Reward and Time

On the Valuation of Delayed Rewards

Zarish Abbas

A Thesis

In the Department

of

Psychology

Presented in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy at

Concordia University

Montreal. Quebec, Canada

August 2021

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Zarish Abbas

Entitled: Reward and Time: On the Valuation of Delayed Rewards

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy Psychology

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Chair

Dr. Dylan Fraser

External Examiner

Dr. Meg Waraczynski

External to Program

Dr. Rachael Mansbach

Examiner

Dr. Peter Shizgal

Examiner

Dr. Dave Mumby

Thesis Supervisor (s)

Dr. Andreas Arvanitogiannis

Approved by

Dr. Aaron Johnson Chair of Department or Graduate Program Director

Date of Defence

Dr. Pascale Sicotte Dean, Faculty of Arts and Science

Abstract

Reward and Time: On the Valuation of Delayed Rewards

Zarish Abbas, Ph.D.

Concordia University, 2021

Time plays an essential role in reward valuation. While animals prefer large rewards to small rewards, the preference reverses if the large reward is delayed too long. This thesis is divided into four chapters that explore this curtailing effect of delay on reward valuation. Chapter 1 introduces delay discounting and neural models of interval timing to provide context for the questions that are asked in this thesis. In Chapter 2, we ask whether repeated treatment with the psychostimulant methylphenidate alters delay discounting in rats in the long run and whether any such effect is contingent on exposure at certain developmental stages. Compared with controls, rats treated with moderate doses of methylphenidate during early adolescence showed a long-term increase in delay discounting, indicating they had become impulsive. This brought up the question of what the drug was affecting: the magnitude of the reward, the perception of delay or how they combine. To gain insight, some of the more rudimentary assumptions about the combination of subjective reward and subjective delay needed to be confirmed. Chapter 3 addresses one of those assumptions: whether the computation of reward magnitude is independent from reward delay. We used intracranial self-stimulation and the reward-mountain model to tackle this question. The reward-mountain model describes how various dimensions of value combine. Through the model, we can computationally derive whether manipulations change the reward sensitivity by affecting the directly stimulated neurons or their integration, or whether they change the output of the integration. In contrast to the common assumption that relative subjective reward intensity and reward delay are computed independently and combined in a simple multiplicative manner, we show that delay affected reward processing at multiple stages during the computation of reward value. When reward was delayed by 2 s and 4 s, not only were later stages of the reward circuitry affected, but a reduction in reward sensitivity was also observed at early stages of processing, at the directly stimulated neurons or their integration. The final chapter discusses the implications of these results, including the possibility that the neural signal for reward magnitude and its delay may be interwoven.

Acknowledgements

This work was completed with scholarship support from Concordia University, from the Fonds de Recherche du Québec – Nature et Technologies (FRQNT) and from the laboratory of Dr. Arvanitogiannis. The research projects were financed by the Natural Sciences and Engineering Research Council (NSERC) and the Office of the Vice-President, Research and Graduate Studies (OVPRGS) Individual Seed Program. I am thankful that they gave these projects sustenance.

My supervisor, Andreas Arvanitogiannis, has been the greatest advocate a graduate student can have. I have evolved academically and personally because of his encouragement, contagious curiosity, and clarity and because he patiently gave me the time and space to do so. For all that and more, I am indebted to him. Peter Shizgal made room for my work in his laboratory and time with no contractual obligation to do so, and my work is much better for it. I am very grateful for his generosity. Thanks also to Dave Mumby for helpful discussions, and to my committee members for reading my work.

All my work was conducted using rats; it feels strange (and wrong) to thank them because this is not something they chose to do, but if credit is due to anyone, it is to them.

I thank Dave Munro for constructing the chambers for my experiment and for his patience through many requests for assistance during my somewhat slow technological learning curve. Thanks to Steve Cabilio for writing the software controlling those chambers. Rebecca Solomon helped me with the methodology of my experiment despite having graduated. I am grateful for her kindness. Arwen Sweet, Daniel Palacios, Ebtisam Afrand, Hajar El Mouddene, and Lisa Giaccari helped me run the experiments, I could not have done it on my own. Czarina Evangelista and Emily Cole helped me with surgeries and with my anxiety-driven requests for assurances during surgeries. Vasilis Pallikaras helped to run my experiment on days that I could not, and he took the emotional burden of my work by putting down the rats in my experiment. I am so very thankful for that. Vasilis, Emily, and Czarina along with Lindsay Sparks, Firas Sedki, Katuschia Germe and many of the other graduate students made long days in the laboratory more interesting, and navigating student politics a bit easier. The generous community of stack overflow, and those who have reached out to ask questions taught me how to code. Without them, I could not have conducted or analyzed this work.

My parents have, of course, made all things possible and I am thankful for that and for my sisters: Rajia, Anam and Sarah. I especially appreciate that they stopped asking me when I will graduate. Thanks to my family and friends for reading, and providing feedback on this document. Ashley MacDonald, Cecilia Piumbini, Gaby Girard, George Squire, Joanna Sciascia, Sahar Alamgir and Umber Zaidi gave their unconditional support through all these years whether in person or virtually. I am fortunate to have them! Thanks also to Ryan McKenna for our valuable conversations, and to him and James Brown (the cat), for consistently creating and maintaining spaces for me that encourage writing, thinking, and trying to figure it out – whatever it may be.

Table of Contents

List of Figures

List of Tables

Chapter 1: General Introduction

This thesis explores how time affects the computation of reward magnitude in the framework of neuroscience. The general introduction and discussion examine the union between those two concepts more broadly with some (limited) conjecture. It proposes that the current concept of reward magnitude as distinct from its temporal context may not be able to preserve autonomy and that instead, the two should be considered as some sort of unified entity.

For as long as animals have sought rewards, they have made assessments of time. While the first choanoflagellates may have aimlessly drifted around absorbing bacteria that were unlucky enough to cross their paths, about 550 million years ago animals changed the course of life by developing a purpose for their movements. This began the fortuitous symbiosis between the brain and the environment. Animals could be guided by their environmental conditions, as they evolved to sense them, to make choices and decisions about their movements. Food promoted approach and consummatory behaviour and could therefore be considered a *reward* for the animal and a *reinforcer* of the animals' decisions and actions. Deciding which reward to seek, or whether to seek one at all, depended on the relative values associated with the actions that were available for the animal to take. The computation of time was of the essence in these vital new decisions; how long the animal would have to wait for the reward, how long it would take to get to the reward, and whether a competing animal could get to it. As brains and the reinforcers guiding behaviour have become more complex, so have decisions – cooking timpano, investing in stocks, playing chess, pausing to deliver a punchline – and, still, the computation of time remains of the essence in these most complex decisions.

Even in a highly controlled laboratory setting, the influence of time on behaviour and decision-making is difficult to eradicate. The problem is, of course, not exclusive to seeking rewards; time is interlinked with all that we perceive and do. It is the mandatory backdrop across which actions and events unfold. No one has noticed something other than at a time or a time other than with something. It is disconcerting then that over a century into extensive research on reward, the expense of countless animal lives, and the collective efforts of some of the most brilliant intellects, we do not yet know how the brain combines reward and time, or even how the brain computes time. Still, it has not all been in vain – careful experimental procedures that allowed us to map the behaviour and decisions of animals have revealed quantitative patterns of behaviour that remain preserved across species. These patterns, combined with seminal research on the neural mechanisms that may underlie interval timing, has now provided us with fertile ground for directing our questions into how the brain might compute and integrate these two crucial parameters of reward and time.

1.1. Delay Discounting

Inquiry into the pattern with which animals integrate durations and reward arose from observing that animals would choose a smaller reward over a larger one if the opportunity to obtain it came sooner (Ainslie, 1975). Delaying the delivery of a reward seemed to discount its subjective value in the language of economists, a phenomenon that has been termed *delay discounting* (Critchfield & Kollins, 2001; Mazur, 1987). Devaluation of future rewards has been observed in all animals that have been tested (Vanderveldt et al., 2016). Rats, pigeons, monkeys, and humans not only discount future rewards, but seem to do so in a similar manner. The psychophysical function across all these species looks remarkably similar (Ainslie, 1992) . The exact mathematical function that describes that devaluation is debatable. Both hyperboloid and exponential functions have been suggested and both fit animal behaviour well, though they make different assumptions. The hyperboloid function, specifically an extended hyperbolic function, is given by

$$V = \frac{M}{(1 + kD)^s} \tag{1.01}$$

where V is the subjective value of the reward, M is the magnitude of the reward, D is the delay of the reward, k represents the rate at which the reward value is discounted, and s represents the sensitivity toward (or nonlinear scaling of) reward magnitude and/or delay. When s = 1, the function is a simple hyperbola. When s is less than 1, the rate of discounting decreases less sharply at higher delays.

The exponential model of delay discounting takes the form,

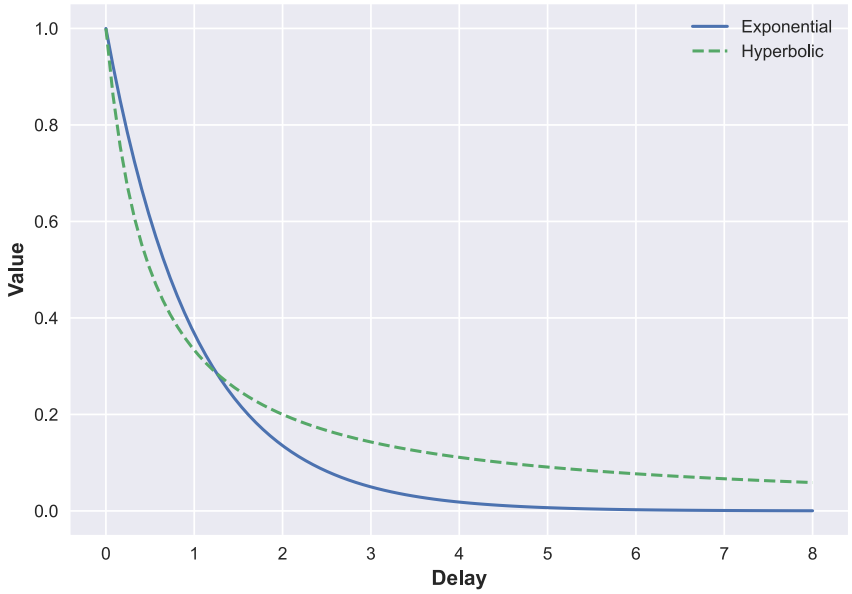
$$V = M.e^{-kD} \tag{1.02}$$

where V, M, and D have the same meaning and k has the same representation as with the hyperbolic function. Figure 1.1 shows a visual depiction of the two functions.

The similarities between the two functions in Figure 1.1 may make the distinction seem trivial. However, the importance of identifying the exact shape of the curve becomes apparent when we attempt to explain behavioural phenomena. The phenomenon of *preference reversals* is an apt example. The difference between subjective values of rewards obtained after 2 and 7 s is far greater than that between subjective values of rewards obtained after 1,002 and 1,007 s. This means that if the animals prefer a smaller reward after 2 s rather than a larger one after 7 s, and the delays to the delivery of both rewards are increased by the same length of time (1,000 s), the animals will reverse their preference and choose the larger reward at 1,007 s (Ainslie & Herrnstein, 1981). An

Figure 1. 1.

Delay Discounting Functions



Note. The figure compares the devaluation of delayed rewards using exponential and hyperbolic functions. The blue solid line shows an exponential function with the rate of discounting equaling one. The green dotted line shows an extended hyperbolic function with the rate of discounting equal to 0.5, and sensitivity equal to one.

underlying hyperbolic function is a simple explanation of this finding because it predicts preference reversals while allowing discount rates (k) and sensitivity parameters (s) to remain constant. This is because a property of the hyperbolic function is that the rate of change is not constant at all points of the curve - the rate is higher at smaller delays such that the curve loses more height per unit of time at first, and then slows down at larger delays, losing less height per unit of time. This can result in a reversal of preference as the absolute delays to both rewards are increased by the same length of time. In an exponential function, on the other hand, the curve loses the same constant proportion of height for every unit of time. To predict preference reversals, had we used an exponential function instead, we would need to allow the discount rate (k) to vary with the magnitude of the reward (Ainslie & Herrnstein, 1981; Green & Myerson, 1996). The assumption of a varying discount rate suggests that the way delay affects reward magnitude depends on the reward magnitude itself. In other words, it suggests that the computation of reward magnitude is interdependent with that of delay. To avoid making that assumption of interdependence between the computation of reward and magnitude, the hyperbolic function is heavily favoured over the exponential function among scientists in the field (see discussion of the concatenated matching law in Section 1.2 for further context on why this assumption is avoided).

While the shape of the discounting curve seems to be the same across species and individuals and is stable within individuals over time, the steepness of the curve is malleable. A steeper curve implies greater devaluation of reward with delay and a greater likelihood of choosing a smaller but more immediate reward. This is colloquially called impulsive choice. A shallower curve is thought to represent greater self-control. Differences in the steepness of the curve exist at both the species and individual levels. The rate at which reward is devalued remains relatively stable over time within an individual, provided that the environmental conditions and neural context remain constant. Environmental conditions that can change the rate at which delay devalues reward include characteristics of the reward itself. The type of reward (e.g., food or heroin), the valence of the reward (whether it is won or lost), and, in humans, the magnitude of the reward all alter the rate at which a delayed reward is devalued. Examples of neural context that have been shown to affect an individual's likelihood to delay gratification for a larger reward include the extent of their working memory being used, the extent of their hunger and their neurochemistry while making the choice (for a review, see Odum & Baumann, 2010) .

A widely used psychostimulant, methylphenidate (commonly known as Ritalin), for instance, slows the rate with which delay devalues reward in humans (Shiels et al., 2009) and other animals (Pitts & McKinney, 2005; van Gaalen et al., 2006). Therefore, it is prescribed to reduce impulsive behaviour and is also one of the most frequently abused drugs by adolescents (Bavarian et al., 2015; Zosel et al., 2013). While the acute effects of stimulant drugs on the rate of delay discounting have been identified, the long-term effects of these drugs have either not been tested or have shown conflicting data (for a review, see de Wit & Mitchell, 2010). The conflicting results in the literature on the long-term effects of methylphenidate exposure during adolescence on delay discounting

may have arisen due to the lack of important procedural controls in those studies. These are detailed in Chapter 2 which shows that rats exposed to methylphenidate at crucial points in their development showed accelerated devaluation of reward value with delay, long after cessation of drug intake.

