hyperbolic form, which he initially conjectured to arise simply from the ratio of reward to delay (Grüne-Yanoff, 2015). This was problematic, however, on two fronts: 1) as the time nears zero, the value curve goes to infinity, and 2) there is no accommodation of differences observed within and between subjects regarding the steepness of discounting. Mazur (Mazur, 1987) addressed these issues by introducing 1 + k into the denominator, providing for the now standard hyperbolic discounting function, $discounting\ function = \frac{sv}{r} = \frac{1}{1+kt}$. Introduction of "1" solved the first issue, though "it never became fully clear how to interpret this 1" (Grüne-Yanoff, 2015 interviewing Ainslie). Introduction of the free-fit parameter, k, accommodated the variability observed across and within subjects by controlling the curvature of temporal discounting, and has become widely interpreted as a psychological trait, such as patience, or willingness to delay gratification (Frederick et al., 2002).

In this way, the Discounting Function framework has often been reified into a function possessed by the brain, an intrinsic property used to reduce, in a manner idiosyncratic to the agent, the value of delayed reward. Indeed, discounting functions have been directly incorporated into numerous models (Laibson, 1997; McClure et al., 2004; al-Nowaihi and Dhami, 2008; Killeen, 2009), motivating the search for its neurophysiological signature (Montague et al., 2006). In addition to accommodating intra- and inter-subject variability through the use of this free-fit parameter, discounting function formulations must also contend with the fact that best fits differ in steepness 1) when the time spent and 2) reward gained outside the pursuit changes (Lea, 1979; Stephens and Dunlap, 2009; Blanchard et al., 2013; Blanchard and Hayden, 2015; Carter et al., 2015; Smethells and Reilly, 2015; Carter and Redish, 2016), 3) as the delay to SS and LL pursuits increase (the Delay Effect), 4) when the reward magnitude of the pursuit changes (the Magnitude Effect), and 5) when considering the sign of the outcome of the pursuit (the Sign Effect). This sensitivity to conditions and variability across and within subjects has spurred a hunt for the 'perfect' discounting function (Namboodiri and Hussain Shuler, 2016) in an effort to better fit behavioral observations, resulting in formulations of increasing complexity (Laibson, 1997; McClure et al., 2004; al-Nowaihi and Dhami, 2008; Killeen, 2009). While such accommodations may provide for better fits of data, the uncertain origins of discounting functions (Hayden, 2016) pose a challenge to the utility of this framework in rationalizing observed behavior.

The apparent discounting function of global reward-rate-optimal agents exhibits purported signs of suboptimality

Of the array of temporal decision-making behaviors commonly observed and viewed through the lens of discounting, what might be better accounted for by a deeper understanding of how a reward-rate-optimal agent would evaluate the worth of initiating a pursuit? To address this, we derived expressions of reward-rate maximization, translated them into subjective value, and then re-expressed subjective value in terms of the apparent discounting function that would be exhibited by a reward-rate-maximizing agent. We demonstrate that a simple and intuitive equation subtracting time's cost is equivalent to a hyperbolic discounting equation that accounts for opportunity costs and the agent's apportionment of time in the environment. This analysis determines that the form and sensitivity to conditions that temporal discounting is experimentally observed to exhibit would actually be expressed by a reward-rate-maximizing agent. In doing so, we emphasize how discounting functions should be considered as descriptions of the result of a process, rather than being the process itself.

Regarding form, our analysis reveals that the apparent discounting function of a reward-rate-maximizing agent is a hyperbolic function…

$$\boldsymbol{Discounting\ Function} = \frac{\mathbf{1} - \boldsymbol{\rho_{out}} * \frac{\boldsymbol{t_{in}}}{\boldsymbol{r_{in}}}}{\mathbf{1} + \frac{\mathbf{1}}{\boldsymbol{t_{out}}}\boldsymbol{t_{in}}}$$

…which resembles the standard hyperbolic discounting function, $\frac{1}{1+kt}$, in the denominator, where $k = \frac{\mathbf{1}}{\boldsymbol{t_{out}}}$.

Whereas Mazur introduced $1 + k$ to t in the denominator to 1) force the function to behave as t approaches zero and 2) provide a means to accommodate differences observed within and between subjects, our derivation gives cause to the terms 1 and k, their relationship to one another, and to t in the denominator. First, from our derivation, "1" actually signifies taking t_{out} amount of time expressed in units of t_{out} ($t_{out}/t_{out} = 1$) and adding it to t_{in} amount of time expressed in units of t_{out} (ie, the total time to make a full pass through the world expressed in terms of how the agent apportions its time under a policy of accepting the considered pursuit). Absent from the numerator in the standard hyperbolic formulation, the solution for a reward-rate-maximizing agent gives rise to a term that accounts for the opportunity cost of time. Together, the diminishment of the value of a pursuit as its time investment increases is thus due to

time's cost—itself hyperbolic—which is shown to be composed of an apportionment (hyperbolic – linear) as well as an opportunity cost (linear) **(Figure 19 & Table 2).**

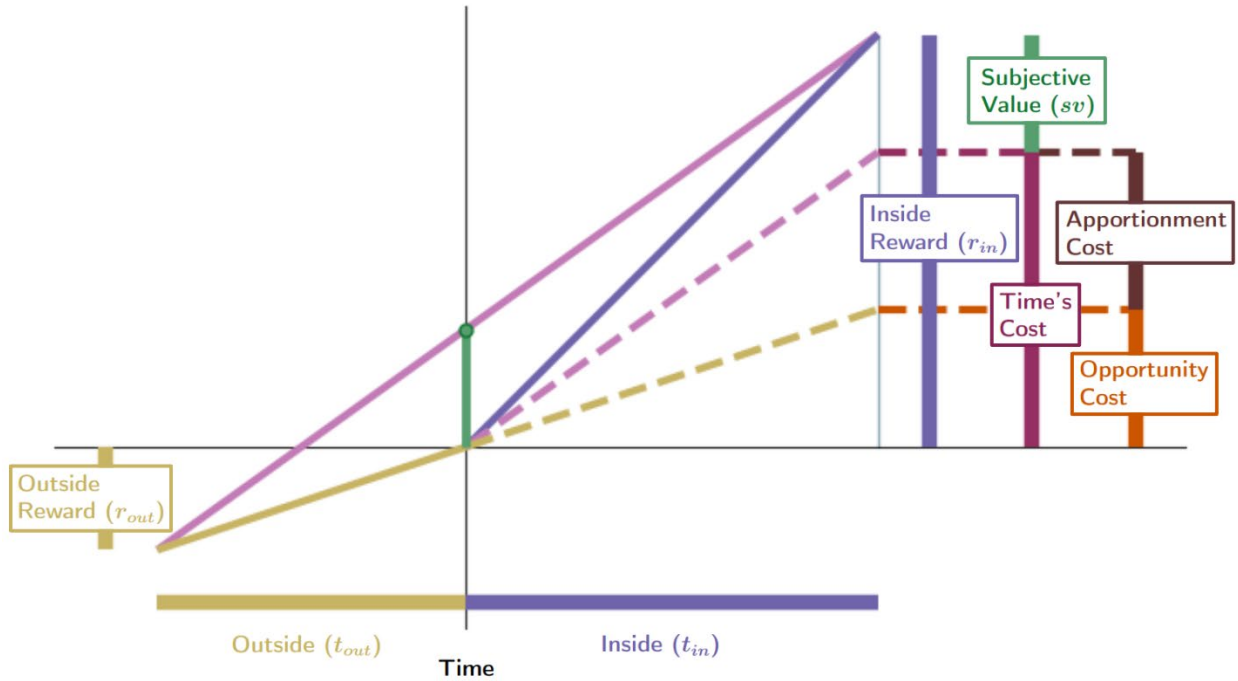


Figure 19. The cost of time of a pursuit comprises both an opportunity as well as an apportionment cost. The global reward rate under a policy of accepting the considered pursuit type (slope of magenta time), times the time that that pursuit takes (t_{in}), is the pursuit's time's cost (height of maroon bar). The subjective value of a pursuit (height of green bar) is its reward magnitude (height of the purple bar) less its cost of time. Opportunity and apportionment costs are shown to compose the cost of time of a pursuit. Opportunity cost associated with a considered pursuit, $\rho_{out}*t_{in}$, (height of orange bar) is the reward rate of the world under a policy of *not accepting* the considered pursuit (its outside rate), ρ_{out}, times the time of the considered pursuit, t_{in}. Therefore, opportunity cost is the amount of reward that would be expected from a policy of not taking the considered pursuit over a time equal to the considered pursuit's duration. The difference in reward that can be expected, on average, between a policy of taking versus a policy of not taking the considered pursuit, over a time equal to its duration, is the apportionment cost of time (height of brown bar). Together, they sum to time's cost.

In addition to demonstrating the form of the discounting function of an optimal agent, we can now also rationalize why it would appear to change in relationship to the features of the temporal decision-making world. **First**, rather than being a free-fit parameter like *k* in hyperbolic discounting models **(Figure 20A)**, the reciprocal of the time spent outside the considered pursuit type controls the degree of curvature in reward-rate optimizing agents **(Figure 20B,** denominator**)**. Therefore, changes in the apparent 'willingness' of a reward-rate-optimal agent to wait for reward would accompany any change in the amount of time that that agent needs to spend outside the considered pursuit, making the agent act as if more patient the greater the time spent outside a pursuit for every instance spent within it. Indeed, experiments with shorter intertrial intervals, and thus higher global reward rates, have a higher cost of time, and therefore exhibit steeper apparent temporal discounting (Lea, 1979; Stephens and Dunlap, 2009; Blanchard et al., 2013; Smethells and Reilly, 2015; Carter and Redish, 2016; Hayden, 2016).

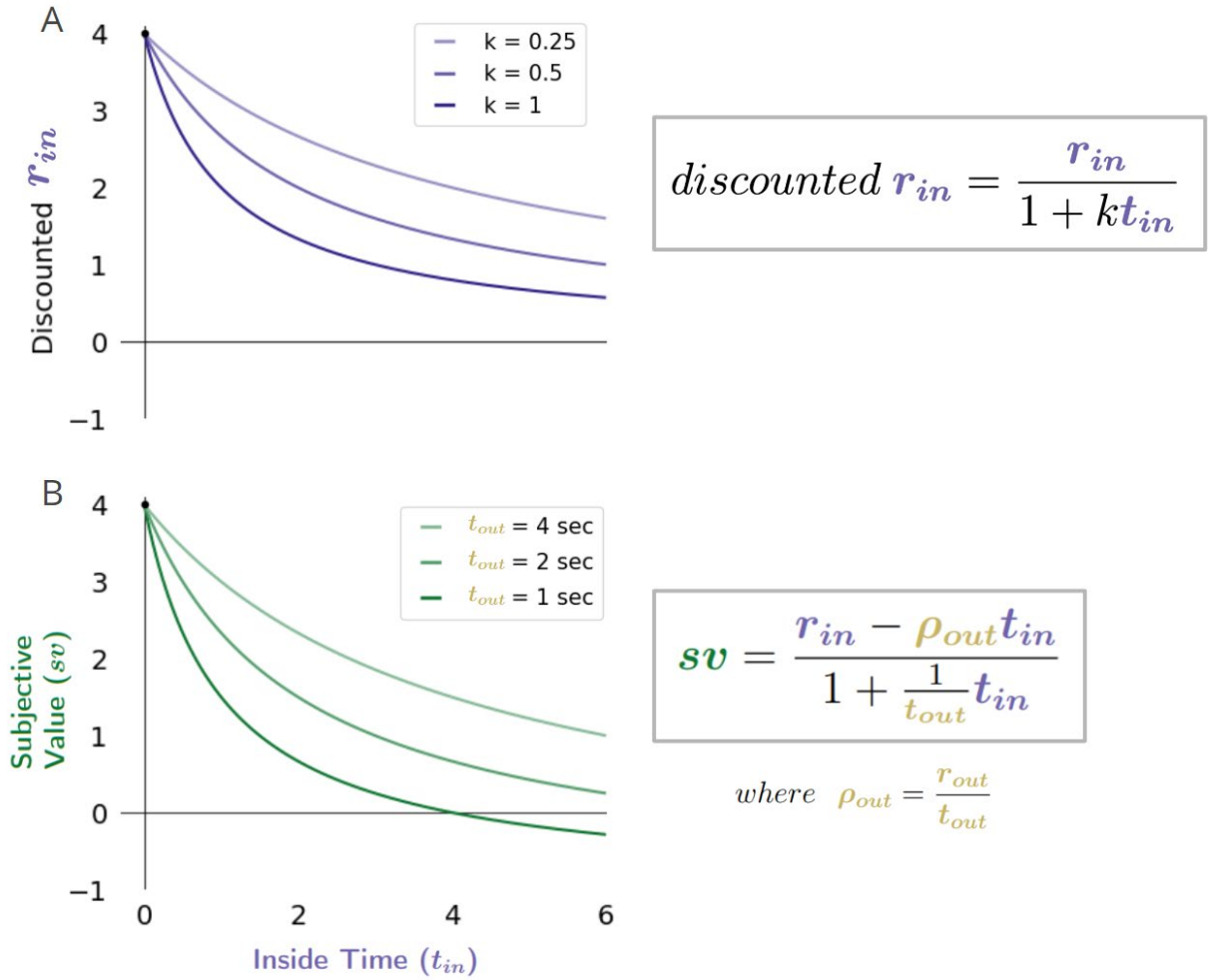


Figure 20. Comparison of typical hyperbolic discounting versus apparent discounting of a reward-rate-optimal agent. Whereas (A) the curvature of hyperbolic discounting models is typically controlled by the free fit parameter *k*, (B) the curvature and steepness of the apparent discounting function of a reward-rate-optimal agent is controlled by the time spent and reward rate obtained outside the considered pursuit. Understanding the shape of discounting models from the perspective of a reward-rate-optimal agent reveals that *k* ought relate to the apportionment of time spent in, versus outside, the considered pursuit, underscoring, how typical hyperbolic discounting models fail to account for the opportunity cost of time (and thus cannot yield negative *sv*'s no matter the temporal displacement of reward). Should *k* be understood as representing time's apportionment cost, the failure to account for the opportunity cost of time would lead to aberrantly high values of *k*.

Second, discounting frameworks must also rationalize why the steepness of apparent discounting changes as the reward rate acquired outside the considered pursuit changes. We show increasing the outside reward rate to be related to the linear opportunity cost of time in a reward rate maximizing agent **(Figure 13)**, subtraction of opportunity cost occurring in the numerator**)**. The greater the opportunity cost of time, the steeper the apparent discounting function and the less patient the agent would appear to be. In animal experiments in which a menu of options with different magnitudes and delays are presented as choice pairs, the outside reward rate will be greater than zero (Lea, 1979; Blanchard et al., 2013; Carter et al., 2015; Carter and Redish, 2016; Hayden, 2016). In these cases, hyperbolic fits should be steeper than would otherwise be the case if there were no outside reward rate. As typical hyperbolic discounting functions are of the form…

$$Discount\ Function\ = \frac{1}{1 + kt}$$

…they lack an accounting of the opportunity cost, and therefore must compensate for the lack of an opportunity cost term by overestimating k (which rightfully should only relate to the apportionment cost). Discounting models of the standard hyperbolic form are therefore only appropriate in worlds with no "outside" reward, as under such conditions opportunity cost is zero. In such worlds, the discounting function of a reward-rate-maximizing agent becomes…

$$Discount\ Function\ = \frac{1}{1 + \frac{1}{t_{out}} t_{in}}$$

… which fits a purely hyperbolic equation, where….

$$k\ = \frac{1}{t_{out}}$$

Relating to the treatment of opportunity cost, we also note that many investigations into temporal discounting often do not make an explicit distinction between situations in which 1) subjects continue to receive the usual rewards from the environment during the delay to a chosen pursuit and 2) situations in which during a chosen pursuit's delay no other rewards or opportunities will occur (Kable & Glimcher, 2007; Kirby & Maraković, 1996; McClure, Laibson, Loewenstein, & Cohen, 2004). Commonly, human subjects are asked to answer questions about their preferences between options for amounts they will not actually earn after delays they will not actually have to wait, during which it is unclear whether they are really investing time away from other options or not (Rosati et al., 2007). In contrast, in most animal experiments, subjects actually receive reward after different delays during which they *do not* receive new options or rewards. By our formulation, when being engaged in a pursuit does not exclude the agent from receiving rewards at the rate that occurs outside, the opportunity cost of time drops out of the subjective value equation **(Ap 12).**

$$\boldsymbol{sv = \frac{r_{in}}{1 + t_{in}/t_{out}}}$$

Equation 12. The value of initiating a pursuit when pursuit does not exclude receiving rewards at the outside rate (Ap 12)

Therefore, the reward-rate-maximizing discounting function in these worlds is functionally equivalent to the situation in which the outside reward rate is zero, and will—lacking an opportunity cost—be less steep. This rationalizes why human discounting functions are often reported to be longer than animal discounting functions: they are typically tested in conditions that negate opportunity cost, whereas animals are typically tested in conditions that enforce opportunity costs. Indeed, when humans are made to wait for actually received reward, their observed discounting functions are much steeper (Jimura et al., 2009).

Another consequence of opportunity cost as it relates to apparent temporal discounting functions is that the presence of opportunity cost means that the apparent temporal discounting function will become negative at sufficiently long delays, so that even rewarding pursuits, if not forced, should be forgone. The greater the opportunity cost of time, the steeper the apparent discounting function, and the earlier a rewarding pursuit will become that the agent ought forgo (when their acceptance would yield rates less than the outside rate, i.e., when $sv < 0$). The standard hyperbolic discounting function, lacking accounting of opportunity cost, cannot fit negative subjective values and therefore would not predict when—or even that—an agent will forgo a rewarding pursuit.

Third, fourth and fifth, discounting frameworks must make an accounting of the Delay Effect, Magnitude Effect, and Sign Effect, respectively, as they are considered important "anomalous" departures from microeconomic theory (Loewenstein and Thaler, 1989; Cruz Rambaud et al., 2023). To do so, rationalizations from previous works have invoked additional assumptions, such as separate processes for small and large rewards (Thaler, 1981), reference points (Kahneman and Tversky, 1979), or the inclusion of an elastic utility function (Loewenstein and Prelec, 1992; Killeen, 2009), among a myriad of other explanations (see (Frederick et al., 2002; Kalenscher and Pennartz, 2008) for reviews), including scalar timing (Gibbon, 1977) accounts of non-stationary time preference. While any of these proposals may indeed impact valuation, we emphasize here that qualitative features that they are invoked to explain are consistent with expectations of a reward-rate-maximizing agent.

Delay Effect. The inside/outside pursuit perspective makes clear that the 'Delay Effect', for instance, is not to be thought of as a sign of suboptimal decision making, but rather is what one should expect from a reward-rate-maximizing agent. As the resulting global reward rates drop due to an increasing delay to SS and LL pursuits, a switch will occur (see **Figure 14)** from a policy of choosing the SS pursuit to a policy of choosing the LL when the difference in the reward rates of these pursuits equals the global reward rates that would result from their selection. From the perspective of temporal discounting, this is understood to result from a hyperbolic discounting form (Ainslie, 1975; Kalenscher and Pennartz, 2008; Fawcett et al., 2012), shown here to be exhibited by a reward-rate-maximizing agent. Preference reversal, then, is not a sign of irrational decision making, as it would be exhibited by a reward-rate-maximizing agent. Rather, a sign of irrationality is that a preference reversal would occur at delays greater or less than what a reward-rate-maximizing agent would exhibit.

Magnitude Effect. We also demonstrate how the 'Magnitude Effect' would be a natural consequence of a process that would maximize reward rate, without invoking specialized processes or additional functions **(Figure 15)**. Further, from our reward-rate-maximizing framework, how the size of the Magnitude Effect would be affected by changing, experimentally, the outside time, reward, and rate parameters can be predicted. The Magnitude Effect should be observed, experimentally, to diminish when 1) increasing the outside time while holding the outside reward constant, (thus decreasing the outside reward rate), or when 2) decreasing the outside reward while holding the outside time constant (thus decreasing the outside reward rate). However, 3) the Magnitude Effect would exaggerate as the outside time increased while holding the outside reward rate constant.

Sign Effect. Whereas discounting frameworks need to invoke separate discounting functions to contend with different discounting rates for positive (rewarding) and negative (punishing) outcomes of the same magnitude (the Sign Effect), here too, we demonstrate how this is consistent with a reward-rate-maximizing process **(Figure 16)**. The asymmetry in the steepness of apparent discounting to rewards and punishments results from the average time and magnitude of rewards (or punishments) received outside the considered pursuit, forming a bias in evaluating equivalently sized outcomes of opposite sign. From the global reward-rate-maximizing perspective, we then also predict that the size of the Sign effect would diminish as the outside reward rate decreases (and as the outside time increases), and in fact would invert should the outside reward rate turn negative (become net punishing), such that punishments would appear to discount more steeply than rewards.

Collectively, our analysis of discounting functions reveals that features typically taken as signs of suboptimal/irrational decision-making are, in fact, consistent with reward-rate maximization. In this way, the general form and sensitivity to conditions of discounting functions, as observed experimentally, can be better understood from the perspective of a reward-rate-optimal agent **(Table 2),** providing a more parsimonious accounting of a confusing array of temporal decision-making behaviors reported.

	Reward		**Time**	
	Outside	**Inside**	**Outside**	**Inside**
Opportunity Cost[1]	Linear Positive slope	No Effect	Hyperbolic Negative slope	Linear Positive slope
Apportionment Cost	Linear Negative slope	Linear Positive slope	Hyperbolic - Hyperbolic [2] Negative slope	Hyperbolic - Linear [2] Negative slope

Time's Cost	Linear Positive slope	Linear Positive slope	Hyperbolic Negative slope	Hyperbolic Positive slope
Subjective Value	Linear Negative slope	Linear Positive slope	Hyperbolic Positive Slope	Hyperbolic Negative slope

Table 2. Opportunity cost, apportionment cost, time cost, and subjective value functions by change in outside and inside reward and time. Functions assume positive inside and outside rewards and times. [1]If outside reward rate is zero, opportunity cost becomes a constant at zero. [2]If outside reward rate is zero, as outside or inside time is varied, apportionment cost becomes purely hyperbolic.

Humans and animals are nonetheless suboptimal. What is the nature of this suboptimality?

These insights into the behavior of a reward-rate-maximizing agent inform on the meaning of the concept "patience". Patience oughtn't imply a willingness to wait a longer time, as it is not correct to say that an agent that chooses a pursuit requiring a long time investment is more patient that one that does not, for the amount of time a reward-rate-maximizing agent is willing to invest isn't an intrinsic property of the agent itself. Rather, it is a consequence of the temporal decision-making world's reward-time structure. So, if patience is to mean investing the 'correct' amount of time (i.e., the reward-rate-maximizing time), then a reward-rate-optimal agent doesn't *become* more or less patient as the context of what is otherwise the same pursuit changes; rather, it is *precisely* patient, under all circumstances. Impatience and over-patience then are terms to describe the behavior of a global reward-rate *suboptimal* agent that invests either too little, or too much time into a pursuit policy than one that would maximize global reward rate.

Having clarified what behaviors are and are not signs of suboptimality, actual differences to optimal performance exhibited by humans and animals can now be identified and quantified. So, what then are the decision-making behaviors of humans and animals when tasked with valuing the initiation of a pursuit, as in forgo and choice decisions? In controlled experimental situations, forgo decision-making is observed to be near optimal, consistent with observations from the field of behavioral ecology (Stephens and Dunlap, 2009; Smethells and Reilly, 2015; Hayden, 2016). In contrast, a suboptimal bias for smaller-sooner rewards is widely reported in Choice decision-making in situations where selection of later-larger rewards would maximize global reward rate (Logue et al., 1985; Blanchard and Hayden, 2015; Carter and Redish, 2016; Kane et al., 2019). Collectively, the pattern of temporal decision-making behavior observed under forgo and choice decisions shows that humans and animals act as if sub-optimally impatience under choice, while exhibiting near-optimal decision-making under forgo decisions.

The Malapportionment Hypothesis

How can animals and humans be sub-optimally impatient in choice, but optimal in forgo decisions? We postulated that previous behavioral findings of suboptimality can be understood from the perspective of overestimating the global reward rate. While misestimation of any variable underlying global reward rate calculation will lead to errors, not all misestimations will lead to errors that match the behavioral pattern of decisions observed experimentally. Having identified equations and their variables enabling reward-rate maximization, we sought to identify the likely source of error committed by animals and humans by analyzing the pattern of behavior consequent to misestimating one or another parameter. To do so, we identified the reward rate obtained outside a considered pursuit type as a useful variable to characterize departure from optimal decision-making behavior. Sweeping over a range of these values as the independent variable, we determined change points in decision-making behavior that would arise from misestimation (over- and under-estimations) of given reward-rate-maximizing parameters.

Our analysis shows how, precisely, misestimation of the inside and outside time or reward will lead to suboptimal temporal decision-making behavior. What errors, however, result in decisions that best accord with what is observed experimentally (i.e., result in suboptimal impatience in choice and optimal

forgo decision-making)? Overestimating outside time, underestimating outside reward, underestimating inside time, or overestimating inside reward would fail to match suboptimal 'impatience' in Choice *and* would result in suboptimal Forgo. Underestimating outside time, overestimating outside reward, overestimating inside time, or underestimating inside reward would match experimentally observed 'impatience' in Choice, but fail to match experimentally observed optimal Forgo behavior. To exhibit optimal forgo behavior, the inside and outside reward rates must be accurately appreciated. Therefore, misestimations of reward *and* time that preserve the true reward rates in and outside the pursuit would permit optimal forgo decisions while still misestimating the global reward rate. Overestimation of the outside time or underestimation of the inside time—while maintaining reward rates—fails to match experimentally observed 'impatience' in choice tasks while achieving optimal forgo decisions. However, underestimation of the outside time or overestimation of the inside time—while maintaining true inside and outside reward rates—*would* allow optimal forgo decision-making behavior while resulting in impatient choice behavior, as experimentally observed.

Previous experimental observations are consistent with, and have been interpreted as an agent underestimating the time spent outside the considered pursuit (Stephens and Dunlap, 2009; Blanchard et al., 2013; Smethells and Reilly, 2015), as would occur with underestimation of post-reward delays (Bateson and Kacelnik, 1996; Blanchard et al., 2013; Namboodiri et al., 2014a; Fung et al., 2021). Therefore, observed behavioral errors point to misestimating time apportionment in/outside the pursuit, either by 1) overestimating the occupancy of the considered choice or 2) underestimating the time spent outside the considered pursuit type, but not by 3) a misestimation of either the inside or outside reward rate. Only errors in time apportionment that underweight the outside time, (or, equivalently, overweight the inside time)—while maintaining the true inside and outside reward rates—will accord with experimentally observed temporal decision-making regarding whether to initiate a pursuit.

Thus, when a temporal decision world can effectively be bisected into two components, as often the case in experimental situations, only the reward rates, *but not the weights* of those portions need be accurately appreciated for the agent to optimally perform forgo decisions. Therefore, when tested in such situations, even agents that misestimate the apportionment of time can still make optimal forgo decisions based solely from a comparison of the reward rate in versus outside the pursuit. However, when faced with a choice between two or more pursuits when emerging from a path in common to any choice policy, optimal pursuit selection based on relative rate comparisons is no longer guaranteed, as *not only* the reward rates of pursuits, but *also their weights* must then be accurately appreciated. Misestimation of the weights of pursuits comprising a world would then result in errors in valuation regarding the initiation of a pursuit under choice instances.