Several testable research questions are generated from these results:

- Would methylphenidate affect the developing human brain in a similar way?
- If so, what dose of methylphenidate produces these results in humans?
- Which of the neurotransmitters affected is responsible for these results?
- What is the critical period in the developing brain when it is most susceptible to this drug's effects?
- How is the development of the circuitry underlying delay discounting interrupted by this drug's exposure?
- Does methylphenidate exposure change the animal's perception of reward or that of time, or does it change the way the values of time and reward are integrated?

That last question seems to be the most significant as well as the most neglected in the literature. It is also the most elementary. To understand how manipulations alter the mechanism underlying delay discounting, we must first determine where the manipulation is taking effect in the circuitry. Accelerated devaluation of reward by delay caused by methylphenidate exposure may stem from one or more of the following:

- a nonlinear change in the valuation of reward such that the difference between values of a smaller and larger reward is no longer the same.
- an increase in the perceived elapsed time.
- a change in the process through which reward value and elapsed time are combined.

This clear division in the mechanism involved in reward devaluation rests on the assumption that the brain computes delay and reward independently. These two values are then combined to guide the animal's choice. This assumption is made so predominantly in the literature that it is very rarely explicitly stated. The proclivity for describing the delay discounting function as hyperbolic (Equation 1.01) over an exponential one (Equation 1.02) also arises from making the same assumption. The idea that time and reward must be computed independently and later integrated may have roots (at least partially) in our view of the brain's modularity, in the way that we consciously intuit, name, and use the two constructs, and in the way that we have historically discussed the constructs in philosophy and physics. Such speculations aside, the idea is firmly rooted in literature and research on the *matching law* – a lawful account of choice behaviour.

The rest of this chapter focuses on the modular perspective of the computation of reward and time. It starts by first providing the historical context for the prevalence of the modular perspective through the matching law and its extensions. Second, it will outline reasons why that perspective

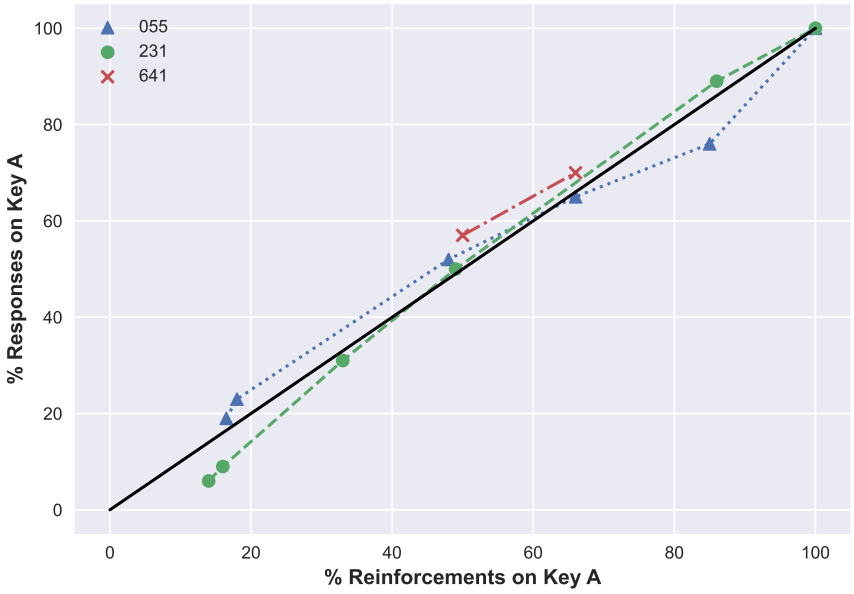
may be problematic. Third, it will present the argument that the assumption of modularity can and should be tested using the technical and theoretical advantages gained from brain stimulation reward. The chapter ends by outlining insights provided by ground-breaking research on the neural mechanisms of interval timing that further explain my hesitation in adopting that modular perspective.

1.2. Matching Law: Behaviour Goes Where Reinforcement Flows.

A vast body of studies on various species using various laboratory and outside tasks have confirmed that the matching law and its extensions fit behaviour quite well and can be successfully used to predict behaviour. The theory of matching can explain a large area of behavioural psychology because, according to the theory, all choice behaviour obeys the matching law. Furthermore, all voluntary behaviour is viewed as choice behaviour (McDowell, 2005). This section outlines three seminal experiments on the matching law by Herrnstein, Baum and Rachlin relevant to our discussion about reward and time. The three experiments elaborate on nuances in the following general message of the matching law: animals judge the relative worth of an outcome and proportionately distribute their behaviour to it rather than exclusively choosing the worthier outcome. Consider two patches with varying food density. The amount of time that animals spend in the patches is not limited to the patch with the larger food density but is proportional to (matches) the food density obtained from the patches.

Through an experimental paradigm that modeled two non-depleting foraging patches with varying food densities, Herrnstein (1961) coined and extensively studied the matching law. In this paradigm, pigeons could work for reinforcement that was available on two concurrent variable-interval schedules. On variable-interval schedules, reinforcement is made available at a variable interval after a reward is obtained. Only responses made after the interval has elapsed are reinforced. The schedules are said to run concurrently when two options are presented continuously and simultaneously, much like two different patches with different food densities. Pigeons in Herrnstein's experiment chose between pecking two keys (B_1 and B_2) that operated on different variable-interval schedules with the average interval on each key ranging from 1.5 to 9 mins. A changeover delay was instated where the pigeon could not be reinforced upon changing from one key to the other, to keep the pigeon from switching rapidly between the levers. Pigeons could keep choosing between the two keys until they had received 60 reinforcements. In this experiment, Herrnstein showed that when choosing between two different acts over the session (pecking key B_1 and key B_2), animals perfectly matched their behaviour to the reinforcement they received. The pecking ratio on the two keys equalled the ratio of the reinforcers obtained by pecking on those keys. This is called *simple matching*, is visually depicted in Figure 1.2, and can be represented mathematically as:

Figure 1. 2.
Simple Matching



Note. The x-axis represents the percentage of the reinforcers obtained by pecking key A out of the total reinforcers obtained by pecking both keys. The y-axis represents the percentage of pecks on key A out of the pecks on both keys. The solid diagonal line in the middle represents perfect matching. The ratio of pecks on the two keys absolutely equals the ratio of the reinforcers obtained by pecking the two keys. The figure shows that for three pigeons, the pecks fall very close to the line, demonstrating simple matching (adapted from Herrnstein, 1961)

$$\frac{B_1}{B_2} = \frac{R_1}{R_2} = \frac{V_1}{V_2} \tag{1.03}$$

where B_1 and B_2 represent the two different behaviours (in this case, the two different keys being pecked), R_1 and R_2 represent the rates of the reward obtained by the animal for each behaviour, and V_1 and V_2 represent the values that the animal assigns to each behaviour.

Despite the numerous variables affecting animals' choices, Herrnstein showed that their behaviour was orderly, and could be quantified, predicted, and controlled. The law of simple matching quantified behaviour in a way that had only been done with psychophysics at that time in the field of psychology (M. Davison & McCarthy, 2016). That level of quantitative analysis of behaviour allowed psychologists to control and predict the rate of the behaviour of animals, opening a new avenue of research into voluntary behaviour that has since been immensely productive (Plaud, 1992). The following two experiments show how the law was refined to represent behaviour in more diverse situations.

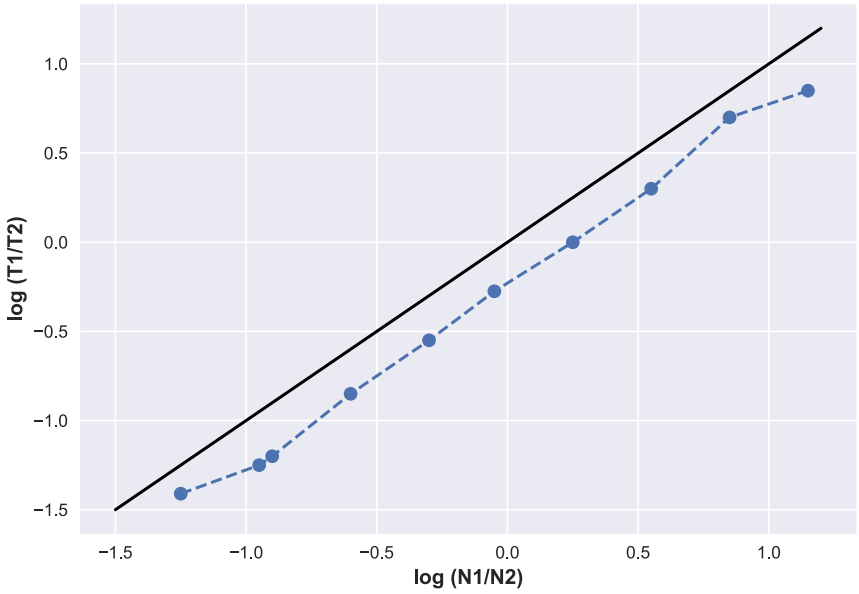
First observed in 1967 by Shull and Pliskoff, two years later, Baum and Rachlin replicated the finding that the matching law could be applied to the more general way animals allocated their time, rather than the way they allocated their time-discrete responses. The study by Baum and Rachlin is outlined next as it is more informative for our purposes. In this experiment, animals no longer had to peck keys to get a reward. Instead, their mere presence at the correct side of the chamber was sufficient for food to appear in the food hopper. Thus, the dependent variable in their experiment was the time that the pigeon spent on each side of the chamber. The reinforcement rate for each side (variable-interval schedule) was different, and food at the food hopper would only appear if the animal were near it after the interval had elapsed. The authors found *biased matching*: when the reinforcement rate from each hopper was equal, the time spent on each side was unequal. Although the animals were biased towards one side, the bias remained the same throughout the range that the reinforcement rate was varied. This is depicted visually in Figure 1.3 and is represented mathematically as:

$$\frac{T_1}{T_2} = b\left(\frac{R_1}{R_2}\right) \tag{1.04}$$

where T_1 and T_2 represent time allocation at each side of the chamber, R_1 and R_2 represent the reinforcement rates for the two sides of the chamber, and b refers to bias.

This next experiment shows another amendment to the simple matching law (Equation 1.03), and it shows that the matching law also applies to choices between qualitatively different behavioural responses. Rachlin and Baum (1972) employed a paradigm in which a single key

Figure 1. 3.
Biased Matching



Note. The x-axis represents the ratio of the reinforcers obtained from the hopper on one side of the chamber compared to reinforcers obtained by the hopper on the other side. The y-axis represents the ratio of the time that the pigeon spent on one side compared to the other. The solid diagonal line in the middle represents simple matching: the ratio of the time spent on the two sides perfectly equals the ratio of the reinforcers obtained from the two hoppers. The blue dotted line and data points represent averaged data and show biased matching: when the reinforcer rate is equal, the animals were biased towards one side. However, the bias remained the same throughout the range that the reinforcer rate was varied (adapted from Baum and Rachlin, 1969)

was presented to pigeons that, upon being pecked, led to food delivery after a variable-interval schedule. This time, they also presented food to the pigeons freely and unsignalled at a variable-interval schedule if the pigeon had stopped pecking the key for at least 2 s. When only one key was available for a pigeon to peck and alternate reinforcements were made available for not pecking that key, the pigeon still chose between pecking the key and not pecking that key in a lawful manner. Not pecking the key could then be considered an alternate action associated with its own contingencies. The pigeons thus matched their pecking and not pecking to the reinforcement that they obtained from the two actions. This shows that single responses and what seems like an absolute response rate can still be viewed as relative rates of response and, importantly, can be quantified and studied similarly. All voluntary behaviour could, therefore, now be considered choice behaviour.

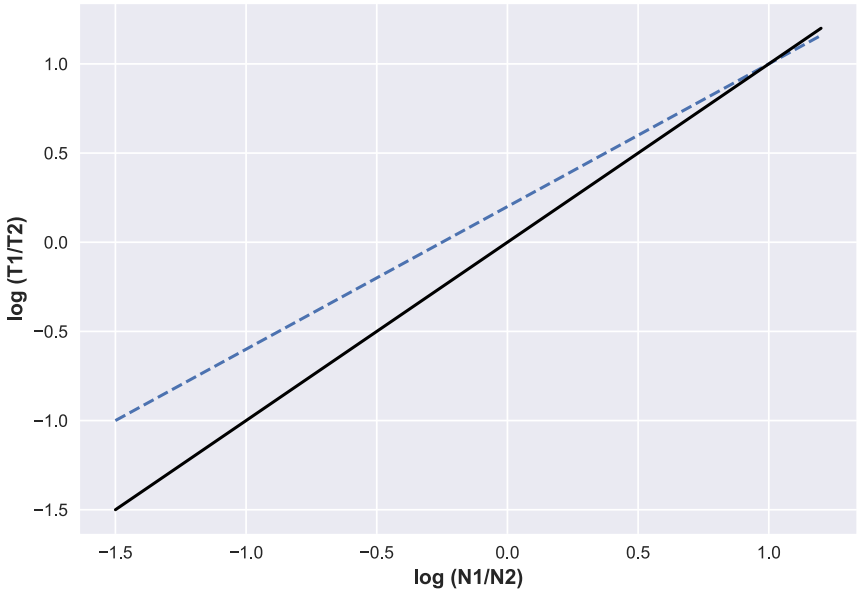
When the rates of the two reinforcers were equal, the responses of the pigeons were not. Animals showed a bias towards one response as the animals did in Baum and Rachlin's (1969) experiment. The pigeons also seemed to *undermatch* their responses to the rates of the reinforcers: the slope did not follow perfect matching. Studies since then show that animals often undermatch their responses with variations in reinforcement. This has been interpreted as a reduction in the sensitivity of the behavioural measure to variations in the reinforcement measure. In other words, if the slope were equal to zero, the key-pecking would be completely insensitive to changes in the reinforcer. On the other hand, a very steep line would imply that a slight change in the reinforcer would drastically change key-pecking. Biased undermatching is exemplified in Figure 1.4, and can be represented mathematically as:

$$\frac{B_1}{B_2} = b\left(\frac{R_1}{R_2}\right)^s \tag{1.05}$$

where B_1 refers to pecking the key, B_2 refers to all behaviours of the pigeon while not pecking the key, b refers to bias, R_1 represents the rate of reinforcement obtained from pecking the key, and R_2 represents the rate of the reinforcement obtained from the free food, in addition to any other reinforcement the pigeon may have obtained from not pecking (e.g., from grooming or resting). This is represented as R_e, where the subscript e is used to denote everything else. The additional exponent s represents the sensitivity of the behavioural measure to variations in the reinforcement measure.

This last extension of the matching equation has been termed the generalized matching law and seems to fit choice behaviour in far more contexts. Baum (1979) reviewed over a hundred datasets and found that Equation 1.05 accounted for 90.5% of the variance in behavioural allocation. It has been suggested that simple matching is a special case of generalized matching in scenarios where the bias and the sensitivity parameters equal unity (McDowell, 2005). The utility of the generalized matching law is not limited to describing and predicting how animals allocate

Figure 1. 4.
Biased Undermatching



Note. The x-axis shows the ratio of the reinforcements obtained by pressing a key that delivers food rewards compared to the reinforcements obtained from not pressing the key (this includes food given to the pigeon freely and everything else that the pigeon could do that is reinforcing, and is represented by a constant R_e). The y-axis shows the ratio of the time that the pigeon spends pecking the key compared to the time spent not pecking the key (session time minus time spent pecking the key). The blue dotted line shows an example of biased undermatching compared to simple matching depicted by the black solid line.

their time with variations in reinforcement rates. Multiple studies have confirmed that animals match their behaviour to various parameters of the reinforcement: rate (Herrnstein, 1961 and countless others), magnitude (Catania, 1963; Neuringer, 1967; Rachlin & Baum, 1972), and immediacy (Chung & Herrnstein, 1967).

Stepping out of the laboratory, however, those reinforcement parameters are more likely to vary simultaneously than in isolation. A foraging animal must choose between rewards of different magnitudes in different places (therefore, with different times to get to the reward, different levels of effort the animal may have to exert, and differences in the number and behaviour of competing animals). Animals can easily choose amongst multiple rewards with several simultaneously varying parameters. To do so, they must be able to combine the different parameters of reinforcement to compute a single value for each behaviour, so that values of competing behaviours can be compared to each other, and behaviour can be allocated according to their relative worth. Then, how do animals combine multiple parameters of reinforcement to obtain the values to which they match their behaviour?

Although the matching law describes overall distributions of behaviour rather than single choices, suppose that we were interested in examining a single choice that the animal was making in the overall context of a concurrent variable-interval (VI) schedule. A VI schedule varies the rate of reinforcement which is, by definition, the amount of the reinforcer per unit of time. If we could communicate with the animal to be aware of the reward rate and make just one choice alone, the animal would obtain, say, two food pellets delivered after a 10-second delay. Through such a simplistic view of the way animals make choices during concurrent VI schedules, matching time allocation to rate would be equivalent to matching time allocation to the product of reward magnitude and reciprocal of delay $\left(\frac{M}{D}\right)$. Notably, such a computation, should it exist, would be consistent with preference reversals that are observed due to the hyperbolic version of the delay discounting function (see Section 1.1 for a discussion about the hyperbolic vs exponential function, and Equations 1.01 and 1.02 for their respective symbolic forms). Consider the following scenario where the pigeon must choose between the following contingencies of behaviours B_1 and B_2:

M_1 = 2 food pellets, D_1 = 1 s

M_2 = 4 food pellets, D_2 = 5 s

Since $B_1 = \frac{M_1}{D_1} = \frac{2}{1} = 2$, and

$B_2 = \frac{M_2}{D_2} = \frac{4}{5} = 0.8$,

B_1 has a higher value than B_2. Consequently, the pigeon in this scenario prefers the reinforcement available from B_1 over that from B_2. Consider if the delay to both rewards were to increase by 9 s and the pigeons were asked to make another single choice, such that,

M_1 = 2 food pellets, D_1 = 10 s

M_2 = 4 food pellets, D_2 = 14 s

Since $B_1 = \frac{M_1}{D_1} = \frac{2}{10} = 0.20$, and

$B_2 = \frac{M_2}{D_2} = \frac{4}{14} = 0.29$,

the pigeon now reverses its preference so that the reinforcement obtained from B_2 is more valuable, and B_2 is carried out more often than B_1. Baum & Rachlin (1969) used this logic to model how animals combine multiple parameters of the reinforcement (see also, BGM channel, 2015). Baum and Rachlin (1969) suggested that the simplest form of the relationship between ratios of different variations in reinforcement and the ratios of behaviour produced by them would be a simple multiplication, represented as

$$\frac{B_1}{B_2} = \frac{R_1}{R_2} \cdot \frac{M_1}{M_2} \cdot \frac{I_1}{I_2} = \frac{V_1}{V_2}, \tag{1.06}$$

where *B*s represent rates of behaviour, *R*s represent rates of reinforcement, *M*s represent magnitudes of the reinforcer, *I*s represent immediacies (reciprocals of delays to reinforcement), and *V*s represent values. The subscripts 1 and 2 refer to the two alternate choices of the parameters of the reinforcement and their associated behaviours. Applying the generalized matching law (Equation 1.05) to this equation transforms it into

$$\frac{B_1}{B_2} = b \left(\frac{R_1}{R_2}\right)^{s_1} \left(\frac{M_1}{M_2}\right)^{s_2} \left(\frac{I_1}{I_2}\right)^{s_3} = \frac{V_1}{V_2}, \tag{1.07}$$

where b is the overall bias towards one alternative, and the *S*s represent the sensitivities of the behavioural measure to the variation in each parameter of the reinforcement. This has been termed the concatenated generalized matching law. Importantly, for a simple scalar combination to occur, the value of one parameter must have access to the functions output, which derives the other parameter. Since the computed reward magnitude must have access to the computed duration for them to be combined in this manner, it is assumed that the two parameters are computed separately and then integrated.

Compared to the generalized matching law tests where only one variable is manipulated, the concatenated generalized matching law tests have been much less frequent and the results far more equivocal. However, perhaps due to the success of the generalized matching law, the concatenated generalized matching law in partnership with its underlying assumptions seem to be heavily favoured among scientists in the field. The assumption is testable. For the ratios of the parameters to be integrated multiplicatively, an increase in the ratio of one parameter should be offset by an equivalent proportional increase in the ratio of another parameter along the whole range of the parameters in the same manner. In other words, the magnitude of changes observed in behavioural

allocation toward the two options should equal the magnitude of changes in the ratios of a parameter, independent of the ratios of the other parameters.

1.3. Mixed Evidence for the Concatenated Matching Law

Experimentally, the assumption has not always been supported. Interactions between rate and magnitude (M. Davison, 1988; M. Davison & Hogsden, 1984; Elliffe et al., 2008), delay and magnitude (Green & Snyderman, 1980; Ito & Asaki, 1982; Killeen, 1985; Navarick & Fantino, 1976; Snyderman, 1983; White & Pipe, 1987), and rate and delay (Alsop & Davison, 1986; M. Davison, 1983; M. C. Davison, 1976; Logue & Chavarro, 1987; Squires & Fantino, 1971) have been reported in studies that simultaneously varied both parameters. However, other studies that also simultaneously varied the two parameters have argued for multiplicative independence between all three combinations: rate and magnitude (Berg & Grace, 2004; Harper, 1982; Keller & Gollub, 1977; M. I. Leon & Gallistel, 1998; McLean & Blampied, 2001; Todorov, 1973), delay and magnitude (R. Grace, 1999; R. C. Grace, 1995; R. C. Grace et al., 2002; Rodriguez & Logue, 1986) and rate and delay (Berg & Grace, 2004).

Conflicting results from different laboratories may have arisen due to one or more of the several procedural difficulties the researchers faced in simultaneously varying both parameters. First, to be able to map behaviour onto two independent variables rather than one accurately requires a substantial increase in the number of experimental conditions. The studies conducted by Herrnstein, Rachlin and Baum varied a minimum of six different levels of one parameter of the reinforcer; requiring six experimental conditions that the animals must learn and stabilize behaviour on. The addition of a second variable would require $6^2 = 36$ experimental conditions to get the same level of accuracy. Multiple sessions for each of the 36 experimental conditions would be necessary to train the animals and obtain accurate statistical fits. This is, indeed, quite difficult using the procedures and reinforcers described so far. From personal experience, not only does it become harder to train the animal on the task, satiety and fatigue begin to affect behaviour as the session progresses. If the sessions are divided over days to combat issues of satiety and fatigue, other confounding variables, such as age will accumulate. Perhaps due to these concerns, experimenters who have studied the interaction between two reinforcer parameters have studied far fewer conditions by testing far fewer levels of the manipulated variable over a much smaller range. Most of the studies cited above consisted of between 6-12 experimental conditions. This is simply not enough to accurately represent behaviour with two simultaneously varying parameters while using an experimental procedure that has never been validated. It could be argued that 36 conditions are unnecessarily excessive, and that some generalization is acceptable. With properly chosen values of the manipulated variables, there is truth to that, but it becomes increasingly important to obtain evidence that those values are properly chosen, and the procedure sufficiently validated. Moreover, the parameter variation ranges must be large enough to show a complete representation of the animals' behaviour and understand the significance or seriousness of any

deviations from the concatenated generalized matching law observed (M. I. Leon & Gallistel, 1998).

Second, as mentioned earlier, it is not easy to remove the influence of time from behavioural measures in the laboratory. The importance of doing so while testing for interdependencies of time with other variables requires no clarification. In most of the above studies, the magnitude of the reward was varied by increasing the duration that the food was accessible (e.g. M. Davison, 1988; Michael Davison & Hogsden, 1984). Duration of food availability as an operational definition for the reward magnitude is problematic for two reasons. It has been shown that the amount of food is not a constant proportion of the duration that the food is available (Epstein, 1981), and increasing the duration that the food is available inherently increases the delay to the reward as well. Some studies used reinforcement with pauses to solve the former issue, and ensure that the food receptacle was emptied. However, it could be argued that they converted the duration into a rate, effectively replacing one interdependency with another (Elliffe et al., 2008). The addition of pauses to the measure also worsens the confound between time and magnitude because it increases the overall duration that the reward is available and introduces short delays between availabilities. Similarly, some studies looking at interactions between rate and delay extracted the average delay from a variable-interval schedule, the same schedule used to manipulate rate (e.g. Berg & Grace, 2004). Therefore, any interactions between the independent variables observed in studies using those operational definitions would be dubious.

Third, studies that simultaneously manipulated delay and rate have almost exclusively used a concurrent-chains procedure – a complicated procedure that obscures the interpretation of results even further. In a concurrent-chains procedure, the animal is confronted with a chain of schedules; the initial link followed by a terminal link. The experimenter sets equal concurrent VI schedules on both available alternatives in the initial link. In contrast, the terminal link schedule is on a fixed-interval schedule, so that a reinforcer is delivered a fixed interval (delay) after a response is made. Since the VI schedules during the initial link are set to be equal, it is assumed that any observed differences in experimental conditions are due to the schedule in the terminal link. Thus, responses during the initial link are thought to show a preference for the resulting fixed interval in the terminal link. Rate is then varied in the initial links to observe interactions between rate and delay. The response-contingencies between a concurrent schedule and a concurrent-chains schedule are quite different. During a concurrent schedule, animals choose between two different contingencies in the present, whereas during a concurrent-chains schedule, animals are thought to choose between two different contingencies in the future. There is some evidence of computational differences in how those two choices are made. Rider (1983) showed that in a concurrent fixed ratio-mixed ratio (FR-MR) schedule, animals chose to respond to the FR schedule more than the MR schedule. However, on a concurrent-chains schedule, with an equal VI-VI schedule as the initial link, and the same FR-MR schedule as the terminal link, animals switched their choice to an MR schedule, for reasons unbeknownst to us. Moreover, according to the matching law, animals match their responses to the

reinforcements they obtain rather than the reinforcements that the experimenter schedules (Herrnstein, 1961). The VIs set by the experimenter may not equal the VIs experienced by the animals. Hence, their choice in the initial link may not be as independent of their choice in the terminal link as intended. There is also evidence that the order of presentation of different schedules can influence animal behaviour (Keller & Gollub, 1977). It becomes unclear whether the results observed are due to a genuine interaction (or lack thereof) between rate and delay or due to the presentation of the different schedules.

Thus, a more thorough manipulation over wider ranges of rate, delay, and magnitude is required to test the concatenated generalized matching law. The combinations of variations must be presented randomly and with constant references to control for carryover effects (see methodology in Chapter 3). Changing the magnitude of a reinforcer while keeping its duration constant is essential to test interactions between magnitude and delay. Varying the concentration of sucrose, the dose of heroin or the frequency of brain stimulation reward are examples where that is achievable. Leon & Gallistel (1998), using brain stimulation reward as a reinforcer, avoided many of the procedural difficulties outlined above and showed that the ratios of rate and magnitude of brain stimulation reward combine multiplicatively over wide ranges (see Section 1.4.2).

1.4. Concatenated Matching Law and Brain Stimulation Reward

1.4.1. What is Brain Stimulation Reward?

Spread across brains are areas that show reinforcing properties upon electrical stimulation. The medial forebrain bundle (MFB) at the level of the lateral hypothalamus is one such area in the brains of rats. The MFB, which has been a site of much research, is a complex and heterogeneous bundle of fibers that interconnects forebrain, midbrain and hindbrain structures (Nieuwenhuys et al., 1982). To control impulse flow with precision, electrical stimulation to the brain in these studies has been delivered in bursts (termed trains) that typically last a short time (0.5 s). They typically consist of a series of pulses: 0.1ms of rectangular, cathodal, monophasic pulses separated by fixed interpulse intervals, the length of which depend on the frequency set by the experimenter.