We term this reckoning of the source of error committed by animals and humans the ***Malapportionment Hypothesis***, which identifies the underweighting of the time spent outside versus inside a considered pursuit *but not the misestimation of pursuit rates*, as the source of error committed by animals and humans **(Figure 21).** This hypothesis therefore captures previously published behavioral observations **(Figure 21A)** showing that animals can make decisions to take or forgo reward options that optimize reward accumulation (Krebs et al., 1977; Stephens and Krebs, 1986; Blanchard and Hayden, 2014), but make suboptimal decisions when presented with simultaneous and mutually exclusive choices between rewards of different delays (Logue et al., 1985; Blanchard and Hayden, 2015; Carter and Redish, 2016; Kane et al., 2019). The Malapportionment Hypothesis further predicts that apparent discounting functions will present with greater curvature than what a reward-rate-maximizing agent would exhibit **(Figure 21B)**. While experimentally observed temporal discounting would have greater curvature, the Malapportionment Hypothesis also predicts that the Magnitude **(Figure 21C)** and Sign Effect **(Figure 21D)** would be less pronounced than what a reward-rate-maximizing agent would exhibit, with these effects becoming less pronounced the greater the underweighting. Finally, with regards to the Delay Effect **(Figure 21E)**, the Malapportionment Hypothesis predicts that preference reversal would occur at delays greater than that exhibited by a reward-rate-maximizing agent, with the delay becoming more pronounced the greater the underweighting outside versus inside the considered pursuit by the agent.

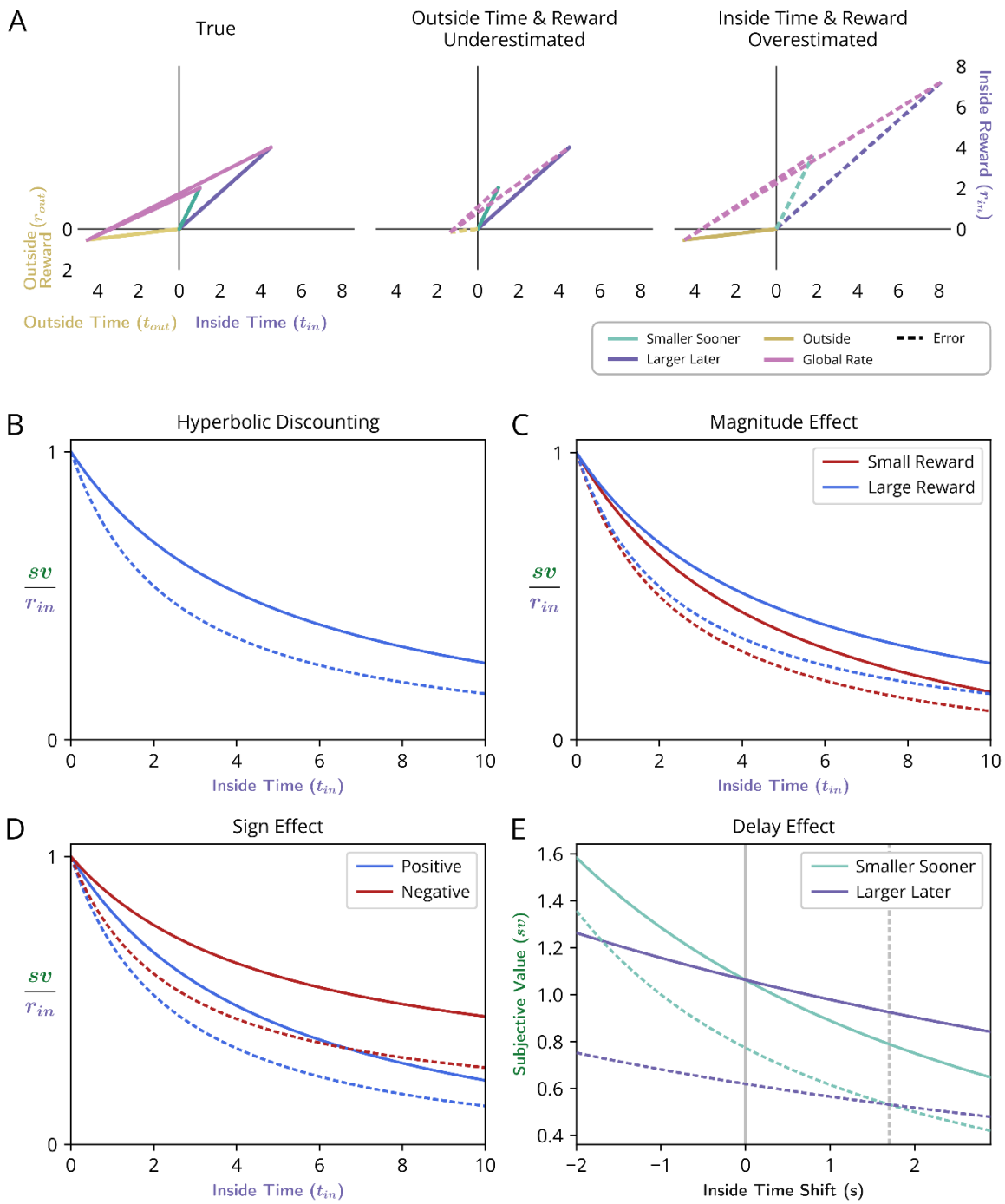


Figure 21. The Malapportionment Hypothesis. A-E Solid lines indicate true reward-rate maximizing values. Dashed lines indicate those of an agent described by the Malapportionment Hypothesis that underweights the apportionment of time outside relative to inside the considered pursuit type. The Malapportionment Hypothesis predicts the following: **A)** Suboptimal Choice, Optimal Forgo. Suboptimally "impatient" decision-making, as revealed under Choice decision-making, arises in humans and animals as a consequence of the valuation process underweighting the contribution of accurately assessed pursuit reward rates outside versus inside the considered pursuit type. **Top Left)** An example Choice situation where the global reward rate is maximized

by choosing a larger later reward over a smaller sooner reward. **Top Middle)** An agent that underweights the outside time but accurately appreciates the outside and inside reward rates, overestimates the global reward rate resulting from each policy, and thus exhibits suboptimal impatience by selecting the smaller sooner reward. $\omega = 0.3$. **Top Right)** Similarly, an agent that overweights the time inside the considered pursuit but accurately appreciates the outside and inside reward rates also overestimates the global reward rate and selects the smaller sooner reward. As inside and outside reward rates are accurately assessed, forgo decisions can correctly be made despite any misappreciation of the relative time spent in/outside the considered pursuit. $\omega = 1.8$. **B)** Hyperbolic discounting with greater curvature. The Malapportionment Hypothesis predicts that humans and animals exhibit temporal discounting with greater curvature than a reward-rate maximizing agent. $\omega = 0.5$, $r_{in} = 5$, $r_{out} = 0.5$, $t_{out} = 5$. **C)** Less pronounced Magnitude Effect**.** The Malapportionment Hypothesis predicts that the difference in apparent discounting related to the magnitude of the pursuit will be less pronounced than a reward-rate-maximizing agent. $\omega = 0.5$, $r_{out} = 0.5$, $t_{out} = 5$, $r_{SS} = 2$, $r_{LL} = 5$. **D)** Less pronounced Sign Effect**.** The Sign Effect is also predicted to be less pronounced than a reward-rate-maximizing agent. $\omega = 0.5$, $r_{out} = 0.5$, $t_{out} = 5$, positive reward = 3, negative reward (punishment) = -3. **E)** More pronounced Delay Effect. The Malapportionment Hypothesis predicts that the Delay Effect will be more pronounced than that exhibited by a reward-rate-maximizing agent. $\omega = 0.5$, $r_{out} = 0.5$, $t_{out} = 5$, $r_{SS} = 2$, $t_{SS} = 3$, $r_{LL} = 5$, $t_{LL} = 12.6$.

Comparisons to prior models

As our description of global reward rate-optimizing valuation is motivated by the same normative principle, how is our formalism unique from OFT, and, more generally, from other models proposing some form of reward-rate maximization? Firstly, the specific formulation proponents of OFT have used fails to adequately recognize how outside rewards influence the value of considered pursuits. Additionally, the relationship between time's cost and apparent temporal discounting has not been explicitly identified in prior OFT explanations. By contrast, our formulation, because of its specificity, can potentially align with neural representations of the variables we propose, and their misestimations may explain the ways in which observed animal behavior may deviate from optimality. Models inspired by OFT's objective of global reward-rate maximization but that seek to make a better accounting of observed deviations make the concession that, while global reward-rate maximization is sought, it is not achieved. Rather, some *non*-global reward-rate maximization is obtained by the agent (Bateson and Kacelnik, 1996; Blanchard et al., 2013; Namboodiri et al., 2014a; Fung et al., 2021). Of particular interest, the Heuristic model (Blanchard et al., 2013) and the TIMERR model (Namboodiri et al., 2014c) both assume non-global reward-rate maximization.

Heuristic Model

In the "Heuristic" model (Blanchard et al., 2013), as in Ecological Rationality Theory, ERT (Stephens et al., 2004), it is thought that animals prioritize the local reward rate of considered pursuits, rather than the global reward rate. In the Heuristic model, however, suboptimal "impatience" is rationalized as being the consequence of the animal's inability to fully appreciate post-reward delays (time subsequent to reward until re-entry into states/pursuits common to one or another policy). Indeed, while animals are demonstrated to be sensitive to post-reward delays, they act as if they significantly underestimate post-reward delays incurred, exhibiting a suboptimal bias for SS pursuits when LL pursuits would maximize global reward rate (Blanchard et al., 2013). Through a parameter, ω, which adjusts the degree in which post-reinforcer delays are underestimated, the Heuristic model can be sufficient to capture observed animal behavior in intertemporal choice tasks that have been assessed (Blanchard et al., 2013). However, as the Heuristic model is quite specific as to the source of error—the underestimation post-reward delays—it would well fit observed behavior only in certain experimental conditions. Should appreciable 1) reward be obtained or 2) time be spent outside of a considered pursuit type and its post-reward interval, then the Heuristic model would fail to make a good accounting of observed behavior.

The Heuristic model can be modified to specify the uniform downscaling of all *non*-pursuit intervals (rather than just post-reward delays), as in the implementation by Carter and Redish (Carter and Redish, 2016). This modification would bring the Heuristic model closer into alignment with the Malapportionment Hypothesis. But, as temporal underestimation would not apply to pursuits occurring outside the currently considered one, fits to observed behavior would be strained in worlds composed predominantly of pursuits with little non-pursuit time. Further, by underestimating the time spent outside

the considered pursuit without a corresponding underestimation of reward earned outside the considered pursuit, the Heuristic model ought to overestimate the outside reward rate and thus the global reward rate. So, while impatience under Choice could be fit under some experimental circumstances, behavior under Forgo instances would then be expected to also be sub-optimally impatient. Therefore, to bring the Heuristic model fully into alignment with the Malapportionment Hypothesis, it must be further assumed that the reward rate from the considered pursuit can be compared to the true outside or true global reward rate of the environment (as assumed in (Carter and Redish, 2016)), *as well as* expanded to underestimate *all* intervals of time occurring outside a considered pursuit type.

TIMERR Model

The essential feature of the TIMERR model (Namboodiri et al., 2014b) is that the agent looks back into its near past to estimate the reward rate of the environment, with this 'look-back' time, T_{ime}, being the model's free-fit parameter. In contrast to the reward-rate-optimal agent, this look-back time, then, is not a basic feature of the external world, but rather is related to how the animal uses its experience. TIMERR's policy is then determined by the reward rate obtained across this interval and that of the considered pursuit. In this way, TIMERR includes sources outside of the considered pursuit type in its evaluation, and because of this, exhibits many of the behaviors that the reward-rate-optimal agent is demonstrated here to express (Blanchard et al., 2013). Indeed, the TIMERR model and the optimal agent share the same mathematical form, though, critically, the meaning of their terms differs. An important additional difference is that IMERR is specific in the manner in which reward obtained outside the current instance of the considered pursuit is used: as recently experienced rewards from the past contribute to the estimation of the average reward rate of the environment, this 'look-back' time can include rewards from the pursuit type currently under consideration. Therefore, TIMERR commits an overestimation of the outside reward rate, and thus, an overestimation of global reward rate, manifesting as suboptimal impatience in choice *and* forgo decisions. In this way, while TIMERR is appealing in assuming that the recent past is used to estimate the global reward rate, and reproduces a number of sensitivities to conditions observed behaviorally, it is not in accordance with the Malapportionment Hypothesis as it mistakes pursuits' rates as well as their weights—something is still amiss.

Conclusion

An enriched understanding of how a reward-rate-optimal agent evaluates temporal decision-making empowers insight into the nature of human and animal valuation. It does so not by advancing the claim that we are optimal, but rather by clarifying what are and are not signs of optimality, which then permits quantification of the intriguing pattern of adherence and deviation from this normative expectation. Therein lies clues for deducing the learning algorithm and representational architecture used by brains to attribute value to representations of the temporal structure of the world. Here we have conceptualized and generalized temporal decision-making worlds as composed of pursuits, described by their rates and weights, and in so doing, come to better appreciate the cost of time, how policies impact the reward rates reaped from those worlds, and how processes that fail to accurately appreciate those features would misvalue the worth of initiating pursuits. We propose the Malapportionment Hypothesis, which identifies a failure to accurately appreciate the weights rather than the rates of pursuits, as the root cause of errors made, to reckon with the curious pattern of behavior observed regarding whether to initiate a pursuit. We postulate that the value learning algorithm and representational architecture selected for by evolution has favored the ability to appreciate the reward rates of pursuits over that of their weights.

Appendices

Ap 1. Derivation of equation for global reward rate under multiple pursuits

$E(r)$: the expected reward magnitude for each reward opportunity

$E(t)$: the expected time between the initiation of reward pursuits

$\rho_g = \frac{E(r)}{E(t)}$ global reward rate: the average reward per pursuit divided by the average time per pursuit.