At high enough parameter values, rats will work for the stimulation of electrodes implanted in the MFB, over food and water, even when facing starvation or exhaustion (Frank & Stutz, 1984; Olds, 1958). The approach and consummatory behaviour towards stimulation has been termed *intracranial self-stimulation* (ICSS) and the reward that rats seek by engaging in ICSS has been termed *brain stimulation reward* (BSR). In an elegant series of experiments, Shizgal and colleagues tested whether the neural signal invoked by electrical stimulation was comparable to the reward magnitude that animals compute for natural rewards. Conover and Shizgal (1994a), studied animal choice between BSR and food reward after carefully minimizing the differences in external stimuli, post-ingestive feedback and the response timing and requirements for the two rewards. They showed that the underlying value of BSR and food rewards could undergo reward summation and the two rewards compete with each other to control the animal's behaviour. Conover, Woodside

and Shizgal (1994) replicated this finding with intraoral saline. Moreover, they showed that negative sodium balance affected the value of the saline reward but not that of BSR. Conover and Shizgal (1994b) also showed that in contrast to gustatory rewards, short-term post-ingestive feedback does not affect the value of BSR. Based on these results, the authors argue that through electrical stimulation, we are tapping into a neural signal representing subjective reward magnitude at a level of neural processing that shares properties with natural rewards. This level of processing occurs downstream from the stage where physiological feedback can modulate the gustatory value. These signals are eventually mapped into reward values (Shizgal, 1997).

Their argument is logical. For competition and summation to occur between the different types of rewards and ultimately for one of those rewards to be selected, the neural representation of the rewards must share properties that allow such computations to occur. Competition and summation of rewards could then occur even though the modality with which they were detected did not share those properties. They called these properties a *common neural currency*. Additionally, the findings that the post-ingestive feedback and negative sodium balance did not affect the value of BSR showed that the point of the reward circuitry we tap into with the stimulation is at a stage of processing which occurs after the stage responsible for the modulation of reward value by post-ingestive feedback and deprivation states. The signal for food, for example, must consist of multiple different dimensions, such as colour, smell, and taste. Post-ingestive feedback and deprivation states seem to have disparate effects on different dimensions. Therefore they must affect encoding at a multidimensional stage of processing where that modality-specific information can be extracted for computation. Gallistel (1978) and Simmons & Gallistel, (1994) conducted experiments to describe how the effects of the number of stimulated neurons and their firing rate are integrated over time and neuronal space. They showed that a simple principle describes this integration process. With a fixed duration of stimulation, the integrator simply counts spikes (the *counter model*). Specifically, the output of the integrator is the product of the number of stimulated neurons and the number of firings in each neuron. Therefore, increasing the pulse frequency and increasing the stimulation current changes the aggregate spike count in the same manner. Shizgal and his colleagues proposed that the simplest explanation of the counter model and the series of experiments is that the electrical stimulation in the directly stimulated neurons is integrated using a rate code, the output of which represents the subjective intensity of reward. This construct is considered unidimensional; the stimulation invokes the signal after the multiple dimensions of rewards affected by post-ingestive feedback and deprivation states have been converted into a single dimension.

1.4.2. BSR Magnitude and Rate Combine Multiplicatively.

This puts an extraordinary opportunity at our electrode tips. We can access a single dimension within a multidimensional value structure to probe how values in those different dimensions interact. We can use this to isolate in which dimension other parameters or experimental manipulations affect the value structure. In other words, we can bypass the myriad of variables that

make up the subjective magnitude of reward, and directly tap into a simplified stage of circuitry, allowing us precise control over a signal that is comparable to that representation. We can thus study and quantify how the brain encodes reward magnitude and computes with that representation by deriving a measurement that is no longer confounded by interacting parameters of the reward. Doing so results in low variability in the animal's behaviour across a single experimental session and over days in the experiment. The technical advantages of using BSR over food reward are enormous. The experimental sessions can run for hours because short-term physiological responses do not seem to affect the stimulation, and thus the stimulation is not subject to satiety. Given that small changes in the stimulation properties can lead to significant mappable changes in the animal's behaviour, the precision of control over stimulation parameters is a paradigm shift in neuroscientific studies.

Leon & Gallistel (1998) collected a large dataset by taking advantage of these benefits. They varied both the rate and magnitude of the stimulation over wide ranges to accurately capture the rats' behaviour and test whether any deviations from the concatenated matching law, if observed, were significant. Since they could change the magnitude of the stimulation without changing the duration that the stimulation was available, they removed the confounding effect of duration on magnitude present in previous studies. The ability to directly tap into this relatively unadulterated and unconfounded stage of reward processing, with all its technical advantages, allowed them to show that the ratios of the two parameters, rate and magnitude, indeed, combine multiplicatively to determine the time animals allocate to the various options. Whether that relationship holds for magnitude and delay has not been published, to my knowledge, using the advantages conferred by BSR.

1.4.3. Does BSR Magnitude and Delay Combine Multiplicatively?

Animals readily discount the value of BSR as the delay to obtain the reward is increased. Fouriezos and Randall (1997), for instance, mapped the performance of rats on a series of trials with decreasing pulse frequencies. They then imposed 12 different delays on each of those frequency sweeps in random order. They showed that BSR, like natural rewards, is subject to delay discounting; the frequency threshold for all rats increased in a consistent and lawful manner. The hyperbolic function and the exponential function both fit the behaviour of rats equally well. The experiment, however, was not designed to test whether delay and reward magnitude interact. The rats in this study were placed on a fixed-interval schedule. Their preference for various combinations of reward magnitudes and reward delays was extracted from their performance rate on a single lever. Whether the concatenated generalized matching law applied to their data, cannot be deduced using this methodology. Both a multiplicative combination, where magnitude and delay are computed independently by the animal, and a more complex relationship where magnitude and delay interact, would lead to the same shift of the rate-frequency curve. This exemplifies the general problem with rate-frequency curves. It is difficult to determine the parameter of the reward that is shifting the rate-frequency curve. This mathematical inconvenience arises because the

subjective magnitude of the reward is not represented by the rate-frequency function but is, instead, only one of many constituent functions that make up the rate-frequency function. For example, trade-off experiments have shown that the preference for a reward (and therefore the subjective reward magnitude) continues to increase while performance for that reward has reached a plateau. This suggests that performance for the reward and the subjective magnitude of the reward have different ceilings. Moreover, Sonnenschein et al., (2003) showed that when observing behaviour through performance as opposed to the degree of preference for a reward, the parallelism of the rate frequency curves requires that delay discount curves that are derived at different reward intensities have different parameter values.

Mazur (1987) developed an indifference-point procedure to obtain the underlying magnitude of a reward at various delays. They tested by varying delays of a large reward, how much the delay of a smaller reward would need to be adjusted for the rat to be indifferent towards it. Mazur et al., (1987) used this procedure to conduct a psychophysical study on how the value of BSR decreases with increasing delays. They concluded from their analysis that a hyperbolic function fit their data better than an exponential one. While their analysis addresses some of the issues with rate frequency curves, they conducted their analysis under the assumption that the rate of discounting would be the same for both the small and the large reward, and thus that reward delay and reward magnitude are independent. Thus, they varied the delay over a wide range, but there were only two different reward strengths in the study. It is difficult to deduce from that procedure whether data deviates from the concatenated matching law.

Arvanitogiannis & Shizgal (2008) partially addressed the issues with rate frequency curves by integrating findings by Leon & Gallistel (1998) about how ratios of reward magnitude and reward rate combine, with the counter model and the extensions of the matching law. They integrated these findings into a unified framework that quantitatively models the transformation of the electrical signal responsible for BSR to animal behaviour. This model, labelled *the reward-mountain model*, is perfect for our purposes to test whether delay is combined with subjective reward magnitude after its computation in a scalar manner. Using this model, we can test at what stage in the reward circuitry delay affects reward value, during or after the integration of the action potentials in the directly stimulated neurons.

1.4.4. The Reward-Mountain Model

The reward-mountain model describes the reward system as a feed-forward information processing system. Each successive step of the model consists of a more complex representation of the reward value, as separate reinforcement parameters are progressively integrated (*Figure 1.4*). Specifically, the model describes how the reward magnitude signal resulting from electrical stimulation of the MFB combines with the costs of working for the stimulation to guide the proportion of their time that rats decide to allocate to pursue BSR. Experimentally, the magnitude of the reward has been manipulated by varying the frequency of the stimulation, and the cost has

been manipulated by varying either the rate of the stimulation (e.g. Arvanitogiannis & Shizgal, 2008) or the opportunity cost associated with working for it (e.g. Hernandez et al., 2010) .

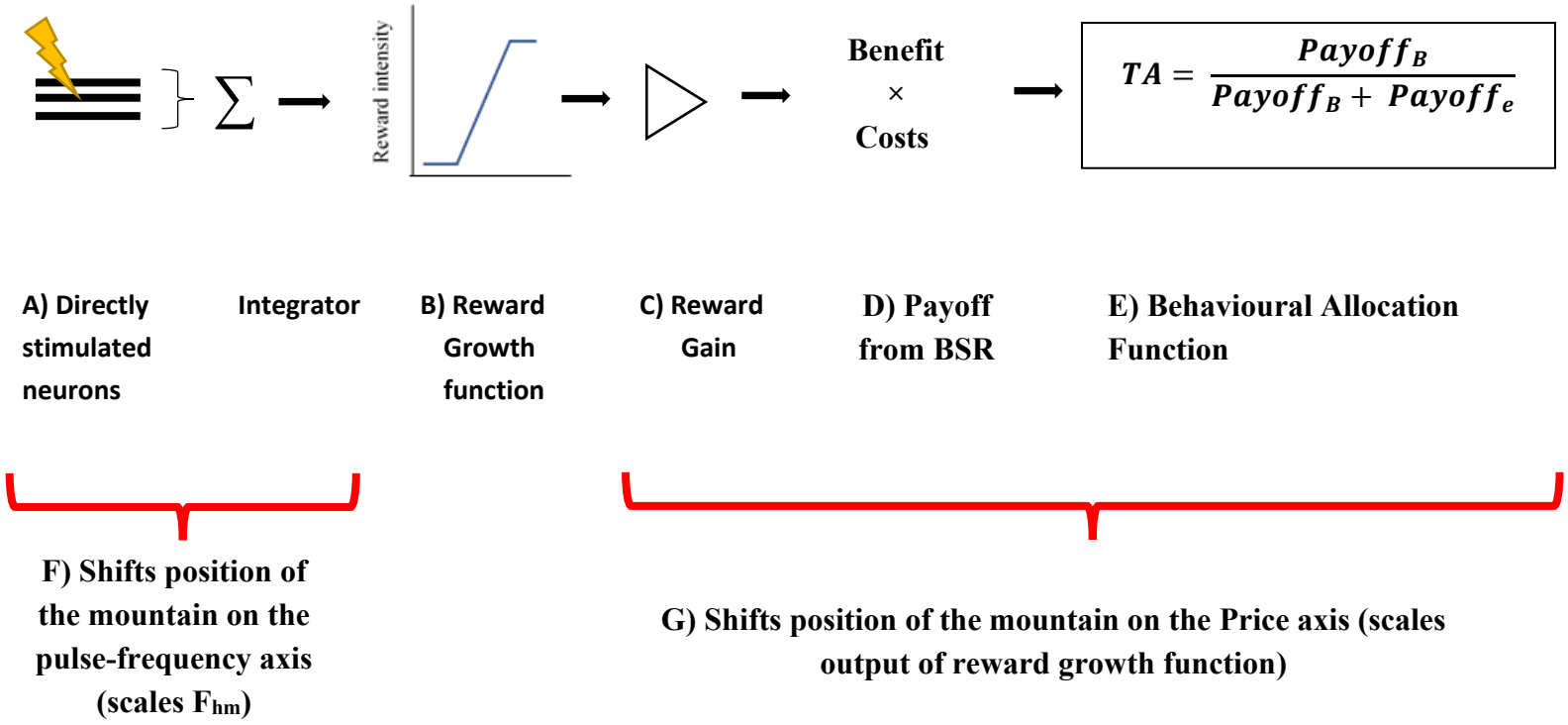
The model has since been refined and ties with the matching law, the findings of Leon and Gallistel (1998), and findings from other meticulously conducted research about how the properties of electrical stimulation in the MFB affect the directly stimulated neurons and their behavioural outcomes. For a full derivation of the model, see Trujillo-Pisanty et al., (2020). Operant matching studies by Gallistel and his colleagues have shown that the relationship between the intensity of the stimulation and the magnitude of the reward (termed the reward-growth function) can be described by a logistic function (Gallistel & Leon, 1991; M. Leon & Gallistel, 1992; Simmons & Gallistel, 1994), which in itself is a function of the strength of the stimulation (frequency and current) and the duration of the train (Sonnenschein et al., 2003). Briefly, they took advantage of the fact that researchers could control the input from which the reward magnitude is derived and then determine the underlying reward magnitude. The way manipulations may affect reward magnitude could then be observed through whether their introduction shifts or scales the function describing reward magnitude. Any changes in the strength or duration of the stimulation or their integration would change how the signal from the stimulation is transformed into reward magnitude, and would shift the reward growth function horizontally. This could then be experimentally compensated by changing the frequency of the stimulation in the opposite direction. Any changes to the output of the function that integrates the subjective reward magnitude would vertically scale the reward growth function. This mathematical deduction was made observable by modelling how animals combine rate of reward and BSR magnitude, as evidenced by Leon and Gallistel (1998). By doing so, they made observing the shifts or scaling of the reward-growth function accessible to us through the animal's behaviour. A rich enough understanding of the dimensions of reward value and how they combine allowed the researchers to derive a unidimensional scale of reward intensity as an accessible axis of the multidimensional construct of reward value.

Up till now, I have been using the term subjective reward magnitude somewhat vaguely. However, the specification of the unidimensional construct obtained from the output of the reward growth function, and the possibility of scaling that symbol, both of which would represent the benefit obtained from BSR (or the reward magnitude) necessitates the use of more precise language.

In the reward-mountain model, any experimental manipulations that affect the directly stimulated neurons or the integration of the resulting action potentials (Figure 1.5A) are said to change the reward *sensitivity.* That is, any such manipulations change the stimulation strength that is required to produce a reward of a given intensity. This results in horizontal shifts in the reward growth function (Figure 1.5B*)* and is observed as horizontal shifts of the mountain along the pulse-frequency axis. Once the signal in the directly stimulated neurons has been integrated,

Figure 1. 5.

Simplified Schematic of the Reward-mountain Model



Note. A) According to the mountain model, the number of action potentials in the directly stimulated neurons is counted by a spike counter in space and time and transformed into reward intensity. B) The reward-growth function describes this transformation. C) Once counted, the reward intensity can be scaled by multiplication by a constant of the value that was integrated; this has been called the reward gain. D) The resulting value is weighed by cost to yield a payoff from BSR. E) Comparing payoff from BSR to the payoff from all other competing rewards determines behaviour. F) Experimental manipulations that affect the activity of the directly stimulated neurons, or their integration (A) shift the mountain along the pulse-frequency axis changing reward sensitivity. G) All other manipulations (C, D, E) shift the mountain along the price axis (for instance, changes in cost, or value of competing rewards).

manipulations that would affect the reward value only have access to the integrated signal and hence, can no longer change how the neurons are firing or being integrated but can only affect the integrated intensity. Thus, they can no longer affect reward sensitivity and instead, they change reward value by scaling (multiplying) the integrated reward magnitude by the *reward gain*. The reward gain determines the maximal reward that is attainable (Figure 1.5C). As such, manipulations that either alter reward gain during its computation or scale it after it has been computed to determine payoff, for instance, the values of competing rewards or the stimulation cost (Figure 1.5 D, E), are observed as shifts of the mountain along the price axis. Hereafter, I will use subjective reward intensity or reward magnitude (interchangeably) to represent the output of the reward growth function, I will use the term reward sensitivity to describe the input to the reward growth function and reward gain to describe the scaling of the output of the reward-growth function. The reward-mountain model assumes that the subjective reward magnitude is combined with rate, opportunity costs, and payoff from other rewards in a simple scalar manner to yield the net payoff. This assumption has been validated along with the reward-mountain model on numerous occasions. The model has also been used to make predictions and can be used experimentally to learn about the reward circuitry (Arvanitogiannis & Shizgal, 2008; Breton et al., 2013, 2014; Hernandez et al., 2010, 2012; Solomon et al., 2015; Trujillo-Pisanty et al., 2011). If ratios of delay and reward intensity similarly combine multiplicatively, keeping with the concatenated generalized matching law, the reward-mountain model would predict that delaying the delivery of BSR would shift the mountain along the price axis. If ratios of delay and magnitude do not combine multiplicatively, and instead, the effect of delay changes the sensitivity of BSR, the mountain, in contrast, would shift along the frequency axis.

One might argue that evidence for the scalar combination of magnitude and rate, magnitude, and opportunity cost, and between the two payoffs should suffice or, at the least, produce strong support for a similar scalar combination of magnitude and delay. That argument is strengthened because the dominant theory of the computation of durations by the brain, the scalar expectancy theory, assumes a dedicated neural circuitry with which time is computed. This supports the assumption inherent in the concatenated matching law that reward magnitude, rate and delay are independent. If these parameters were calculated separately by individual dedicated circuits and then the calculated values combined to make decisions, an interaction between them would not be expected. Manipulating the delay at which the neurons are stimulated following satisfaction of the response requirement is not expected to alter the induced frequency of firing or the way that it is integrated.

Despite the merit to this argument and the intuitive sense that it makes, I hope to have conveyed by the discussion so far that our assumptions may not be accurate. A specific test of the independent effect of reward delay on reward magnitude is, therefore, warranted because:

- There is recurring evidence of interactions between delay and rate, and delay and magnitude using the concurrent-chains and concurrent VI schedules (notwithstanding the challenges encountered with their methodologies). Therefore, it is important and pertinent to conduct further tests to determine the interaction between the two parameters of reinforcement using a validated procedure and methodology that minimizes confounds.
- There is evidence of interaction between the magnitude of a reward and the delay to its delivery in humans. No other species-specific differences have been observed in the many studies examining the matching law. Thus, there is no *a priori* reason to believe there would be species-specific differences in the neural integration of magnitude and delay.
- The importance and ubiquity of the interaction of reward magnitude and reward delay in decision-making is well-established. Our ability to quantify the way animals make choices and to understand the brain's treatment of durations thus rests on understanding that interaction. Thus, it seems unacceptable to continue modeling animal decision-making or behaviour under the assumption that subjective reward magnitude and subjective reward delay act independently, without an explicit test of this phenomenon.

To further explain my lack of confidence in the assumption that the computation of delay is independent from that of magnitude, I must digress and provide some background on our current understanding of the neural computation of time. There is an ongoing investigation into a field that has grown up simultaneously but for the most part independently in different laboratories.

1.5. Fundamentals of Time.

It seems pertinent to begin the discussion with an attempt to attain some common notion of what time is and whether time is a unitary construct in the animal brain.

1.5.1. What is Time?

We can feel time passing by, but the question of what has passed is difficult to answer. That distinction between subjective time and objective time is important. It distinguishes between the work of psychologists from that of physicists, and between a measure that is idiosyncratic and notoriously stochastic from one that is objective and universal (given identical frames of reference). Some physicists consider objective time, along with space, to be a fundamental property of the world. In the same vein, a dualist view of the body-mind allows for subjective time to be a fundamental property of our minds. However, I have approached this discussion and my research from a materialistic perspective, in which time must be constructed from more basic properties of the brain.

According to most textbooks, physical time is a linear continuum of instants and can be most easily defined by its measurement through clocks, essentially any repetitive, continuous, process. Thus, time could be measured in days using the earth's rotation around the sun, or 1.2×10^{-17} s by measuring emissions of light pulses by lasers. Subjective time – the representation and estimation of physical time in the brain – is also sensitive to various timescales. The timescales

may be divided into three main categories: large intervals (days to years), small intervals (seconds to minutes), and very small sub-second intervals (less than a second). I will call the small intervals of seconds to minutes *interval timing* for the rest of this discussion. While subjective time may be categorized differently or with different timescales, I have chosen this categorization because it is most frequently observed in the literature and because it provides the basic context required for the present discussion. Estimations of these timescales seem to serve different functions. Timing within milliseconds is essential for coordinated movement. Timing within seconds to minutes allows for the prediction of events and plays a role in decision-making. Timing within days is necessary to regulate appetite and sleep-wake cycles (Buhusi & Meck, 2005). Whether there is one common mechanism in the brain or different ones to measure all timescales the brain is capable of is a matter of heated debate in the current timing literature. I outline what we know about the brain's measurement processes for these timescales in the following subsections and discuss why this categorization is the most appropriate.

1.5.2. Large Intervals - Circadian Timing and Ultradian Timing vs. Interval Timing.

For many species, modifying behaviour somewhat precisely with recurrent environmental changes is essential for survival and reproductive success. Annual events (e.g. seasons) dictate some species reproduction, migration or hibernation (Kreitzman & Foster, 2010). Marine animals often exhibit circalunar or circasemilunar reproductive rhythms (Raible et al., 2017). The rising and setting of the sun clearly affect how animals capitalize on gaining essential resources like light and food. Circadian timing is an essential factor in most species of animals and can also be seen in plants, fungi, and cyanobacteria. Circadian rhythms affect a multitude of mammal behaviours, including sleep-wake cycles, hormonal changes, mood changes, and the effects of drugs. There is also some evidence that the disruption of circadian rhythms is associated with harmful effects on human health (Patke et al., 2020)

Animals have biological rhythms that can guide their behaviour over days, months, or years. These rhythms have been explained by self-supporting endogenous *oscillators* which are, functionally like clocks. One of the defining characteristics of those oscillators is maintaining a rhythm even in the absence of external influences (Gibbon et al., 1997; Schibler & Sassone-Corsi, 2002). Extensive research on the circadian oscillator has identified some of the neurophysiological and biochemical bases of these endogenous clocks. Like physical clocks, the brain employs repetitive, continuous processes to measure these set intervals of time. For example, circadian rhythms are thought to be encoded by an area of the brain called the suprachiasmatic nucleus, made up of approximately 20,000 neurons. Within each neuron, feedback loops that regulate transcription cause recurrent rhythms in some RNAs and proteins that last approximately 24 hrs (Reppert & Weaver, 2002). Without external influences, these rhythms deviate slightly from 24 hrs (the standard deviation in laboratory mice can be as low as 10-20 mins). However, external influences, such as sunlight, can entrain the oscillators so that the activity occurs, on average, every 24 hrs and becomes synchronous with the external rhythm of

the sun (Herzog et al., 2004; Schibler & Sassone-Corsi, 2002; Welsh et al., 1986). Thus, these endogenous oscillators can control processes occurring every 24 hrs or at some regular point within those 24 hrs but do not count the number of 24-hr oscillations that occur. Circadian oscillators exist in almost all cell types in the brain and body and are controlled by the suprachiasmatic nucleus. The suprachiasmatic nucleus is thus the master clock that controls all other endogenous circadian clocks in mammals. Extensive research, especially on circadian clocks, has given us a very clear picture of these endogenous timing processes on cellular, tissue and organismal levels. For a review, see Patke et al. (2020).

The mechanisms underlying circadian rhythms and those that underlie interval timing are distinct and seem to have no direct relationship. For instance, lesions of the suprachiasmatic nucleus, do not affect the ability to estimate intervals in the seconds to minutes range but extinguish the rhythmicity of 24-hr cyclic behaviours (Lewis et al., 2003). Specific proteins that alter the precision of circadian timing also seem to have no effect on the precision or accuracy of interval timing in rats (Cordes & Gallistel, 2008; Papachristos et al., 2011). The difference between the constructs that circadian and interval timing measure are clear. Through circadian timing animals anticipate instances that are cyclic and occur at fixed times in the cycle.

In contrast, through interval timing, the animals do not rely on any such obvious cycle and events can be timed while measuring intervals without any external vantage point. As such, circadian timing can be compared to the position of the hour hand on a clock, whereas interval timing can be compared to the duration obtained from a stopwatch started at one instance and stopped at another. While circadian timing is automatic and takes several days to alter (for instance, it takes days to recover from jet lag), interval timing is measured from the presentation of a stimulus. It can therefore be reset on a moment-to-moment basis. While there has been much progress over recent years in understanding the mechanisms that underlie circadian timing, the mechanisms that underlie interval timing have remained elusive. The idea of a master clock similar to the suprachiasmatic nucleus seems to have inspired the investigation of mechanisms that underlie timing in the sub-second to interval range. However, given the very different problems being solved by circadian timing compared to those solved by interval timing, and the different selection pressures being faced at those timescales, there seems to be no reason to assume that animals would appropriate similar solutions for the two issues.

1.5.3. Sub-second Timing vs. Interval Timing.

The categorization of intervals as under a second in sub-second timing and over a second in interval timing (used interchangeably with supra-second timing) is largely based on our definition of a second in objective time. If a cut-off point exists, it likely occurs somewhere in the millisecond range (Karmarkar & Buonomano, 2007). I use the terms sub-second timing and interval timing in this discussion to reduce the stress on that arbitrary cut-off point. The distinction between the two timescales is generally made for the following two reasons.

First, the problems that need to be solved during sub-second and interval timing are different, though more subtle than the comparison we made with circadian timing. Interval timing plays a prominent role in judgment and decision-making, whereas sub-second timing is essential for perception and motor behaviour (Buhusi & Meck, 2005). For instance, we can detect the direction of a sound because sound reaches one ear milliseconds before another if it is not coming head-on. Speaking, dancing, playing music, playing various sports, and driving a car all require precise sub-second timing to execute. Second, unlike interval timing, sub-second timing occurs outside conscious awareness and cognitive control (Michon, 1985). Increasing cognitive load affects performance on tasks requiring interval timing, but has no effect on tasks requiring sub-second timing (Rammsayer & Lima, 1991). Lewis et al. (2003) argue for two separate systems that control timing on the two timescales: an automatic one that controls sub-second timing, and a more cognitively controlled system that exists for interval timing.

Despite the differences outlined above, the abundant similarities between sub-second timing and interval timing cannot be ignored in our search for mechanisms that underlie interval timing. Indeed, timing in both the milliseconds and seconds range have often been considered to rely on the same underlying mechanisms by other researchers (Church, 1984a; Macar et al., 2002). The differences in functions and conscious awareness of the two time periods could result from overlapping or interacting mechanisms. The functions underlying sub-second and interval timing are indeed different, but in contrast to circadian timing, the construct of durations being measured remains, arguably, the same in both. The functions underlying both sub-second and interval timing require greater continuity, more precise discrimination, the ability to time from an unknown vantage point, and thus a greater involvement of memory, than circadian or ultradian timing. Our inability to identify a precise cut-off point between the two categories also casts doubt on the complete independence of the two mechanisms underlying sub-second and interval timing.

There is no known neurological disorder or experimental lesion that results in complete elimination of one's ability to time within the sub-second or interval range, in contrast to lesions of the master clock, which abolish the circadian rhythms. Sub-second and interval timing depend heavily on the task at hand. Encoding several different stimuli from different sensory modalities can start the duration to be timed, necessitating a more widespread neural code. Some studies have argued for different underlying neural circuitry between the two timescales (Lewis & Miall, 2003; Rammsayer, 1999; Rammsayer & Lima, 1991). However, the recruitment of differing circuits in those studies can be interpreted to arise from different dependent variables. If timing at these scales is indeed dependent on the task at hand, the recruitment of different neural circuits is conceivable, even if the mechanism used to time those tasks were the same. Thus, while the distinction between the mechanisms underlying circadian timing and interval timing is quite clear, the same confidence can not be attained about the distinction between mechanisms underlying sub-second timing and interval timing.