ρ_d: the average rate of collecting rewards while in the default pursuit

p_i: reward opportunities i as a proportion of total pursued rewards

$$p_i = \frac{f_i}{\sum_{j=1}^{n} f_j}$$

$E(r_{pursuit}) = \sum_{i=1}^{n} p_i r_i$ the average reward received per reward opportunity

$$E(r_{pursuit}) = \sum_{i=1}^{n} \frac{f_i r_i}{\sum_{j=1}^{n} f_j} = \frac{\sum_{i=1}^{n} f_i r_i}{\sum_{i=1}^{n} f_i}$$

$E(t_{pursuit}) = \sum_{i=1}^{n} p_i t_i$ the average time invested per reward opportunity

$$E(t_{pursuit}) = \sum_{i=1}^{n} \frac{f_i t_i}{\sum_{j=1}^{n} f_j} = \frac{\sum_{i=1}^{n} f_i t_i}{\sum_{i=1}^{n} f_i}$$

$E(t_{default}) = \frac{1}{\sum_{i=1}^{n} f_i}$: the average time spent in the default pursuit between reward opportunities

$E(r_{default}) = \rho_d \frac{1}{\sum_{i=1}^{n} f_i}$: the average reward received in the default pursuit between reward opportunities

$\rho_g = \frac{E(r_{pursuit}) + E(r_{default})}{E(t_{pursuit}) + E(t_{default})}$ the global reward rate of the reward opportunity landscape

$$\rho_g = \frac{\frac{\sum_{i=1}^{n} f_i r_i}{\sum_{i=1}^{n} f_i} + \frac{\rho_d}{\sum_{i=1}^{n} f_i}}{\frac{\sum_{i=1}^{n} f_i t_i}{\sum_{i=1}^{n} f_i} + \frac{1}{\sum_{i=1}^{n} f_i}}$$

$$\boldsymbol{\rho_g = \frac{\sum_{i=1}^{n} f_i r_i + \rho_d}{\sum_{i=1}^{n} f_i t_i + 1}} \textbf{ (Equation 1 in Main test)}$$

Thus, **Equation 1** is formulated to calculate the average reward received and average time spent per unit time spent in the default pursuit. So, f_i is the encounter rate of pursuit i for one unit of time spent in the default pursuit. Added to the summation in the numerator, we have the average reward obtained in the default pursuit per unit time ($\rho_d * 1$) and in the denominator we have the time spent in the default pursuit per unit time (1).

Ap 2. Average time spent outside t_{out} the considered pursuit type, in, and the average reward rate earned outside that pursuit type, ρ_{out}

In order to simplify representations of policies governing any given pursuit opportunity, the expression for global reward rate, ρ_g, can be reformulated from the perspective of a policy of accepting any given pursuit.

$$\rho_{\forall i} = \frac{f_{in} r_{in} + \sum_{i \neq in} f_i r_i + \rho_d}{f_{in} t_{in} + \sum_{i \neq in} f_i t_i + 1}$$

$$\boldsymbol{\rho_g = \frac{r_{in} + (\sum_{\neq in} f_i r_i + \rho_d)/f_{in}}{t_{in} + (\sum_{i \neq in} f_i t_i + 1)/f_{in}}} \qquad \textbf{(Equation 2 in Main Text)}$$

$$t_{out} = t_{\forall i \neq in} = E(t_{invested, \forall i \neq in}) + E(t_{avail, \forall i \neq in})$$

$$\boldsymbol{t_{out} = t_{\forall i \neq in} = (\sum_{i \neq in} f_i t_i + 1)/f_{in}} \quad \textbf{(Equation 3 in Main Text)}$$

f_i is the probability that pursuit i will be encountered during a single unit of time spent in the default pursuit. The numerator of the expression is the average amount of time spent across all pursuits, excepting the considered pursuit, per unit time spent in the default pursuit. Note that the + 1 in the numerator is accounting for the unit of time spent in the default pursuit and is added outside of the sum. Since f_{in} is the probability that the considered pursuit will be encountered per unit of time spent in the default pursuit, $\frac{1}{f_{in}}$ is the average amount of time spent in the default pursuit between encounters of the considered pursuit. By multiplying the average time spent across all outside pursuits per unit of time in the default pursuit by the average amount of time spent in the default pursuit between encounters of the considered pursuit, we get the average amount of time spent outside the considered pursuit per encounter of the considered pursuit. This is calculated as if the pursuit encounters are mutually exclusive within a single unit of time spent within the default pursuit, as this is the case as the length of our unit time (delta t) approaches zero.

$$\rho_{out} = \rho_{\forall i \neq in} = \frac{\sum_{i \neq in} f_i r_i + \rho_d}{\sum_{i \neq in} f_i t_i + 1}$$

ρ_{out} is the reward rate achieved from all the time spent outside the considered pursuit, in, which is also the reward rate achieved if the considered pursuit, in, is never pursued.

Ap 3. Reformulation of global reward rate in terms of ρ_{out} and t_{out}

Parceling a pursuit world into "inside" the considered pursuit type and everything "outside" the considered pursuit type, gives a generalized form for the reward rate of an environment under a given policy.

$$\rho_{\forall i} = \frac{r_{in} + (\sum_{i \neq in} f_i r_i + \rho_d)/f_{in}}{t_{in} + (\sum_{i \neq in} f_i t_i + 1)/f_{in}}$$

$$\boldsymbol{r_{out} = (\sum_{i \neq in} f_i r_i + \rho_d)/f_{in}} \qquad \textbf{(Equation 4 in Main Text)}$$

$$t_{out} = (\sum_{i \neq in} f_i t_i + 1)/f_{in}$$

$$\boldsymbol{\rho_g = \rho_{\forall i} = \frac{r_{in} + r_{out}}{t_{in} + t_{out}}} \qquad \textbf{(Equation 5 in Main text)}$$

Ap 4. Global reward rate is a weighted average of an option's reward rate and its outside reward rate

$$\rho_g = \frac{r_{in} + \rho_{out} t_{out}}{t_{in} + t_{out}}$$

$$\rho_g = \frac{r_{in}}{t_{in}} \frac{t_{in}}{t_{in} + t_{out}} + \rho_{out} \frac{t_{out}}{t_{in} + t_{out}}$$

$$\text{Let } w = \frac{t_{in}}{t_{in} + t_{out}}$$

$$\boldsymbol{\rho_g = w \cdot \rho_{in} + (1 - w) \cdot \rho_{out}} \qquad \textbf{(Equation 6 in Main Text)}$$

Ap 5. Derivation of reward-rate-maximizing forgo policies

Forgo the considered pursuit in if $\rho_{\forall i} < \rho_{\forall i \neq in}$

$$\rho_{\forall i} < \rho_{\forall i \neq in}$$

$$\frac{r_{in}+(\sum_{i\neq in} f_i r_i+\rho_d)/f_{in}}{t_{in}+(\sum_{i\neq in} f_i t_i+1)/f_{in}} < \frac{(\sum_{i\neq in} f_i r_i+\rho_d)/f_{in}}{(\sum_{i\neq in} f_i t_i+1)/f_{in}}$$

$$\frac{r_{in}+ r_{out}}{t_{in}+ t_{out}} < \rho_{out}$$

$$r_{in} + r_{out} < \rho_{out} t_{in} + \rho_{out} t_{out}$$

$$r_{in} + r_{out} < \rho_{out} t_{in} + r_{out}$$

$$r_{in} < \rho_{out} t_{in}$$

$$\frac{r_{in}}{t_{in}} < \rho_{out}$$

$$\rho_{in} < \rho_{out}$$

$$\rho_{out} = \rho_g{}^*$$

$$\rho_{in} < \rho_g{}^*$$

$$\rho_{\forall i} < \rho_{\forall i\neq in} \leftrightarrow \rho_{in} < \rho_{out} \leftrightarrow \rho_{in} < \rho_g{}^*$$

Choose considered pursuit in if $\rho_{\forall i} > \rho_{\forall i\neq in}$

$$\rho_{\forall i} > \rho_{\forall i\neq in}$$

$$\frac{r_{in}+(\sum_{i\neq in} f_i r_i+\rho_d)/f_{in}}{t_{in}+(\sum_{i\neq in} f_i t_i+1)/f_{in}} > \frac{(\sum_{i\neq in} f_i r_i+\rho_d)/f_{in}}{(\sum_{i\neq in} f_i t_i+1)/f_{in}}$$

$$\frac{r_{in} + r_{out}}{t_{in} + t_{out}} > \rho_{out}$$

$$r_{in} + r_{out} > \rho_{out} t_{in} + \rho_{out} t_{out}$$

$$r_{in} + r_{out} > \rho_{out} t_{in} + r_{out}$$

$$r_{in} > \rho_{out} t_{in}$$

$$\frac{r_{in}}{t_{in}} > \rho_{out}$$

$$\rho_{in} > \rho_{out}$$

$$\frac{r_{in}}{t_{in}} \frac{t_{out}}{t_{in} + t_{out}} > \rho_{out} \frac{t_{out}}{t_{in} + t_{out}}$$

$$\frac{r_{in}}{t_{in}} \frac{t_{out}}{t_{in} + t_{out}} + \frac{r_{in}}{t_{in} + t_{out}} > \frac{r_{in} + \rho_{out}\, t_{out}}{t_{in} + t_{out}}$$

$$\frac{r_{in}}{t_{in}} \frac{t_{out}}{t_{in} + t_{out}} + \frac{r_{in}}{t_{in}} \frac{t_{in}}{t_{in} + t_{out}} > \frac{r_{in} + \rho_{out}\, t_{out}}{t_{in} + t_{out}}$$

$$\frac{r_{in}}{t_{in}} > \frac{r_{in} + \rho_{out}\, t_{out}}{t_{in} + t_{out}}, \rho_g{}^* = \frac{r_{in} + \rho_{out}\, t_{out}}{t_{in} + t_{out}}$$

$$\frac{r_k}{t_k} > \rho_g{}^*$$

$$\rho_{\forall i} > \rho_{\forall i \neq in} \leftrightarrow \rho_{in} > \rho_g{}^*$$

$$\rho_{\forall i} > \rho_{\forall i \neq in} \leftrightarrow \rho_{in} > \rho_{out} \leftrightarrow \rho_{in} > \rho_g{}^*$$

Choosing and forgoing the considered option in are equivalent if

$$\rho_{\forall i} = \rho_{\forall i \neq in}$$

$$\frac{r_{in} + (\sum_{i \neq in} f_i r_i + \rho_d)/f_{in}}{t_{in} + (\sum_{i \neq in} f_i t_i + 1)/f_{in}} = \frac{(\sum_{i \neq in} f_i r_i + \rho_d)/f_{in}}{(\sum_{i \neq in} f_i t_i + 1)/f_{in}}$$

$$\frac{r_{in} + r_{out}}{t_{in} + t_{out}} = \rho_{out}$$

$$r_{in} + r_{out} = \rho_{out} t_{in} + \rho_{out} t_{out}$$

$$r_{in} + r_{out} = \rho_{out} t_{in} + r_{out}$$

$$r_{in} = \rho_{out} t_{in}$$

$$\frac{r_{in}}{t_{in}} = \rho_{out}$$

$$\rho_{in} = \rho_{out}$$

$$\frac{r_{in}}{t_{in}} \frac{t_{out}}{t_{in} + t_{out}} = \rho_{out} \frac{t_{out}}{t_{in} + t_{out}}$$

$$\frac{r_{in}}{t_{in}}\frac{t_{out}}{t_{in}+t_{out}}+\frac{r_{in}}{t_{in}+\ t_{out}}=\frac{r_{in}+\ \rho_{out}t_{out}}{t_k+t_{out}}$$

$$\frac{r_{in}}{t_{in}}\frac{t_{out}}{t_{in}+\ t_{out}}+\frac{r_{in}}{t_{in}}\frac{t_{in}}{t_{in}+\ t_{out}}=\frac{r_{in}+\ \rho_{out}t_{out}}{t_{in}+\ t_{out}}$$

$$\frac{r_{in}}{t_{in}}=\frac{r_{in}+\rho_{out}\,t_{out}}{t_{in}+t_{out}},\ \rho_g{}^*=\frac{r_{in}+\rho_{out}\,t_{out}}{t_{in}+t_{out}}$$

$$\frac{r_{in}}{t_{in}}=\rho_g{}^*$$

$$\rho_{in}=\rho_g{}^*$$

$$\rho_{\forall i}=\rho_{\forall i\neq in}\leftrightarrow\rho_{in}=\rho_{out}\leftrightarrow\rho_{in}=\rho_g{}^*$$

Ap 6. Derivation of the equivalent immediate reward (i.e. the subjective value) for optimal global reward rate

Pursuit $in1$ and pursuit $in2$ produce the equivalent global reward rate if $\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}=\frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}=\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}}$

By definition, if $t_{in2}=0$, pursuit $in2$ is an immediate reward. Finding r_{in2} such that $\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}=\frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}=\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}}$ describes the equivalent immediate subjective value of pursuit $in1$.

If $\frac{r_{in2}+\rho_{out}t_{out}}{t_{out}}=\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}}$, $sv_{in1}=r_{in2}$

$$\frac{r_{in1}-sv_{in1}}{t_{in1}}=\frac{sv_{in1}+\rho_{ou}t_{out}}{t_{out}}=\frac{r_{in}+\rho_{out}t_{out}}{t_{in1}+t_{out}}=\rho_g$$

$$\frac{r_{in1}-sv_{in1}}{t_{in1}}=\rho_g$$

$$r_{in1}-sv_{in1}=\rho_g t_{in1}$$

$$sv_{in1}=r_{in1}-\rho_g t_{in1}$$

Therefore, for a considered pursuit, in,…

$$\boldsymbol{sv = r_{in} - \rho_g t_{in}}$$ **(Equation 7 in Main Text)**

Ap 7. Equivalent immediate subjective value need not be calculated from option-specific estimations of global reward rate

If $sv_{in1} < sv_{in2}$

$$sv_{in1} < sv_{in2}$$

$$r_{in1} - \rho_g(in1)t_{in1} < r_{in2} - \rho_g(in2)t_{in2}$$

$$\rho_g(in1) = \frac{r_{in1} + \rho_{out}t_{out}}{t_{in1} + t_{out}}$$

$$\rho_g(in1) < \rho_g(in2) = \rho_g{}^*$$

$$\rho_g(in1)t_{in1} < \rho_g(in2)t_{in1}$$

$$r_{in1} - \rho_g(in2)t_{in1} < r_{in1} - \rho_g(in1)t_{in1}$$

$$r_{in1} - \rho_g(in2)t_{in1} < r_{in1} - \rho_g(in1)t_{in1} < r_{in2} - \rho_g(in2)t_{in2}$$

$$r_{in1} - \rho_g{}^*t_{in1} < r_{in2} - \rho_g{}^*t_{in2}$$

$$sv_{in1}{}^* < sv_k{}^*$$

$$sv_{in1} < sv_{in2} \leftrightarrow sv_{in1}{}^* < sv_{in2}{}^*$$

If $sv_{in1} = sv_{in2}$

$$sv_{in1} = sv_{in2}$$

$$r_{in1} - \rho_g(in1)t_{in1} = r_{in2} - \rho_g(in2)t_{in2}$$

$$\rho_g(in1) = \frac{r_{in1} + \rho_{out}t_{out}}{t_{in1} + t_{out}}$$

$$\rho_g(in1) = \rho_g(in2) = \rho_g{}^*$$

$$r_{in1} - \rho_g(in2)t_{in1} = r_{in1} - \rho_g(in1)t_{in1}$$

$$r_{in1} - \rho_g{}^*t_{in1} = r_{in2} - \rho_g{}^*t_{in2}$$

$$sv_{in1}{}^* = sv_{in2}{}^*$$

$$sv_{in1} = sv_{in2} \leftrightarrow {sv_{in1}}^* = {sv_{in2}}^*$$

If $sv_{in1} > sv_{in2}$

$$sv_{in1} > sv_{in2}$$

$$r_{in1} - \rho_g(in1)t_{in1} > r_{in2} - \rho_g(in2)t_{in2}$$

$$\rho_g(in1) = \frac{r_{in1} + \rho_{out}t_{out}}{t_{in1} + t_{out}}$$

$${\rho_g}^* = \rho_g(in1) > \rho_g(in2)$$

$$\rho_g(in1)t_{in2} > \rho_g(in2)t_{in2}$$

$$r_{in2} - \rho_g(in2)t_{in2} > r_{in2} - \rho_g(in1)t_{in2}$$

$$r_{in1} - \rho_g(in1)t_{in1} > r_{in2} - \rho_g(in1)t_{in2}$$

$$r_{in1} - {\rho_g}^*t_{in1} > r_{in2} - {\rho_g}^*t_{in2}$$

$${sv_{in1}}^* > {sv_{in2}}^*$$

$$sv_{in1} > sv_{in2} \leftrightarrow {sv_{in1}}^* > {sv_{in2}}^*$$

If $\rho_{in} < \rho_{out}$

$$\rho_{in} < \rho_{out}$$

$${\rho_g}^* = \rho_{out}$$

$$\rho_{in} < {\rho_g}^* = \rho_{out}$$

$$\rho_{in} < {\rho_g}^*, \rho_{in} < \rho_{out}$$

$$r_{in} - {\rho_g}^*t_{in} < 0, r_{in} - \rho_{out}t_{in} < 0$$

$${sv_{in}}^* < 0, sv_{in} < 0$$

If $\rho_{in} = \rho_{out}$

$$\rho_{in} = \rho_{out}$$

$${\rho_g}^* = \rho_{in} = \rho_{out}$$

$$\rho_{in} = \rho_g{}^* = \rho_{out}$$

$$\rho_{in} = \rho_g{}^*, \rho_{in} = \rho_{out}$$

$$r_{in} - \rho_g{}^* t_{in} = 0, r_{in} - \rho_{out} t_{in} = 0$$

$$sv_{in}{}^* = 0, sv_{in} = 0$$

If $\rho_{in} > \rho_{out}$

$$\rho_{in} > \rho_{out}$$

$$\rho_g{}^* > \rho_{out}$$

$$\rho_{in} > \rho_g{}^* > \rho_{out}$$

$$\rho_{in} > \rho_g{}^*, \rho_{in} > \rho_{out}$$

$$r_{in} - \rho_g{}^* t_{in} > 0, r_{in} - \rho_{out} t_{in} > 0$$

$$sv_{in}{}^* > 0, sv_{in} > 0$$

Ap 8. Reformulation of equivalent immediate subjective value in terms of outside parameters

$$sv_{in} = r_{in} - \rho_g t_{in}$$

$$\rho_g = \frac{r_{in} + \frac{r_{out}}{t_{out}} t_{out}}{t_{in} + t_{out}}$$

$$sv_{in} = r_{in} - \frac{r_{in} + \frac{r_{out}}{t_{out}} t_{out}}{t_{in} + t_{out}} t_{in}$$

$$sv_{in} = r_{in} \frac{t_{in} + t_{out}}{t_{in} + t_{out}} - \frac{r_{in} t_{in} + (\frac{r_{out}}{t_{out}} t_{out}) t_{in}}{t_{in} + t_{out}}$$

$$sv_{in} = \frac{r_{in} t_{out} - \frac{r_{out}}{t_{out}} t_{in} t_{out}}{t_{in} + t_{out}}$$

$$\boldsymbol{sv_{in} = (r_{in} - \frac{r_{out}}{t_{out}} t_{in})(\frac{t_{out}}{t_{in}+t_{out}})} \quad \textbf{(Equation 8 in Main Text)}$$

The above expression of subjective value is arranged to emphasize an understanding of value in terms of an opportunity cost subtraction (the left parenthetical clause) and an apportionment scaling (the right parenthetical clause). Alternatively, the expression of subjective value below is re-arranged so that an opportunity cost subtraction appears in the numerator and where the time of the pursuit appears in the denominator with a scaling factor, $\frac{1}{t_{out}}$, resembling the standard temporal discounting form.

$$\boldsymbol{sv_{in} = \frac{r_{in} - \frac{r_{out}}{t_{out}} t_{in}}{1 + \frac{1}{t_{out}} * t_{in}}} \quad \textbf{(Equation 10, RHS, in Main Text)}$$

AP 9. Derivation of choice policies that optimize global reward rate

Let $t_{in1} > t_{in2}$

Choose option $in1$ if $\rho_{in1+\forall i \neq in1,in2} > \rho_{in2j+\forall i \neq in1,in2}$

$$\rho_{in1+\forall i \neq in1,in2} > \rho_{in2+\forall i \neq in1,in2}$$

$$\frac{r_{in1} + (\sum_{i \neq in1,in2} f_i r_i + \rho_d) / f_{in1,in2}}{t_{in} + (\sum_{i \neq in1,in2} f_i t_i + 1) / f_{in1}} > \frac{r_{in2} + (\sum_{i \neq in1,in2} f_i r_i + \rho_d) / f_{in1,in2}}{t_{in2} + (\sum_{i \neq in1,in2} f_i t_i + 1) / f_{in1,in2}}$$

$f_{in1,in2}$: the frequency at which the choice between option $in1$ and $in2$ are presented.

ρ_{out}: the reward rate earned outside of the $in1$ v. $in2$ choice

t_{out}: the average time per choice spent outside of $in1$ or $in2$.

$$\frac{r_{in1} + \rho_{out} t_{out}}{t_{in1} + t_{out}} > \frac{r_{in2} + \rho_{out} t_{out}}{t_{in2} + t_{out}}$$

$$(r_{in1} + \rho_{out} t_{out})(t_{in2} + t_{out}) > (r_{in2} + \rho_{out} t_{out})(t_{in1} + t_{out})$$

$$r_{in1} t_{in2} + r_{in1} t_{out} + \rho_{out} t_{out} t_{in2} + \rho_{out} {t_{out}}^2 > r_{in2} t_{in1} + r_{in2} t_{out} + \rho_{out} t_{out} t_{in1} + \rho_{out} {t_{out}}^2$$

$$r_{in1} t_{in2} + t_{out}(r_{in1} - r_{in2}) > r_{in2} t_{in1} + \rho_{out} t_{out}(t_{in1} - t_{in2})$$

$$(r_{in1} - r_{in2}) t_{in2} + t_{out}(r_{in1} - r_{in2}) > r_{in2}(t_{in1} - t_{in2}) + \rho_{out} t_{out}(t_{in1} - t_{in2})$$

$$(r_{in1} - r_{in2})(t_{in2} + t_{out}) > (r_{in2} + \rho_{out} t_{out})(t_{in1} - t_{in2})$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} > \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$$

$$\rho_{in1+\forall i\neq in1,in2} > \rho_{in2+\forall i\neq in1,in2} \leftrightarrow \frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} > \rho_{in2+\forall i\neq in1,in2}$$

relationship between $\rho_{in1+\forall i\neq in1,in2}, \rho_{in1+\forall i\neq in1,in2}$

$$\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} = \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}\frac{t_{in2}+t_{out}}{t_{in1}+t_{out}} + \frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}\frac{t_{in1}-t_{in2}}{t_{in1}+t_{out}}$$

$$\frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}\frac{t_{in2}+t_{out}}{t_{in1}+t_{out}} = \frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} - \frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}\frac{t_{in1}-t_{in2}}{t_{in1}+t_{out}}$$

$$\frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}} = \frac{t_{in1}+t_{out}}{t_{in2}+t_{out}}\left(\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} - \frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}\frac{t_{in1}-t_{in2}}{t_{in1}+t_{out}}\right)$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} > \frac{t_{in1}+t_{out}}{t_{in2j}+t_{out}}\left(\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} - \frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}\frac{t_{in1}-t_{in2}}{t_{in1}+t_{out}}\right)$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}\left(1+\frac{t_{in1}-t_{in2}}{t_{in2}+t_{out}}\right) > \frac{t_{in1}+t_{out}}{t_{in2}+t_{out}}\left(\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}}\right)$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}}\frac{t_{in1}+t_{out}}{t_{in2}+t_{out}} > \frac{t_{in1}+t_{out}}{t_{in2j}+t_{out}}\left(\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}}\right)$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} > \frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}}$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} > \frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} > \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$$

$\rho_g{}^*$: the maximum reward rate

If $\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} > \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$, $\rho_g{}^* = \frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}}$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} > \rho_g{}^*$$

Choose option $in2$ if $\rho_{in1+\forall i\neq in1,in2} < \rho_{in2+\forall i\neq in1,in2}$

If $\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} < \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$, $\rho_g{}^* = \frac{r_{in2}+\rho_{out}t_{out}}{t_{in1}+t_{out}}$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} < \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} < \rho_g{}^*$$

Option $in2$ and option $in1$ are equivalent if $\rho_{in1+\forall i\neq in1,in2} = \rho_{in2+\forall i\neq in1,in2}$

$$\text{If } \frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} = \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}, \rho_g{}^* = \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} = \frac{r_{in2}+\rho_{out}t_{out}}{t_{in}+t_{out}} = \frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} = \rho_g{}^*$$

$$\frac{r_{in1}-r_{in2}}{t_{in1}-t_{in2}} = \rho_g{}^*$$

Ap 10. Equivalent immediate subjective value policies that optimize global reward rate

Choose option $in1$ over pursuit $in2$ if $\rho_g(in1) > \rho_g(in2)$

$$\rho_g(in1) > \rho_g(in2)$$

$$\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} > \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$$

$$\frac{r_{in1}+\rho_{out}t_{out}}{t_{ink}+\ t_{out}} - \rho_{out} > \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}} - \rho_{out}$$

$$\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} - \rho_{out}\frac{t_{in1}+t_{out}}{t_{in1}+t_{out}} > \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}} - \rho_{out}\frac{t_{in2}+t_{out}}{t_{in2}+t_{out}}$$

$$\frac{r_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} - \frac{\rho_{out}t_{in1}+\rho_{out}t_{out}}{t_{in1}+t_{out}} > \frac{r_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}} - \frac{\rho_{out}t_{in2}+\rho_{out}t_{out}}{t_{in2}+t_{out}}$$

$$\frac{r_{in1}-\rho_{out}t_{in1}}{t_{in1}+t_{out}} > \frac{r_{in2}-\rho_{out}t_{in2}}{t_{in2}+t_{out}}$$

$$\frac{r_{in1}-\rho_{out}t_{in1}}{t_{in1}/t_{out}+1} > \frac{r_{in2}-\rho_{out}t_{in2}}{t_{in2}/t_{out}+1}$$

$$sv_{in1} > sv_{in2}$$

$$\boldsymbol{\rho_g(in1) > \rho_g(in2) \leftrightarrow sv_{in1} > sv_{in2}}$$

Choose option $in2$ over option $in1$ if $\rho_g(in2) > \rho_g(in1)$

$$\rho_g(in2) > \rho_g(in1)$$

$$\frac{r_{in2} + \rho_{out} t_{out}}{t_{in2} + t_{out}} > \frac{r_{in1} + \rho_{out} t_{out}}{t_{in1} + t_{out}}$$

$$\frac{r_{in2} - \rho_{out} t_{in2}}{t_{in2}/t_{out} + 1} > \frac{r_{in1} - \rho_{out} t_{in1}}{t_{in1}/t_{out} + 1}$$

$$\boldsymbol{\rho_g(in2) > \rho_g(in1) \leftrightarrow sv_{in2} > sv_{in1}}$$

Option $in2$ and option $in1$ are equivalent if $\rho_g(in1) = \rho_g(in2)$

$$\rho_g(in1) = \rho_g(in2)$$

$$\frac{r_{in1} + \rho_{out} t_{out}}{t_{in1} + t_{out}} = \frac{r_{in2} + \rho_{out} t_{out}}{t_{in2} + t_{out}}$$

$$\frac{r_{in1} - \rho_{out} t_{in1}}{t_{in1}/t_{out} + 1} = \frac{r_{in2} - \rho_{out} t_{in2}}{t_{in2}/t_{out} + 1}$$

$$sv_{in1} = sv_{in2}$$

$$\boldsymbol{\rho_g(in1) = \rho_g(in2) \leftrightarrow sv_{in1} = sv_{in2}}$$

Ap 11. Conditions wherein overestimation of global reward rate leads to suboptimal choice behavior