1.6. Neural Models of Interval timing

Returning to the main discussion, we must ask ourselves whether the neurological mechanisms proposed to underlie the computation of interval timing support the idea of a scalar combination of reward magnitude and delay. Neurologically, a scalar integration of the ratios of magnitude and delay implies that the two are being computed and stored separately. Those stored values are combined later, as in the reward-mountain model with magnitude and rate (see Figure 1.5). This idea thus supports a modular view of the way the brain computes and stores the two magnitudes. A theory of timing called the *scalar expectancy theory* has largely dominated the history of research on timing and decision-making, which follows this modular perspective.

1.6.1. Scalar Expectancy Theory

Gibbon proposed the scalar expectancy theory in 1977. It gained popularity because of its clarity, precision, and ability to account for the behavioural properties of interval timing that have been observed both in humans and animals. The scalar expectancy theory was developed and consequently named after its ability to explain scalar variability, the well-replicated finding that the error that animals show in estimating a duration scales with the magnitude of the interval to be estimated. The greater the magnitude of the interval, the higher the expectancy of estimation error. Over the years, researchers have collaborated with Gibbon to build the model, further enhancing its explanatory power. This model is straightforward and comprises multiple distinct components that allow the mapping of each component onto specific brain structures or neurotransmitter systems. It consists of three main processes: a clock (comprising a pacemaker, a switch, and an accumulator), a mechanism for memory, and a mechanism for decision-making.

Briefly, as the interval to be timed begins, the pacemaker emits pulses that are collected in an accumulator. The switch can turn on or off to control whether pulses can pass through to the accumulator. As the interval to be timed ends, for example, at the delivery of a reward, the magnitude of pulses stored in the accumulator is transferred to memory. During subsequent timing in a similar situation, the previously stored memory can be retrieved and compared with the new value in the accumulator. Once a threshold in the comparison is met, action towards obtaining the reward occurs. A short description of the components of the model (for further information see Church, 1984b; van Rijn et al., 2014) is elaborated below. This will facilitate discussion about the prediction arising from the scalar expectancy model that the subjective magnitude of the reward and the subjective delay to its delivery would combine in a scalar manner and some of the problems the model might encounter.

Clock. The clock consists of the *pacemaker*, the *accumulato*r, and the *switch*. The pacemaker is thought to function analogously to ticks of a clock, where it continuously and rhythmically emits pulses (λ) for the duration of an interval (t). The pacemaker has an average rate of pulse generation that can be varied by multiple factors, such as drugs, diet, and stress. The pulses emitted by the pacemaker can be sent to the accumulator if the switch is closed, such that

the number of pulses sent to the accumulator would equal to λt. The accumulator can be reset whenever necessary. The purpose of the switch is to accumulate the pulses until they need to be transferred to memory. Pulses cannot be transferred to memory if the switch is open. The switch closes when a stimulus is presented, indicating the interval needs to be timed, and opens on completion of the interval. The switch, however, takes some time to close after the presentation of the stimulus (t_1), and to open at the end of the interval (t_2) to be timed. Consequently, this affects the number of pulses accumulated, and therefore the accuracy of the estimation of the interval. Thus, the number of pulses sent to the accumulator (a) is formalized as:

$$a = \lambda\,(t + (t_2 - t_1)) \tag{1.08}$$

To summarize, estimation of the interval depends on the pulse rate, the latency for the switch to operate, and of course, the duration of the interval.

Memory. Upon presentation of a stimulus indicating the end of the interval, the switch opens, and the number of pulses counted from the accumulator is shifted into memory. Given that memory is imprecise, the number of pulses in memory (m) does not equate to the number of pulses sent from the accumulator. Instead, those pulses are multiplied by a constant k, which is normally distributed and thought to approximate to 1. Although it approximates to 1, the constant varies in individuals and can also be affected by factors such as drugs, attention, and brain lesions. Thus, the value in memory can be formalized as

$$m = ka \tag{1.09}$$

Each subsequent presentation of the duration updates the value in memory, such that the memory stores multiple representations of the same stimulus. The representations in memory are thought to be normally distributed. The multiplication of a constant with the accumulated value leads to wider distributions in memory, for longer durations relative to shorter durations. That is the basis for the scalar property of interval timing, according to the scalar expectancy theory.

Decision process. The decision to produce the response (coinciding with the end of the interval) requires a comparison between the current value in the accumulator and the value of the duration stored in memory. As mentioned previously, there is no correct value stored in memory, but instead, there are a set of normally distributed possible values. The decision process involves comparing the current value of a to a random value of the normally distributed m's (Church, 1997). That comparison (c) involves taking a ratio of the absolute difference of the value in the accumulator and stored in memory to that of the value stored in memory. Thus:

$$c = \frac{|m - a|}{m} \tag{1.10}$$

Once that ratio reaches a threshold (b), the animal initiates the decision (d), a response coinciding with the end of the timed interval:

$$d = \frac{|m - a|}{m} > b \tag{1.11}$$

The value of b was initially thought to be a constant (Gibbon, 1977), but is now considered normally distributed and is another point in the model where variability could arise through multiple factors (Church, 2003).

The scalar expectancy theory is detailed and precise and considers many of the properties shown by interval timing. The theory has also been extremely useful for making predictions and discovering properties of interval timing (Church, 2002). Not surprisingly then, it has been the leading and most-cited model for interval timing over the past half-century. However, an unconditional general acceptance of the scalar expectancy theory would be somewhat premature.

1.6.2. Issues With Scalar Expectancy Theory

Conceptual and practical problems with the SET model have been noted. Gallistel & King (2009) note that the assumption of a dedicated clock that resets its timer places the model in an unachievable situation. Animals can time a whole host of intervals, whether they know that the interval needs to be timed in advance. Animals can also time intervals between neutral stimuli. For animals to do that using SET's clock, they would either need to have timers for an infinite number of events, or predict the future. Neither scenario seems plausible. Gallistel and King suggest that the duration to be timed must be computed retroactively to circumvent that problem. Thus, for example, what is stored are timestamps of events, that can then be extracted to compute durations. Only what is experienced is recorded to memory, which is still a large number of events, but the number is finite and thus plausible. The former scenario, a dedicated clock, does not allow for shifts on the frequency axis of the reward-mountain model. The latter scenario, in contrast, makes that allowance. Timestamps need not be stored in a centralised, dedicated space. They could be stored and used in neuronal spaces that may affect the magnitude of the reward before or during its computation (see Section 1.6 for further discussion).

Higa & Staddon (1997) pointed out that despite the relative precision of the scalar expectancy theory, it remains vague in some respects. They argue that the theory does not allow the prediction of behaviour before the animal has been trained for the task at hand. Thus, prediction about the dynamics of encoding time is difficult using this theory. The behaviour of animals can only be predicted once the representation of time has been stored in their memories. Staddon & Higa (1999) also noted that scalar variability arises from noise associated with encoding the duration in reference memory according to the model. The pacemaker-accumulator part of the model does not play a role in the scalar variability. Thus, external factors can be explained through changes in the pulse rate of the pacemaker, the constant (Equation 1.08), or remembered time (Equations 1.09, 1.10). The above two considerations make the pacemaker-accumulator essentially redundant in the model's explanatory power. Could the same behaviours and characteristics observed in interval timing procedures be explained without a dedicated endogenous clock? An entire issue of the *Journal of the Experimental Analysis of Behaviour* was

dedicated to this discussion in 1999 (vol 71, issue 2). The scalar expectancy theory remains popular in the field today.

Despite the popularity of the scalar expectancy theory, numerous other models have been proposed for the way that the brain computes time. These models consist of a wide variety of potential biological mechanisms, ranging from properties of single neurons (Grossberg & Schmajuk, 1989; Tieu et al., 1999), neural oscillators read out by coincidence detectors (Matell & Meck, 2004; van Rijn et al., 2014), reverberating loops within the cerebellum (Medina et al., 2000; Yamazaki & Tanaka, 2005), slowly climbing activity in cortical neurons (Durstewitz, 2004), synfire chains (Diesmann et al., 1999) and stochastic decay of memory traces (Kitano et al., 2003). Given the plethora of mechanisms that vary so drastically in their basic postulates, it appears from the current state of the literature that we are absolutely stumped by the way animal brains estimate time or how they store the estimated intervals for further computation.

Some of these models assume endogenous clock-like mechanisms where, following a modular perspective, segmentations of the brain are dedicated to compute durations as a single dimension, which can be stored in memory for later comparison and further computation. These models predict shifts on the price axis of the reward-mountain model. Other models allow for durations to be encoded and/or stored ubiquitously using non-dedicated neural mechanisms, which allow for shifts on the frequency axis and the price axis of the reward-mountain model.

As progress is made in this field, the objective is to map models that provide convincing accounts of behaviour onto specific neural mechanisms. Mapping of the model onto neural mechanisms not only has practical significance but is essential for theoretical reasons. It is the discovery of those neural mechanisms that is the ultimate validation of the veracity of a model. The scalar expectancy theory has not yet met that standard. Over the past 50 years, extensive research has been unable to isolate components of the pacemaker-accumulator model or provide a neural basis for any component. Of course, the absence of evidence does not imply evidence of absence. However, it does make a case for moving some of our eggs into other baskets and extending the search for neural mechanisms beyond an interpretation for a neurally unsubstantiated model. This search, although limited, has already begun, and it supports a non-dedicated mechanism underlying interval timing.

1.6.3. Non-dedicated Models of Interval Timing

Models that assume a non-dedicated circuitry responsible for interval timing are not without their challenges, but they allow for a more ubiquitous encoding of time. Neural mechanisms that support this ubiquity of interval timing have been identified. Through pharmacological and imaging recordings, it has been difficult to find areas of the brain that are *not* correlated with timing tasks (Buonomano, 2017). Firing activities, for instance, within various brain regions are timed.

A particularly striking example of this can be seen through experiments conducted on songbirds. Songs of male songbirds hold structural parallels to speech, where notes are combined to form syllables that can be up to a few hundred milliseconds long. The syllables are then combined in sequences with pauses, usually less than a hundred milliseconds long, to form phrases. These songs are often a few seconds long. Neuronal activity within a nucleus called HVC (formerly, hyperstriatum ventrale, pars caudalis [HVc], and high vocal center, and now, just HVC) in zebra finches is considered critical to song learning and production. Neurons in this nucleus also show timed firing activity; neurons burst at specific times during the song. An approximately 6 ms burst of spikes is timed to a particular moment in every repetition of the song motif. It is thought that the neurons fire in a chain where each neuron's firing is responsible for a sequence of timed firing in other neurons, often compared to using a chain of falling dominoes as a timer. We would know the time elapsed depending on which neurons in the sequence were active or which domino had fallen (Hahnloser et al., 2002; Long & Fee, 2008). To test whether this timed sequential firing activity had a causal link to the timing heard within a song, Long and Fee (2008) cooled the nucleus of songbirds up to 6.5°C below body temperature to slow down the rate of action potentials. They showed that this slowed down the rate of singing uniformly across the entire song without changing the acoustic structure of the song. Slowing down the activity in the motor area (robust nucleus of the arcopallium [RA]), to which the HVC neurons project did not affect song timing. One could argue that a structure upstream or downstream of the HVC acts as the clock and controls the onset of the firing responsible for the sequence of firing that was slowed down. However, cooling the HVC also slowed down the onset of syllables, consistent with the idea that the timing mechanism is generated within the HVC.

It has been suggested that most if not all neural circuits can perform timing as a computation. Even *in vitro* slices of the cortex are sensitive to intervals between stimuli to which they are exposed. Johnson, Goel and Buonomano (2010) trained *in vitro* slices of the cortex to change their activity depending on the duration of two stimuli. Different subsets of neurons in the cortical slices were stimulated to fire twice (first electrically and then optically) at three different intervals (100ms, 250ms or 500ms) for a few hours. When exposed to a brief electrical pulse after the training, the researchers found that the internal dynamics differed depending on the interval at which the slices were trained. The stimulation led to longer bursts of neural activity, the longer the interval between the two stimulations during the training period.

This same phenomenon has also been shown with slightly longer durations. Shuler & Bear (2006) showed that slices of the visual cortex of mice could be taught to report time. In a similar experiment, *in vitro,* visual cortex slices were stimulated twice, first electrically to simulate a visual signal and then with a cholinergic agonist to simulate a reward. The intervals used in this study were 0.5, 1, and 1.5 s long. In line with results from Johnson, Goel and Buonomano (2010), an electrical test stimulus led to longer bursts of neural activity, the longer the interval between the two stimulations during the training period. What is noteworthy here, apart from the finding

that slices of the cortex can time and learn durations, is that the slices were taken from the primary visual cortex very early on in the pathway of sensory processing.

Recent evidence illustrates that neural sequences are not the only dynamics through which the brain can infer time. Johansson and his colleagues (2014) showed that even single neurons could encode timing activity while investigating the neural mechanisms underlying conditioned eye blinks. In a conditioned eye blink study, the animal learns the duration between the conditioned stimulus and the puff of air directed to its eye and can, consequently, blink at the correct time to protect its eye against the puff of air. This learning has been observed in ferrets who have had their entire forebrain removed, leaving only the very primitive hindbrain and midbrain regions intact. Specifically, eyeblink conditioning relies on circuitry in the cerebellar cortex. Information about the conditioned stimulus is received by large Purkinje cells through a bundle of parallel fibers that they are monosynaptically connected with. Whereas information about the unconditioned stimulus is received through single afferent climbing fibers wrapped around them (Moore, 2002). Johannsson et al. (2014) showed that these single cerebellar Purkinje cells could be taught durations. In this experiment, the researchers electrically stimulated the bundle of parallel fibers connected to a Purkinje cell to mimic the presentation of the conditioned stimulus (e.g., a tone). To mimic the unconditioned stimulus (puff of air), they directly stimulated the climbing fiber connected to that Purkinje cell. They could now control both inputs to the Purkinje cell. In doing so, they neurologically recreated the eye conditioning experiment by stimulating the climbing fiber at different intervals following the stimulation of the parallel fibers. Remarkably, the single Purkinje cell learned, remembered, and reproduced those durations on probe trials in which the climbing fiber was not stimulated. The neuron, which has a high endogenous firing rate, stopped firing when the conditioned stimulus was presented, and resumed firing around the time that the unconditioned stimulation was presented in the past. The pause in firing was independent of the parameters of the input stimulation of the parallel fibers after the training but occurred upon the stimulation of the input despite its excitatory nature. After repeated exposure to the pattern of inputs, the cell changed its behaviour in response to the input stimulation, reflecting with high fidelity the information that it had received in the past. If the directly stimulated neurons responsible for BSR or their afferent integrator are similarly capable of timing, the computation of reward magnitude and the delay to obtain it could be interdependent.