If $t_{LL} > t_{SS}$ and $\frac{r_{SS}}{t_{SS}} > \frac{r_{LL}}{t_{LL}}$ and $r_{LL} > r_{SS}$

$$\frac{r_{SS}}{t_{SS}} > \frac{r_{LL}}{t_{LL}}$$

$$r_{SS} t_{LL} > r_{LL} t_{SS}$$

$$r_{SS} t_{LL} - r_{SS} t_{SS} > r_{LL} t_{SS} - r_{SS} t_{SS}$$

$$r_{SS}(t_{LL} - t_{SS}) > (r_{LL} - r_{SS}) t_{SS}$$

$$\frac{r_{SS}}{t_{SS}} > \frac{r_{LL} - r_{SS}}{t_{LL} - t_{SS}}$$

$$r_{SS} t_{LL} > r_{LL} t_{SS}$$

$$r_{SS} t_{LL} + r_{LL} t_{LL} > r_{LL} t_{SS} + r_{LL} t_{LL}$$

$$r_{LL}t_{LL} - r_{LL}t_{SS} > r_{LL}t_{LL} - r_{SS}t_{LL}$$

$$r_{LL}(t_{LL} - t_{SS}) > (r_{LL} - r_{SS})t_{LL}$$

$$r_{LL}(t_{LL} - t_{SS}) > (r_{LL} - r_{SS})t_{LL}$$

$$\frac{r_{LL}}{t_{LL}} > \frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}}$$

$$\frac{r_{SS}}{t_{SS}} > \frac{r_{LL}}{t_{LL}} > \frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}}$$

$$\rho_{SS} > \rho_{LL} > \frac{r_{LL} - r_{SS}}{t_{LL} - t_{SS}}$$

$$\frac{r_{LL}}{t_{LL}} = \frac{r_{LL} - r_{SS}}{t_{LL} - t_{SS}}\frac{t_{LL} - t_{SS}}{t_{LL}} + \frac{r_{SS}}{t_{SS}}\frac{t_{SS}}{t_{LL}}$$

$$\rho_{LL} = \frac{r_{LL} - r_{SS}}{t_{LL} - t_{SS}}\frac{t_{LL} - t_{SS}}{t_{LL}} + \rho_{SS}\frac{t_{SS}}{t_{LL}}$$

$$\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} = \rho_{LL}\frac{t_{LL}}{t_{LL}-t_{SS}} - \rho_{SS}\frac{t_{SS}}{t_{LL}-t_{SS}}$$

pursuit LL is optimal if $\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} > {\rho_g}^*$ and $sv_{LL} > sv_{SS}$

Policy from global reward rate overestimation

$s\hat{v}_{LL} > s\hat{v}_{SS}$, the animal will choose pursuit LL

$$s\hat{v}_{LL} > s\hat{v}_{SS}$$

$$r_{LL} - \hat{\rho}_g t_{LL} > r_{SS} - \hat{\rho}_g t_{SS}$$

$$r_{LL} - r_{SS} > \hat{\rho}_g t_{LL} - \hat{\rho}_g t_{SS}$$

$$r_{LL} - r_{SS} > \hat{\rho}_g(t_{LL} - t_{SS})$$

$$\frac{r_{LL} - r_{SS}}{t_{LL} - t_{SS}} > \hat{\rho}_g$$

$s\hat{v}_{LL} < s\hat{v}_{SS}$, the animal will choose pursuit SS

$$s\hat{v}_{LL} < s\hat{v}_{SS}$$

$$r_{LL} - \hat{\rho}_g t_{LL} < r_{SS} - \hat{\rho}_g t_{SS}$$

$$r_{LL} - r_{SS} < \hat{\rho}_g t_{LL} - \hat{\rho}_g t_{SS}$$

$$r_{LL} - r_{SS} < \hat{\rho}_g(t_{LL} - t_{SS})$$

$$\frac{r_{LL} - r_{SS}}{t_{LL} - t_{SS}} < \hat{\rho}_g$$

pursuit LL is optimal if $\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} > \rho_g{}^*$ and pursuit LL is chosen if $\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} > \hat{\rho}_g$

pursuit LL is optimal if $\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} > \rho_g{}^*$ but pursuit SS is chosen if $\frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} < \hat{\rho}_g$

The policy from overestimation is suboptimal if $\rho_g{}^* < \frac{r_{LL}-r_{SS}}{t_{LL}-t_{SS}} < \hat{\rho}_g$

The policy from overestimation is suboptimal if $s\hat{v}_{LL} < s\hat{v}_{SS}$ but $sv_{LL} > sv_{SS}$

$$s\hat{v}_{LL} = sv_{LL} - t_{LL}(\hat{\rho}_g - \rho_g{}^*)$$

$$s\hat{v}_{LL} < s\hat{v}_{SS}$$

$$sv_{LL} - t_{LL}(\hat{\rho}_g - \rho_g{}^*) < sv_{SS} - t_{SS}(\hat{\rho}_g - \rho_g{}^*)$$

$$sv_{LL} - sv_{SS} < (t_{LL} - t_{SS})(\hat{\rho}_g - \rho_g{}^*)$$

$$sv_{LL} > sv_{SS} \rightarrow 0 < sv_{LL} - sv_{SS}$$

$$0 < sv_{LL} - sv_{SS} < (t_{LL} - t_{SS})(\hat{\rho}_g - \rho_g{}^*)$$

$$0 < \frac{sv_{LL} - sv_{SS}}{t_{LL} - t_{SS}} < \hat{\rho}_g - \rho_g{}^*$$

$$\hat{\rho}_g{}^* = max(\hat{w}_{LL}(\rho_{LL} - \rho_{out}), \hat{w}_{SS}(\rho_{SS} - \rho_{out})) + \rho_{out}$$

$$\hat{\rho}_g{}^* = \rho_g{}^* + (\hat{w}_{LL} - w_{LL})(\rho_{LL} - \rho_{out}) \text{ or } \hat{\rho}_g{}^* = \hat{\rho}_g{}^* + (\hat{w}_l - w_l)(l_l - \rho_{out})$$

$$\hat{\rho}_g{}^* - \rho_g{}^* = (\hat{w}_{LL} - w_{LL})(\rho_{LL} - \rho_{out}) \text{ or } \hat{\rho}_g{}^* - \rho_g{}^* = (\hat{w}_l - w_l)(l_l - \rho_{out})$$

$$0 < \frac{sv_{LL} - sv_{SS}}{t_{LL} - t_{SS}} < max((\hat{w}_{LL} - w_{LL})(l_{LL} - \rho_{out}), (\hat{w}_l - w_l)(l_l - \rho_{out}))$$

$$0 < \frac{sv_{LL} - sv_{SS}}{t_{LL} - t_{SS}} < (\hat{w}_l - w_l)(l_l - \rho_{out})$$

$$0 < \frac{sv_{LL} - sv_{SS}}{(l_l - \rho_{out})(t_{LL} - t_{SS})} < \hat{w}_l - w_l$$

$$w_l < w_l + \frac{sv_{LL}-sv_{SS}}{(l_l-\rho_{out})(t_{LL}-t_{SS})} < \hat{w}_l$$

Ap 12. Situations in which the rewarding option does not exclude the animal from receiving outside reward

$$\frac{sv + \rho_{out} t_{out}}{t_{out}} = \frac{r_{in} + \rho_{out}(t_{in} + t_{out})}{t_{in} + t_{out}}$$

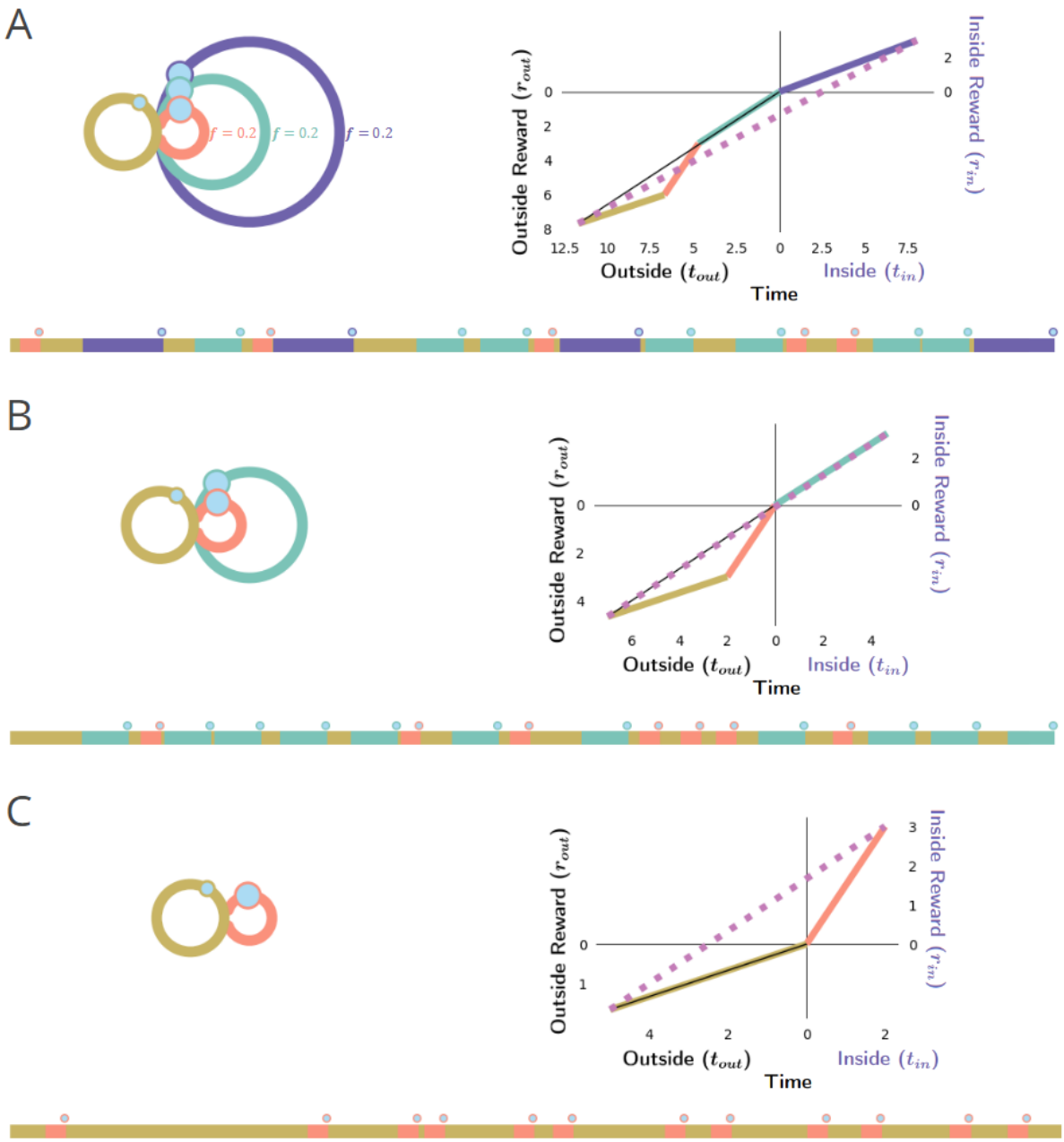
$$\frac{sv}{t_{out}} + \rho_{out} = \frac{r_{in}}{t_{in} + t_{out}} + \rho_{out}$$

$$\frac{sv}{t_{out}} = \frac{r_{in}}{t_{in} + t_{out}}$$

$$\boldsymbol{sv = \frac{r_{in}}{1 + t_{in}/t_{out}}}$$ **(Equation 12 in Main Text)**

Supplemental Figures

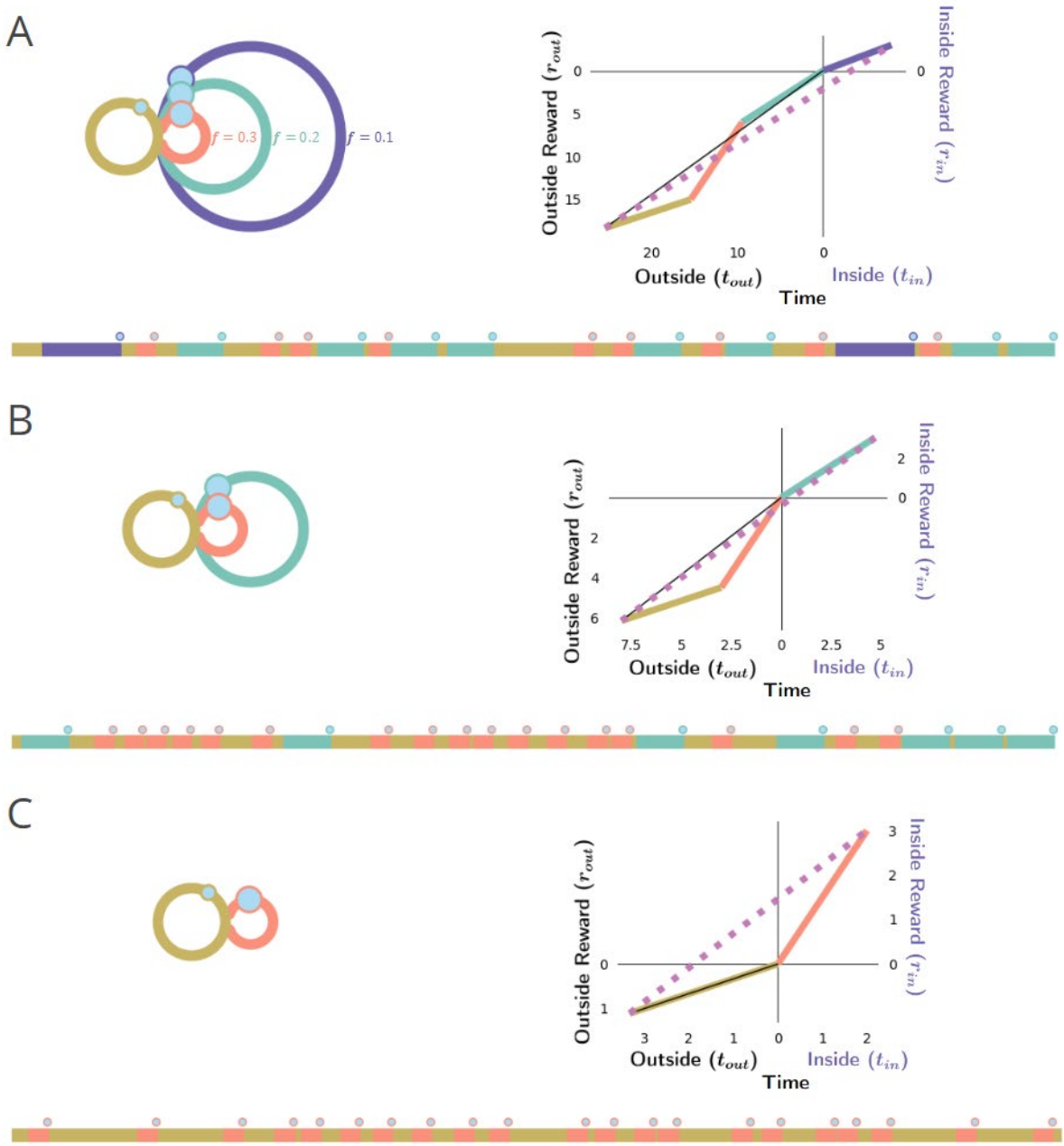
S1. Forgo world with n-pursuits occurring at the same frequency



Supplemental 1. Forgo world with n-pursuits occurring at the same frequency. An agent experiencing a forgo decision making world (topology as in A, left) where n-pursuits of varying times and rewards (purple, aqua, orange) occurring at the same frequency from the default pursuit (gold) are encountered, will forgo considered pursuits that evaluate to negative subjective values using, iteratively, the policy evaluation identified in Ap 7, ultimately producing the global reward-rate-optimal policy. **A)** *Left.* Topology of pursuit world. *Bottom.* Sample of the experienced world under a policy of accepting all offered pursuits. *Right*. The purple pursuit type, plotted with respect to it being the considered pursuit engaged 'in', with gold, orange, and cyan pursuits occurring 'outside' of purple, yields a global reward rate (dashed magenta line) that has a negative y-intercept, ie, a negative subjective value. Pursuits with a negative subjective value will have reward rates that are less than their 'outside' reward rate (slope of thin black line). Thus, when the inside reward rate is less than the outside reward rate, a reward rate-optimal agent will forgo

the considered pursuit, realizing a topology as in B. **B)** This agent, when encountering the aqua pursuit will subsequently forgo the aqua pursuit as well, excluding it from its policy as it would result in negative subjective value. **C)** Ultimately, the agent's policy converges to one of accepting only the orange pursuit, as a policy of its acceptance results in a positive subjective value. x-axis projections of these line segments represent the average time spent in one or another pursuit, whereas y-axis projections of these line segments represent the average reward acquired from one or another pursuit. Pursuit: reward, time. Purple: 3, 8; Aqua: 3, 4.7; Orange 3, 2; each occurring from the default pursuit at a frequency of 0.2 per unit time in the default pursuit.

S2. Forgo world with n-pursuits occurring at different frequencies



Supplemental 2. Forgo world with n-pursuits occurring at different frequencies. An agent experiencing a forgo decision making world (topology as in A, left) where n-pursuits of varying times and rewards (purple, aqua, orange) occurring at different frequency from the default pursuit (gold) are encountered, will forgo considered pursuits that evaluate to negative subjective values using, iteratively, the policy evaluation identified in Ap 7, ultimately producing the global reward-rate-optimal policy. **A)** *Left.* Topology of pursuit world. *Bottom.* Sample of the experienced world under a policy of accepting all offered pursuits. *Right*. The purple pursuit type, plotted with respect to it being the pursuit considered to engaged 'in', with gold, orange, and cyan, occurring 'outside' of purple, yields a global reward rate (dashed magenta line) that has a negative y-intercept, ie, a negative subjective value. Pursuits with a negative subjective value will have reward rates that are less than their 'outside' reward rate (slope of thin black line). Thus, when the inside reward rate is less than the outside reward rate, a reward rate-optimal agent will forgo the considered pursuit realizing a topology as in B. **B)** This agent, when encountering the aqua pursuit will subsequently forgo the aqua pursuit as well, excluding it from its policy as it would result in negative subjective value. **C)** Ultimately, the agent's policy converges to one

of accepting only the orange pursuit, as a policy of its acceptance results in a positive subjective value. x-axis projections of these line segments represent the average time spent in one or another pursuit, whereas y-axis projections of these line segments represent the average reward acquired from one or another pursuit. Pursuit: reward, time = Purple: 3, 8; Aqua: 3, 4.7; Orange 3, 2; occurring from the default pursuit at frequencies of 0.1, 0.2, and 0.3, respectively, per unit time in the default pursuit.

References

1. Ainslie G (1975) Specious reward: A behavioral theory of impulsiveness and impulse control. Psychol Bull 59:257–272.

2. Ainslie GW (1974) Impulse control in pigeons. J Exp Anal Behav 21:485–489 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16811760.

3. al-Nowaihi A, Dhami S (2008) A general theory of time discounting : The reference-time theory of intertemporal choice.

4. Baker F, Johnson MW, Bickel WK (2003) Delay discounting in current and never-before cigarette smokers: similarities and differences across commodity, sign, and magnitude. J Abnorm Psychol 112:382–392 Available at: http://dx.doi.org/10.1037/0021-843x.112.3.382.

5. Bateson M, Kacelnik A (1996) Rate currencies and the foraging starling: the fallacy of the averages revisited. Behav Ecol 7:341–352 Available at: http://dx.doi.org/10.1093/beheco/7.3.341.

6. Bennett SM (2002) Preference reversal and the estimation of indifference points using a fast-adjusting delay procedure with rats.

7. Benzion U, Rapoport A, Yagil J, Science, Source Management, Mar N (1989) Discount Rates Inferred from Decisions : An Experimental Study. Manage Sci 35:270–284.

8. Beran MJ, Evans TA (2009) Delay of gratification by chimpanzees (Pan troglodytes) in working and waiting situations. Behav Processes 80:177–181 Available at: http://dx.doi.org/10.1016/j.beproc.2008.11.008.

9. Berns GS, Laibson D, Loewenstein G (2007) Intertemporal choice--toward an integrative framework. Trends Cogn Sci 11:482–488 Available at: http://dx.doi.org/10.1016/j.tics.2007.08.011.

10. Bickel WK, Jarmolowicz DP, Mueller ET, Koffarnus MN, Gatchalian KM (2012) Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: emerging evidence. Pharmacol Ther 134:287–297 Available at: http://dx.doi.org/10.1016/j.pharmthera.2012.02.004.

11. Bickel WK, Miller ML, Yi R, Kowal BP, Diana M, Pitcock JA (2007) Behavioral and Neuroeconomics of Drug Addiction: Competing Neural Systems and Temporal Discounting Processes. Drug Alcohol Depend 90:S85–S91.

12. Blanchard TC, Hayden BY (2014) Neurons in dorsal anterior cingulate cortex signal postdecisional variables in a foraging task. J Neurosci 34:646–655 Available at: http://dx.doi.org/10.1523/JNEUROSCI.3151-13.2014.

13. Blanchard TC, Hayden BY (2015) Monkeys are more patient in a foraging task than in a standard intertemporal choice task. PLoS One 10:e0117057 Available at: http://dx.doi.org/10.1371/journal.pone.0117057.

14. Blanchard TC, Pearson JM, Hayden BY (2013) Postreward delays and systematic biases in measures of animal temporal discounting. Proc Natl Acad Sci U S A 110:15491–15496 Available at: http://dx.doi.org/10.1073/pnas.1310446110.

15. Bretteville-Jensen AL (1999) Addiction and discounting. J Health Econ 18:393–407 Available at: http://dx.doi.org/10.1016/s0167-6296(98)00057-5.

16. Calhoun AJ, Hayden BY (2015) The foraging brain. Current Opinion in Behavioral Sciences 5:24–31 Available at: https://www.sciencedirect.com/science/article/pii/S235215461500090X.

17. Calvert AL, Green L, Myerson J (2010) Delay discounting of qualitatively different reinforcers in rats. J Exp Anal Behav 93:171–184 Available at: http://dx.doi.org/10.1901/jeab.2010.93-171.

18. Cardinal RN, Pennicott DR, Sugathapala CL, Robbins TW, Everitt BJ (2001) Impulsive choice induced in rats by lesions of the nucleus accumbens core. Science 292:2499–2501 Available at: http://dx.doi.org/10.1126/science.1060818.

19. Carter EC, Pedersen EJ, McCullough ME (2015) Reassessing intertemporal choice: human decision-making is more optimal in a foraging task than in a self-control task. Front Psychol 6:95 Available at: http://dx.doi.org/10.3389/fpsyg.2015.00095.

20. Carter EC, Redish AD (2016) Rats value time differently on equivalent foraging and delay-discounting tasks. J Exp Psychol Gen 145:1093–1101 Available at: https://www.ncbi.nlm.nih.gov/pubmed/27359127.

21. Charnov E, Orians GH (1973) Optimal Foraging: Some Theoretical Explorations. Available at: https://digitalrepository.unm.edu/biol_fsp/45/?sequence [Accessed July 20, 2022].

22. Charnov EL (1976a) Optimal Foraging: Attack Strategy of a Mantid. Am Nat 110:141–151.

23. Charnov EL (1976b) Optimal Foraging, the Marginal Value Theorem. Theor Popul Biol 9:129–136 Available at: https://www.ncbi.nlm.nih.gov/pubmed/1273796.

24. Cheng K, Peña J, Porter MA, Irwin JD (2002) Self-control in honeybees. Psychon Bull Rev 9:259–263 Available at: http://dx.doi.org/10.3758/bf03196280.

25. Chung S-H, Herrnstein RJ (1967) CHOICE AND DELAY OF REINFORCEMENT. J Exp Anal Behav 10:67–74 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16811307.

26. Critchfield TS, Kollins SH (2001) Temporal discounting: basic research and the analysis of socially important behavior. J Appl Behav Anal 34:101–122 Available at: http://dx.doi.org/10.1901/jaba.2001.34-101.

27. Cruz Rambaud S, Ortiz Fernández P, Parra Oller IM (2023) A systematic review of the main anomalies in intertemporal choice. J Behav Exp Econ 104:101999 Available at: http://dx.doi.org/10.1016/j.socec.2023.101999 [Accessed July 11, 2024].

28. Estle SJ, Green L, Myerson J, Holt DD (2006) Differential effects of amount on temporal and probability discounting of gains and losses. Mem Cognit 34:914–928 Available at: http://dx.doi.org/10.3758/bf03193437.

29. Fawcett TW, McNamara JM, Houston AI (2012) When is it adaptive to be patient? A general framework for evaluating delayed rewards. Behav Processes 89:128–136 Available at: http://dx.doi.org/10.1016/j.beproc.2011.08.015.

30. Fedus W, Gelada C, Bengio Y, Bellemare MG, Larochelle H (2019) Hyperbolic Discounting and Learning over Multiple Horizons. arXiv [statML] Available at: http://arxiv.org/abs/1902.06865.

31. Frederick S, Loewenstein G, Donoghue TO, Donoghue TEDO (2002) Time Discounting and Time Preference : A Critical Review. J Econ Lit 40:351–401.

32. Fung BJ, Sutlief E, Hussain Shuler MG (2021) Dopamine and the interdependency of time perception and reward. Neurosci Biobehav Rev 125:380–391 Available at: http://dx.doi.org/10.1016/j.neubiorev.2021.02.030.

33. Gibbon J (1977) Scalar Expectancy Theory and Weber's Law in Animal Timing. Psychol Rev 84:279–325.

34. Glimcher PW, Kable J, Louie K (2007) Neuroeconomic Studies of Impulsivity: Now or Just as Soon as Possible? Am Econ Rev 97:142–147 Available at: https://pubs.aeaweb.org/doi/abs/10.1257/aer.97.2.142 [Accessed July 21, 2022].

35. Grace RC, Sargisson RJ, White KG (2012) Evidence for a magnitude effect in temporal discounting with pigeons. J Exp Psychol Anim Behav Process 38:102–108 Available at: http://dx.doi.org/10.1037/a0026345.

36. Green L, Fristoe N, Myerson J (1994) Temporal discounting and preference reversals in choice between delayed outcomes. Psychon Bull Rev 1:383–389.

37. Green L, Myerson J (2004) A discounting framework for choice with delayed and probabilistic rewards. Psychol Bull 130:769–792 Available at: http://dx.doi.org/10.1037/0033-2909.130.5.769.

38. Green L, Myerson J, McFadden E (1997) Rate of temporal discounting decreases with amount of reward. Mem Cognit 25:715–723.

39. Grossbard CL, Mazur JE (1986) A comparison of delays and ration requirements in self-control choice. J Exp Anal Behav 45:305–315.

40. Grüne-Yanoff T (2015) Models of temporal discounting 1937-2000: An interdisciplinary exchange between economics and psychology. Sci Context 28:675–713 Available at: https://people.kth.se/~gryne/papers/Hyperbol_150401.pdf.

41. Haith AM, Reppert TR, Shadmehr R (2012) Evidence for hyperbolic temporal discounting of reward in control of movements. J Neurosci 32:11727–11736 Available at: http://dx.doi.org/10.1523/JNEUROSCI.0424-12.2012.

42. Hariri AR, Brown SM, Williamson DE, Flory JD, de Wit H, Manuck SB (2006) Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. J Neurosci 26:13213–13217 Available at: http://dx.doi.org/10.1523/JNEUROSCI.3446-06.2006.

43. Hayden BY (2016) Time discounting and time preference in animals: A critical review. Psychon Bull Rev 23:39–53 Available at: http://dx.doi.org/10.3758/s13423-015-0879-3.

44. Hayden BY, Parikh PC, Deaner RO, Platt ML (2007) Economic principles motivating social attention in humans. Proc Biol Sci 274:1751–1756 Available at: http://dx.doi.org/10.1098/rspb.2007.0368.

45. Hayden BY, Pearson JM, Platt ML (2011) Neuronal basis of sequential foraging decisions in a patchy environment. Nat Neurosci 14:933–939 Available at: http://dx.doi.org/10.1038/nn.2856.

46. Hayden BY, Platt ML (2007) Temporal discounting predicts risk sensitivity in rhesus macaques. Curr Biol 17:49–53 Available at: http://dx.doi.org/10.1016/j.cub.2006.10.055.

47. Holt DD, Green L, Myerson J (2003) Is discounting impulsive? Behav Processes 64:355–367 Available at: http://dx.doi.org/10.1016/S0376-6357(03)00141-4.

48. Hwang J, Kim S, Lee D (2009) Temporal discounting and inter-temporal choice in rhesus monkeys. Front Behav Neurosci 3:9 Available at: http://dx.doi.org/10.3389/neuro.08.009.2009.

49. Ito M, Asaki K (1982) CHOICE BEHAVIOR OF RATS IN A CONCURRENT-CHAINS SCHEDULE: AMOUNT AND DELAY OF REINFORCEMENT. J Exp Anal Behav 37:383–392.