Many of the studies mentioned above showing the intrinsic timing ability of neural circuits and neurons have been conducted in the past decade. They support the idea that the timing of durations is ubiquitous in the brain. Given that timing is a ubiquitous problem that animals must solve and that it is required for a whole host of different problems (from perception to learning to decision-making), this is not surprising. The capability in single neurons of storing durations, combined with the ubiquity of timing widespread in brain regions and sensory modalities, makes it seem unlikely that dedicated neural circuits exist for timing durations. Immanuel Kant

describes objective time (along with space) as an irreducible category through which brains perceive reality. That perspective is conceivable. Timing may be built into and extracted from neural dynamics. If that radical stance is conceivable, we can not rest assuming that time and reward are encoded and computed independently. In Chapter 3, I show support for the idea that the neural representation of time and reward may not be independent and that it may now be time (no pun intended) to focus our efforts on a thorough investigation of timing mechanisms beyond the scalar expectancy theory and the belief of a neural circuitry dedicated to this task.

1.7. Summary

I explore the integration of the neural computation of reward magnitude with the delays after which the reward is obtained. Intrigued by the lawful way that delay discounts behaviour across species and how stimulant drugs can decrease that rate of discounting, I first tested whether stimulant drugs prescribed to decrease impulsive behaviours affect the underlying mechanisms and cause long-term effects on those same behaviours. Chapter 2 shows that the developing brains of rats are more susceptible to long-term effects of stimulant drugs on impulsive choice than fully developed brains. This brought up the question of what the drug was affecting, the magnitude of the reward, the perception of duration, or how the two were combined. A review of the literature showed that I was making the above categorization on an often made, yet unsupported, assumption that reward magnitude and durations are encoded separately and combined in a simple scalar manner. Chapter 3 uses intracranial self-stimulation and the reward-mountain model to show that reward magnitude and duration are not as independent as is often assumed.

Chapter 2

Adolescent Exposure to Methylphenidate Increases Impulsive Choice Later in Life

2.1. Abstract

Background: The psychostimulant methylphenidate (MPH) is known to temporarily reduce impulsive choice and promote self-control. What is not sufficiently understood is how repeated treatment with MPH affects impulsive choice in the long run, and whether any such effect is contingent on exposure at certain developmental stages.

Methods: Using an animal model for impulsive choice, we examined first whether giving MPH through early adolescence alters delay discounting, an operational measure of impulsive choice, later in adulthood. We then tested whether equivalent long-term effects are observed if exposure to the drug occurred during adulthood. Starting on postnatal day 25 or postnatal day 60, male rats received one of a range of doses of MPH for 10 consecutive days. Twenty-six days later, all rats were trained to choose between a lever that produced a small immediate reward and a lever that produced a large reward after a range of delays.

Results: Rats showed a long-term decrease in the selection of the delayed larger reward when treated with moderate doses of MPH during early adolescence, but not when treated with the lower or higher doses. In contrast, no differences were observed in the selection of the delayed larger reward in animals that were treated with various doses of MPH during adulthood.

Conclusions: Our findings suggest effects of MPH on impulsive choice that are contingent on dosage and on the developmental period of exposure. When administered during adolescence, moderate doses of MPH increase impulsive choice long after the end of treatment, whereas these same doses administered during adulthood were without effect.

2.2. Introduction

Time is money. A payout depreciates in value the more time it takes to earn it, to the point where a smaller, but more timely payout starts to look more worthwhile. What is true for money is also true for other rewards, be it a glass of wine, a cigarette, or a sugary snack: if a larger reward takes too long to earn, a smaller, quicker one wins out. For some individuals, this critical point is reached more quickly; they are said to show impulsivity. Others will wait longer for the same reward, showing self-control (Ainslie, 1975; Rachlin & Green, 1972).

While these traits tend to be stable, psychostimulant drugs can dislodge and shift them. One such drug, methylphenidate (MPH), temporarily slows the erosion of reward value in humans (Pietras et al., 2003; Shiels et al., 2009) and other animals (Pitts & McKinney, 2005; van Gaalen et al., 2006), effectively curbing impulsivity and imposing self-control. As such, it is

prescribed under trade names like Ritalin© to treat impulse-control disorders, and is commonly used illicitly by adolescents to improve their academic performance (Low & Gendaszek, 2002; Zosel et al., 2013).

Whereas the acute effects of MPH on impulsivity are well known, the long-term effects have received little attention. This is surprising, considering the large body of research elucidating persistent effects of MPH (Andersen et al., 2002; Urban & Gao, 2015) and other psychostimulant drugs (Koob & Moal, 1997) on reward value. These same studies have shown the effects to depend on the stage of development at which drug exposure occurs, with particular vulnerability attributed to early adolescence. While there are no absolute boundaries to the stages of adolescence, earlier in adolescence humans and non-human animals show crucial maturation of brain regions associated with impulsivity. For example, the prefrontal cortex and its dopaminergic connections are still developing and are subject to disruption in rats during that period (Reynolds et al., 2015). So far, three studies in rats have examined the long-term effects of MPH exposure during adolescence, and none have characterized its effects with adult exposure. The three adolescent rat studies showed mixed results; one found no long-term effect (Pardey et al., 2012), and the others reported long-term reductions in basal impulsivity (Adriani et al., 2007; Leo et al., 2009). Their methodologies, however, impose caveats on their interpretation.

To observe the curtailing effects of delay on reward valuation—known in behavioral research as delay discounting— animal studies employ operant conditioning chambers with two levers: one delivers an immediate small reward, and the other a larger reward after varying delays (Ainslie, 1975). Typical control procedures include within-session changes in delay for the large reward, counterbalancing of lever locations, training rats for center nose-poking at the beginning of trials and most importantly, training them for the association of delay with the large reward (Evenden & Ryan, 1996). None of the three adolescent rat studies employed these control procedures. Such procedural differences could lead to contrast effects and selection bias toward one of the levers (Richards et al., 1997), obscuring the findings of the experiments. As a result, it is unclear whether MPH has any long-term effect on delay discounting.

In the present study, we examined whether MPH produces a persistent effect on the devaluation of reward by delay in rats. We used a modified version of the delay discounting paradigm commonly used in the literature in order to circumvent the potential confounds listed above. Given the specificity of psychostimulant drug effects to early developmental stages, we first tested the effects of early adolescent exposure on long-term impulsivity. In addition, we used a smaller sample to determine whether similar effects could be reproduced with adult exposure. We found that moderate doses of MPH administered during early adolescence, but not during adulthood, resulted in persistently reduced choice for the delayed larger reward, indicative of greater impulsivity.

The results of this work have been published (Abbas et al., 2017).

2.3. Methods

2.3.1. Experiment 1

Subjects. Male Long-Evans rats (Charles River, St. Constant, QC) were housed in clear, plastic cages (44.5 × 25.8 × 21.7 cm) containing beta-chip bedding. Rats were at post-natal day (PND) 21 on arrival (n = 50). Cages were kept in the Animal Care Facility (ACF) under reverse 12 h light/dark conditions (lights off at 8 a.m.), at a temperature of approximately 21◦C. Rats were handled daily and enrichment was provided by the addition of shredded paper to the animals' cages. Food and water was provided *ad libitum* and the rats were housed in pairs until PND 58, at which point they were food restricted (as described below) and housed individually for the remainder of the experiment. The rats were treated in accordance with the guidelines of the Canadian Council on Animal Care and as approved by the Concordia University Animal Research Ethics Committee.

Apparatus. Behavioural training and testing took place in operant conditioning chambers (12.5″ × 13.5″ × 13.5″; Med Associates, Georgia, VT) placed within ventilated, sound-attenuating compartments. Each chamber was equipped with a modular food pellet dispenser and a food pellet receptacle centered between two retractable levers (Coulbourn Instruments, Whitehall, PA). A continuous infrared photobeam was horizontally mounted across the entrance of the pellet receptacle to detect nose-pokes into the receptacle. Each chamber contained a house light located at the rear of the chamber and three cue lights, one above each lever and one located above the receptacle. Responses on either lever activated the food pellet dispenser, which delivered food pellets into the receptacle. Equipment was interfaced to a computer for experimental programming and data collection using MED-PC software. Rats were placed in the chambers at approximately the same time every day, during the dark phase of the light-dark cycle, and were returned to the ACF upon completion of the sessions.

Methylphenidate Treatment. Rats at PND 25 were randomly assigned to receive one of four different doses of MPH (1, 2, 4, or 8 mg/kg) or 0.9% isotonic saline (1 ml/kg) intraperitoneally for 10 consecutive days. This period of MPH administration spanning shortly after weaning to approximately PND 35 in rats is termed early adolescence (Spear, 2000). The animals were moved from the ACF to the laboratory for the injections once daily at 11 AM

Food Restriction. The rats' daily food intake was restricted to about 14 g starting 23 days after the last injection and until the end of the experiment. They were fed a combination of 45-mg chocolate-flavored sucrose pellets (Bio-Serv, Frenchtown, NJ) and standard rat chow (Harlan Laboratories, IN, USA). Rats consumed chocolate pellets during the experimental task and rat chow 2 h after task completion. The exact weight of rat chow provided was adjusted daily based on body weight and the number of pellets consumed during the task, so that each rat's weight was maintained at about 80–90% of its original weight prior to food restriction.

Lever-Press Training. Lever-press training began 26 days after the last injection, at which point the rats were 58 days old. Neurobehavioural characteristics and developmental changes typical to adolescents can be seen until PND 55 in male rats, and it is recommended that PND 60 be used as a generous estimate to mark the onset of adulthood (Spear, 2000). Here, rats were trained to perform lever responses for sucrose pellets on a fixed-ratio 1 schedule for reinforcement. Each session began with the random extension of one of the two levers, and illumination of the cue light associated with the extended lever. Responses on the lever resulted in the simultaneous retraction of the lever, extinguishing of the cue light above the lever, illumination of the cue light located above the pellet receptacle, and delivery of a food pellet. Each subsequent lever extension during the session was random so that rats had approximately equal exposure to both levers. The criterion for training was set at a minimum of 60 lever responses in 1 h.

Nose-Poke Training. Once rats reached the lever-press training criterion, they were trained to nose-poke in the pellet receptacle to trigger lever presentations. This ensured that the rats were positioned centrally between the two levers at the start of each trial. Trials began with the illumination of the house light and receptacle cue light. With each successful nose poke, the receptacle cue light extinguished, one random lever cue light illuminated, and its associated lever extended. Responding on the lever initiated the simultaneous retraction of the lever, extinction of the house light and lever cue light, activation of the receptacle cue light, and delivery of a food pellet. The trial ended when the rat poked his nose in the pellet receptacle, causing all lights in the box to turn off for an inter-trial interval of 15 s. After two 2 h sessions, all rats moved on to the delay discounting task.

Delay Discounting Task. The delay discounting task was modeled after Evenden and Ryan (1996). The task consisted of a discrete-trials choice procedure in which one lever was paired with the immediate delivery of one food pellet, and the second lever was paired with the delivery of four food pellets presented either immediately or after a delay. The lever corresponding to the larger outcome was consistently paired with a cue light, while the lever corresponding to the smaller outcome had no uniquely associated stimulus. Once the lever was pressed the cue light extinguished. To prevent rats from associating a particular lever with reward or delay, lever-outcome pairing changed at random across blocks of trials. Each training session consisted of 6 blocks of 14 trials each, with the first block taken as a practice round. Each block began with a pair of forced-choice trials, where the levers were extended one at a time, so that rats had no choice between outcomes. These initial trials allowed rats to learn the lever-outcome pairings for that block while breaking any stereotypy in lever choices. Next, rats underwent 12 free-choice trials, where both levers were presented simultaneously, so that rats could choose between the two outcomes. The inter-trial interval corrected for the delay of the chosen outcome, such that the beginning of one trial and the next were always 73 s apart. Rats were first trained without delay for six identical blocks. Once the animals showed almost exclusive choice of the

large reward, a delay was introduced before its delivery, increasing in length with each block (0.1, 4, 10, 25, and 63 s). Lever responses in each block were checked daily until stable behavior was observed for 10 consecutive days. The minimum period of training for this phase was set at 21 days. Only the last 5 days of stable behavior were used for analysis.

2.3.2. Experiment 2

Using PASS (v.15) and the large effect sizes obtained from Experiment 1, a sample size calculation indicated that for a desired power of 0.9 and including inflation for potential dropouts, a lower sample size per group would be sufficient ($N = 5$). Thus, to test the long-term effects of MPH exposure in adults on delay discounting, 20 rats (PND = 58 on arrival) were used. Since no effect of 1 mg/kg of MPH was observed in adolescent rats, adult rats were randomly assigned to receive one of three different doses of the drug (2, 4, or 8 mg/kg) or 0.9% isotonic saline (1 ml/kg). Drug administration began at PND 60. As in Experiment 1, lever-press training began 26 days after the last injection. All other procedures were the same as in Experiment 1.

2.4. Statistical Analysis

For all analyses, lever choices within a session were quantified as the number of choices on the large reward-lever divided by the total number of lever choices. This yields the ratio V, known as the discounted value of the delayed reward. Behavioural stabilization during training was assessed using intraclass correlations. A rat's lever choices were considered stable when the intraclass correlation over 5 days exceeded 0.75. This value was obtained by considering the typical reliability scores in delay discounting procedures and leaving room for fluctuations in choice behavior. In Experiment 1, 10 rats failed to reach a stable performance and were removed from the study; three were removed in Experiment 2. The final analysis for Experiment 1 was based on 40 rats divided into the five injection conditions: 1 ($n = 10$), 2 ($n = 9$), 4 ($n = 8$), and 8 ($n = 7$) mg/kg of MPH, and vehicle ($n = 6$). For Experiment 2, the final analysis included 17 rats divided into 4 injection conditions: 2 ($n = 4$), 4 ($n = 5$), and 8 ($n = 4$) mg/kg of MPH, and vehicle ($n = 4$).