50. Jimura K, Myerson J, Hilgard J, Braver TS, Green L (2009) Are people really more patient than other animals? Evidence from human discounting of real liquid rewards. Psychon Bull Rev 16:1071–1075 Available at: http://dx.doi.org/10.3758/PBR.16.6.1071.

51. Kacelnik A, Bateson M (1996) Risky Theories—The Effects of Variance on Foraging Decisions. Integr Comp Biol 36:402–434 Available at: http://dx.doi.org/10.1093/icb/36.4.402.

52. Kahneman D, Tversky A (1979) Prospect Theory: An Analysis of Decision under Risk. Econometrica 47:263 Available at: http://dx.doi.org/10.2307/1914185.

53. Kalenscher T, Pennartz CMA (2008) Is a bird in the hand worth two in the future? The neuroeconomics of intertemporal decision-making. Prog Neurobiol 84:284–315 Available at: http://dx.doi.org/10.1016/j.pneurobio.2007.11.004.

54. Kalenscher T, Windmann S, Diekamp B, Rose J, Güntürkün O, 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 Available at: http://dx.doi.org/10.1016/j.cub.2005.02.052.

55. Kane GA, Bornstein AM, Shenhav A, Wilson RC, Daw ND, Cohen JD (2019) Rats exhibit similar biases in foraging and intertemporal choice tasks. Elife 8 Available at: http://dx.doi.org/10.7554/eLife.48429.

56. Killeen PR (2009) An additive-utility model of delay discounting. Psychol Rev 116:602–619 Available at: http://dx.doi.org/10.1037/a0016414.

57. Kim S, Hwang J, Lee D (2008) Prefrontal coding of temporally discounted values during intertemporal choice. Neuron 59:161–172 Available at: http://dx.doi.org/10.1016/j.neuron.2008.05.010.

58. Kinloch JM, White KG (2013) A concurrent-choice analysis of amount-dependent temporal discounting. Behav Processes 97:1–5 Available at: http://dx.doi.org/10.1016/j.beproc.2013.03.007.

59. Kirby KN, Herrnstein RJ (1995) Preference reversals due to myopic discounting of delayed reward. Psychol Sci 6:83–89 Available at: http://dx.doi.org/10.1111/j.1467-9280.1995.tb00311.x.

60. Kobayashi S, Schultz W (2008) Influence of reward delays on responses of dopamine neurons. J Neurosci 28:7837–7846 Available at: http://dx.doi.org/10.1523/JNEUROSCI.1600-08.2008.

61. Koopmans TC (1960) Stationary Ordinal Utility and Impatience. Econometrica 28:287–309.

62. Krebs BYJR, Erichsen JT, Webber MI (1977) OPTIMAL PREY SELECTION IN THE GREAT TIT (PARUS MAJOR). Anim Behav 25:30–38.

63. Laibson D (1997) Golden eggs and hyperbolic discounting. Q J Econ 112:443–477.

64. Lea SEG (1979) Foraging and reinforcement schedules in the pigeon: Optimal and non-optimal aspects of choice. Anim Behav 27:875–886 Available at: https://www.sciencedirect.com/science/article/pii/0003347279900253.

65. Loewenstein G, Thaler RH (1989) Anomalies: Intertemporal Choice. J Econ Perspect 3:181–193 Available at: https://www.aeaweb.org/articles?id=10.1257/jep.3.4.181 [Accessed March 13, 2024].

66. Loewenstein, Prelec (1992) Anomalies in intertemporal choice: Evidence and an interpretation. Q J Econ Available at: https://academic.oup.com/qje/article-abstract/107/2/573/1838331.

67. Logue AW, Smith ME, Rachlin H (1985) Sensitivity of pigeons to prereinforcer and postreinforcer delay. Anim Learn Behav 13:181–186 Available at: http://dx.doi.org/10.3758/bf03199271.

68. Louie K, Glimcher PW (2010) Separating value from choice: delay discounting activity in the lateral intraparietal area. J Neurosci 30:5498–5507 Available at: http://dx.doi.org/10.1523/JNEUROSCI.5742-09.2010.Separating.

69. Madden GF, Bickel WK eds. (2010) Impulsivity: The behavioral and neurological science of discounting.

70. Mazur JE (1987) An adjusting procedure for studying delayed reinforcement. In: The effect of delay and of intervening events on reinforcement value., pp 55–73 Quantitative analyses of behavior, Vol. 5. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

71. Mazur JE (1988) Estimation of indifference points with an adjusting-delay procedure. J Exp Anal Behav 49:37–47.

72. Mazur JE (2006) Mathematical models and the experimental analysis of behavior. J Exp Anal Behav 85:275–291 Available at: http://dx.doi.org/10.1901/jeab.2006.65-05.

73. Mazur JE, Snyderman M, Coe D (1985) Influences of delay and rate of reinforcement on discrete-trial choice. J Exp Psychol Anim Behav Process 11:565–575 Available at: http://dx.doi.org/10.1037//0097-7403.11.4.565.

74. McClure SM, Ericson KM, Laibson DI, Loewenstein G, Cohen JD (2007) Time discounting for primary rewards. J Neurosci 27:5796–5804 Available at: http://dx.doi.org/10.1523/JNEUROSCI.4246-06.2007.

75. McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. Science 306:503–507 Available at: http://dx.doi.org/10.1126/science.1100907.

76. McDiarmid CG, Rilling ME (1965) Reinforcement delay and reinforcement rate as determinants of schedule preference. Psychon Sci 2:195–196 Available at: https://doi.org/10.3758/BF03343402.

77. McNamara J (1982) Optimal patch use in a stochastic environment. Theor Popul Biol 21:269–288 Available at: https://www.sciencedirect.com/science/article/pii/0040580982900181.

78. Mischel W, Grusec J, Masters JC (1969) Effects of Expected Delay Time on Subjective Value of Rewards and Punishments. J Pers Soc Psychol 11:363–373.

79. Montague PR, Berns GS (2002) Neural economics and the biological substrates of valuation. Neuron 36:265–284 Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12383781.

80. Montague PR, King-Casas B, Cohen JD (2006) Imaging valuation models in human choice. Annu Rev Neurosci 29:417–448 Available at: http://dx.doi.org/10.1146/annurev.neuro.29.051605.112903.

81. Monterosso J, Ainslie G (1999) Beyond discounting: possible experimental models of impulse control. Psychopharmacology 146:339–347 Available at: http://dx.doi.org/10.1007/pl00005480.

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

83. Nakahara H, Kaveri S (2010) Internal-time temporal difference model for neural value-based decision making. Neural Comput 22:3062–3106 Available at: http://dx.doi.org/10.1162/NECO_a_00049.

84. Namboodiri VM, Hussain Shuler MG (2016) The hunt for the perfect discounting function and a reckoning of time perception. Curr Opin Neurobiol 40:135–141 Available at: http://dx.doi.org/10.1016/j.conb.2016.06.019.

85. Namboodiri VMK, Mihalas S, Hussain Shuler MG (2014a) A temporal basis for the origin of Weber's law in value perception. Front Integr Neurosci 8:1–11 Available at: http://dx.doi.org/10.3389/fnint.2014.00079.

86. Namboodiri VMK, Mihalas S, Hussain Shuler MG (2014b) Rationalizing decision-making: understanding the cost and perception of time. Timing and Time Perception Reviews 1:1–40 Available at: https://ugp.rug.nl/ttpr/article/view/15503.

87. Namboodiri VMK, Mihalas S, Marton T, Hussain Shuler MG (2014c) A general theory of intertemporal decision-making and the perception of time. Front Behav Neurosci 8:61 Available at: http://dx.doi.org/10.3389/fnbeh.2014.00061.

88. Niv Y (2009) Reinforcement learning in the brain. J Math Psychol 53:139–154 Available at: http://dx.doi.org/10.1016/j.jmp.2008.12.005.

89. Ostaszewski P (1996) The relation between temperament and rate of temporal discounting. Eur J Pers 10:161–172 Available at: https://journals.sagepub.com/doi/abs/10.1002/%28SICI%291099-0984%28199609%2910%3A3%3C161%3A%3AAID-PER259%3E3.0.CO%3B2-R.

90. Pearson JM, Hayden BY, Platt ML (2010) Explicit information reduces discounting behavior in monkeys. Front Psychol 1:237 Available at: http://dx.doi.org/10.3389/fpsyg.2010.00237.

91. Peters J, Büchel C (2011) The neural mechanisms of inter-temporal decision-making: understanding variability. Trends Cogn Sci 15:227–239 Available at: http://dx.doi.org/10.1016/j.tics.2011.03.002.

92. Pyke GH (1984) OPTIMAL FORAGING THEORY : A CRITICAL REVIEW. Annu Rev Ecol Syst 15:523–575.

93. Pyke GH, Pulliam HR, Charnov EL (1977) Optimal Foraging: A selective review of theory and tests. Q Rev Biol 52.

94. Rachlin H, Brown J, Cross D (2000) Discounting in judgments of delay and probability. J Behav Decis Mak 13:145–159 Available at: http://dx.doi.org/10.1002/(SICI)1099-0771(200004/06)13:2<145::AID-BDM320>3.0.CO;2-4.

95. Rachlin H, Green L, Vi AD (1972) Commitment, choice and self-control. J Exp Anal Behav 17:15–22.

96. Reynolds B, Schiffbauer R (2004) Measuring state changes in human delay discounting: an experiential discounting task. Behav Processes 67:343–356 Available at: http://dx.doi.org/10.1016/j.beproc.2004.06.003.

97. Richards JB, Mitchell SH, de Wit H, Seiden LS (1997) Determination of discount functions in rats with an adjusting-amount procedure. J Exp Anal Behav 67:353–366 Available at: http://dx.doi.org/10.1901/jeab.1997.67-353.

98. Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nat Neurosci 10:1615–1624 Available at: http://dx.doi.org/10.1038/nn2013.

99. Rosati AG, Stevens JR, Hare B, Hauser MD (2007) The evolutionary origins of human patience: temporal preferences in chimpanzees, bonobos, and human adults. Curr Biol 17:1663–1668 Available at: http://dx.doi.org/10.1016/j.cub.2007.08.033.

100. Samuelson PA (1937) A Note on Measurement of Utility. Rev Econ Stud 4:155–161 Available at: http://dx.doi.org/10.2307/2967612.

101. Samuelson PA (1938) A Note on the Pure Theory of Consumer's Behaviour. Economica 5:61–71 Available at: http://www.jstor.org/stable/2548836.

102. Schweighofer N, Shishida K, Han CE, Okamoto Y, Tanaka SC, Yamawaki S, Doya K (2006) Humans can adopt optimal discounting strategy under real-time constraints. PLoS Comput Biol 2:e152 Available at: http://dx.doi.org/10.1371/journal.pcbi.0020152.

103. Shuler M, Namboodiri V (2018) Think Tank: Forty Neuroscientists Explore the Biological Roots of Human Experience. In: Time's weird in the brain-that's a good thing, and here's why (Linden D, ed), pp 135–144. Yale University Press.

104. Smethells JR, Reilly MP (2015) Intertrial interval duration and impulsive choice. J Exp Anal Behav 103:153–165 Available at: http://dx.doi.org/10.1002/jeab.131.

105. Snyderman M (1983) Delay and amount of reward in a concurrent chain. J Exp Anal Behav 39:437–447 Available at: http://dx.doi.org/10.1901/jeab.1983.39-437.

106. Stephens DW (2008) Decision ecology: foraging and the ecology of animal decision making. Cogn Affect Behav Neurosci 8:475–484 Available at: http://dx.doi.org/10.3758/CABN.8.4.475.

107. Stephens DW, Anderson D (2001) The adaptive value of preference for immediacy : when shortsighted rules have farsighted consequences. Behav Ecol 12:330–339.

108. Stephens DW, Dunlap AS (2009) Why do animals make better choices in patch-leaving problems? Behav Processes 80:252–260 Available at: http://dx.doi.org/10.1016/j.beproc.2008.11.014.

109. Stephens DW, Kerr B, Fernández-Juricic E (2004) Impulsiveness without discounting: the ecological rationality hypothesis. Proc Biol Sci 271:2459–2465 Available at: http://dx.doi.org/10.1098/rspb.2004.2871.

110. Stephens DW, Krebs JR (1986) Foraging Theory.

111. Stevens JR, Mühlhoff N (2012) Intertemporal choice in lemurs. Behav Processes 89:121–127 Available at: http://dx.doi.org/10.1016/j.beproc.2011.10.002.

112. Story GW, Vlaev I, Seymour B, Darzi A, Dolan RJ (2014) Does temporal discounting explain unhealthy behavior? A systematic review and reinforcement learning perspective. Front Behav Neurosci 8:76 Available at: http://dx.doi.org/10.3389/fnbeh.2014.00076.

113. Strotz RH (1956) Myopia and Inconsistency in Dynamic Utility Maximization. Rev Econ Stud 23:165–180.

114. Takahashi T, Han R (2012) Tempospect theory of intertemporal choice. Psychology 3:555–557 Available at: http://dx.doi.org/10.4236/psych.2012.38082.

115. Thaler R (1981) Some empirical evidence on dynamic inconsistency. Econ Lett 8:201–207.

116. Thaler RH, Shefrin HM (1981) An Economic Theory of Self-Control. J Polit Econ 89:392–406 Available at: http://www.jstor.org/stable/1833317.

117. Wikenheiser AM, Stephens DW, Redish AD (2013) Subjective costs drive overly patient foraging strategies in rats on an intertemporal foraging task. Proc Natl Acad Sci U S A 110:8308–8313 Available at: https://www.ncbi.nlm.nih.gov/pubmed/23630289.

118. Winstanley CA, Theobald DEH, Cardinal RN, Robbins TW (2004) Contrasting roles of basolateral amygdala and orbitofrontal cortex in impulsive choice. J Neurosci 24:4718–4722 Available at: http://dx.doi.org/10.1523/JNEUROSCI.5606-03.2004.

119. Yi R, de la Piedad X, Bickel WK (2006) The combined effects of delay and probability in discounting. Behav Processes 73:149–155 Available at: http://dx.doi.org/10.1016/j.beproc.2006.05.001.