Using MATLAB (Mathworks, R2012b), lever choice averages for each rat across the 5 days were fit to the delay discounting equation, to estimate the rate at which the value of the delayed reward was discounted:

$$V = \frac{1}{1 + kd^b}$$

where V is the discounted value of the delayed reward (obtained from the lever choice averages), d is the delay until reward delivery, b is the discounting exponent reflecting the shape of the curve and k is the discounting rate (Rachlin, 2006).

Results from the Levene's test demonstrated a violation of the homogeneity of variance assumption for the derived k parameter. Thus, all significance testing was carried out on natural

log-transformed k-values, which showed homogeneity of variance. Data were analyzed using a between-subjects ANOVA in SPSS (IBM SPSS Statistics, version 20). The cut-off point for statistically significant results was set at $\alpha = 0.05$. The between subjects variable was the drug dose with five factors in Experiment 1 (0, 1, 2, 4, and 8 mg/kg MPH during adolescence) and four factors in Experiment 2 (0, 2, 4, and 8 mg/kg MPH during adulthood). The magnitude of the effect was calculated using partial η^2. Cohen's d effect sizes were calculated to compare whether each group exposed to MPH was meaningfully different from the control group.

2.5. Results

2.5.1. Experiment 1

Choices of the large reward at increasing delays for each MPH-pretreated group are compared to the control group in Figures 2.1 A-D. The figures also illustrate the fit of these data points to the delay discounting equation. Of the two free parameters in the delay discounting equation, only the discounting rate *k* was significantly different across groups. Specifically, different doses of MPH given intraperitoneally during adolescence had a statistically significant effect on how quickly delay decreased reward value ($F_{(4, 35)} = 2.73$, $p = .045$, $\eta^2 = .24$; see Fig. 1E). Cohen's d effect sizes were calculated to compare the discounting rate of groups that received each dose of MPH with the group of rats that received saline during adolescence. Rats that received 1 mg/kg of MPH and 8 mg/kg of MPH showed an effect that was small in magnitude ($d = 0.13$ and $d = 0.33$, respectively). On the other hand, rats that had received moderate doses of MPH (2 and 4 mg/kg) showed a robust effect with higher discounting rates compared to the saline group ($d = 0.98$ and $d = 1.00$, respectively). Rats showed an increased discounting rate compared to their vehicle counterparts, long after exposure to the drug had discontinued.

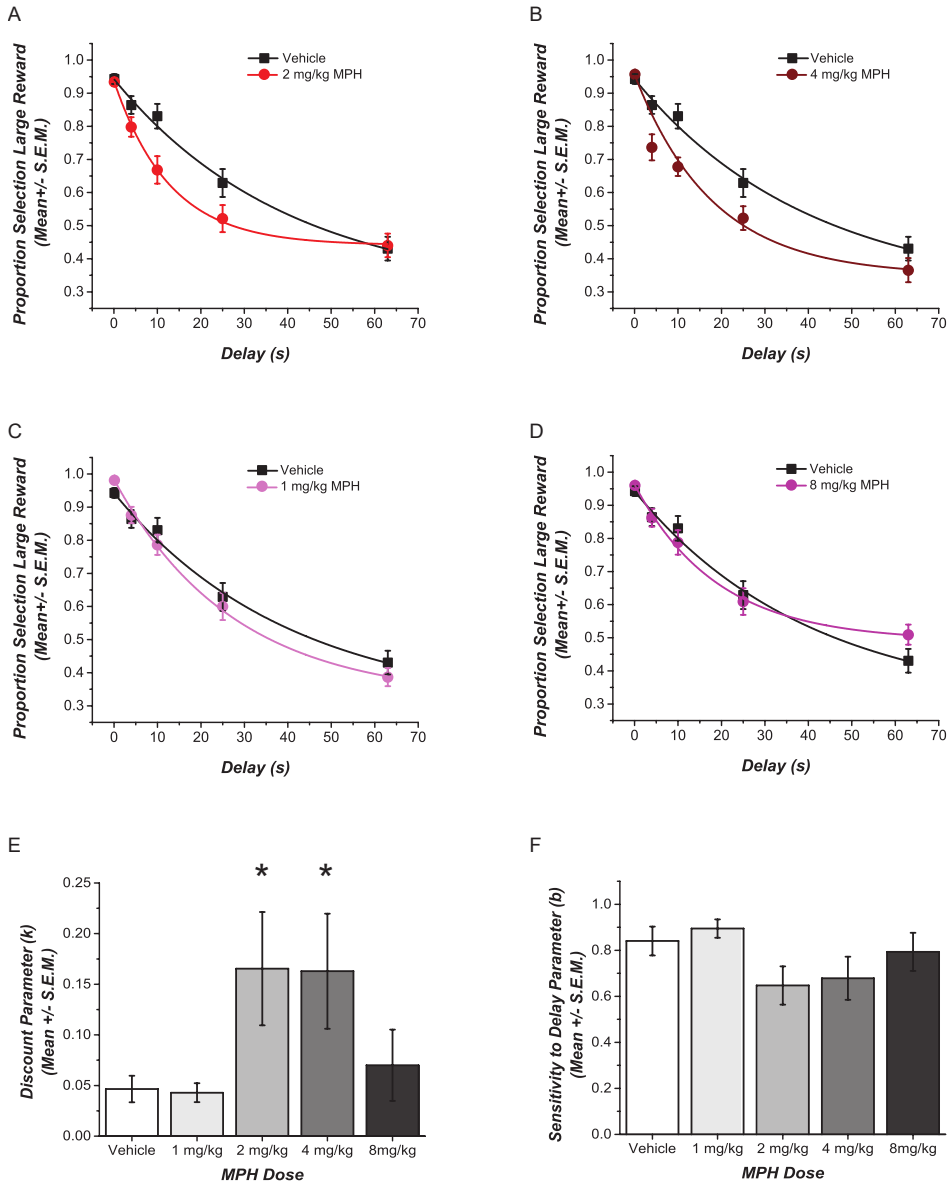
Figure 2.1F shows that no difference was observed for the other free parameter, the discounting exponent *b* ($F_{(4, 35)} = 1.88$, $p = .13$, $\eta^2 = .08$). There were no statistically significant differences across the groups in the number of sessions until stable performance was reached ($F_{(4, 35)} = 1.54$, $p = .21$, $\eta^2 = .15$), and there were no differences across the groups between the choice of the larger reward compared to the smaller reward when there was no delay to the larger reward ($F_{(4, 35)} = 1.66$, $p = .18$, $\eta^2 = .16$). Finally, there were no statistically significant differences across groups in the weight of the rats at the time of testing ($F_{(4, 35)} = 2.81$, $p = .88$, $\eta^2 = .03$).

2.5.2. Experiment 2

Choices of the large reward at increasing delays for each MPH-pretreated group are compared to the control group in Figures 2.2A-C. The figures also illustrate the fit of these data points to the delay discounting equation. Unlike those rats that were exposed to MPH during adolescence, rats that were treated with the drug during adulthood were resilient to the different doses of MPH ($F_{(3, 13)} = 0.30$, $p = .83$, $\eta^2 = .06$). Figure 2D shows no statistically significant differences in the discounting rate *k.* Rats that received higher doses of 4 mg/kg of MPH and 8

Figure 2. 1.

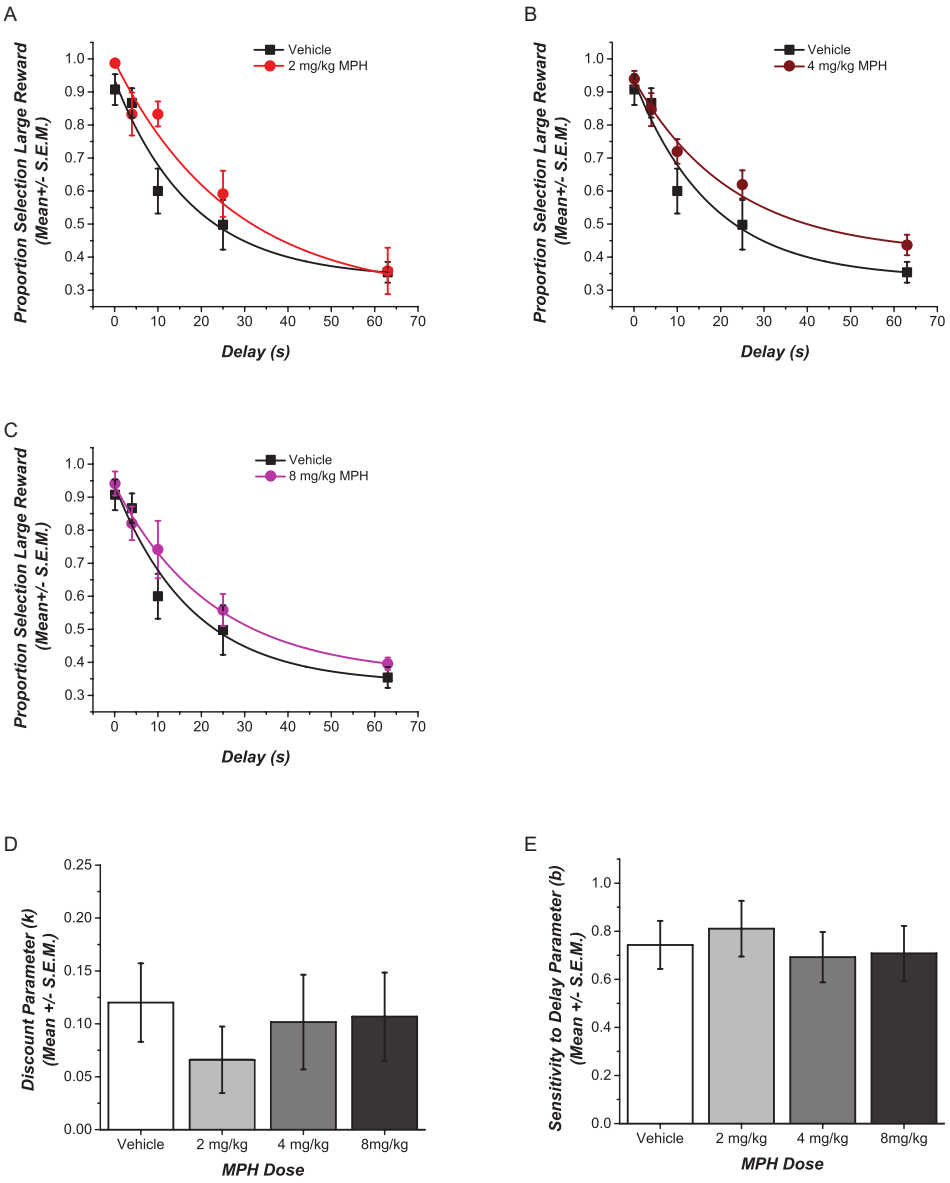
Effect of Adolescent Pre-treatment with MPH on Delay Discounting



Note. (A–D) Data for each MPH-pretreated group is plotted against the vehicle to depict the dose dependent differences in choice for the delayed larger reward. Data points show the mean (± SEM) proportional choice of the large reward. The curve shows the fit of Rachlin's power function to those points. (E, F) Magnitudes of the shifts in the two free parameters of the function are contrasted in the bar graphs. A statistically significant main effect of dose was observed ($p <$ 0.05) for the log transformed discounting rate *k*. *Cohen's *d* effect sizes compared to the saline group were .98 and 1 SD for groups that received 2 and 4 mg/kg respectively. No statistically significant difference was observed in the discounting exponent b. Error bars represent SEM.

Figure 2. 2.

Effect of Adult Pre-treatment with MPH on Delay Discounting



Note. (A–C) Data for each MPH-pretreated group is plotted against the vehicle to depict the dose dependent differences in choice for the delayed larger reward. Data points show the mean (± SEM) proportional choice of the large reward. The curve shows the fit of Rachlin's power function to those points. (D, E) Magnitudes of the shifts in the two free parameters of the function are contrasted in the bar graphs. No statistically significant main effect of dose was observed for the different groups in their discounting rate *k* or in the discounting exponent *b* ($p > 0.05$).

mg/kg of MPH showed effects that were small in magnitude (d = 0.20 and d = 0.16, respectively). Rats that had received the lower dose of 2 mg/kg MPH showed a relatively greater effect with a lower discounting rate, on average, compared to the saline group (d = 0.74). The decrease in the discounting rate indicates more self-controlled behaviour, although this was not statistically significant.

As in Experiment 1, Figure 2.2E shows that rats across the groups had similar discounting exponents b ($F_{(3, 13)} = 0.24$, $p = 0.86$, $\eta^2 = .02$). There were no statistically significant differences across the groups in the number of sessions until stable performance was reached ($F_{(3, 13)} = 0.21$, $p = 0.21$, $\eta^2 = .05$), and there were no differences across the groups between the choice of the larger reward compared to the smaller reward when there was no delay to the larger reward ($F_{(3, 13)} = 1.70$, $p = .22$, $\eta^2 = .28$). Finally, there were no statistically significant differences across groups in the weight of the rats at the time of testing ($F_{(3, 13)} = 1.19$, $p = .35$, $\eta^2 = .22$).

2.6. Discussion

We have demonstrated that MPH exposure during early adolescence can persistently speed the devaluation of reward by delay. Rats that were exposed to moderate doses of MPH during adolescence showed a quicker devaluation of reward with increasing delays compared to rats that did not receive the drug. Importantly, this effect was observed long after the cessation of drug treatment.

The persistent effects of MPH observed here contrast with previously published work (Adriani et al., 2007; Leo et al., 2009; Pardey et al., 2012), a discrepancy that may be explained by methodological differences. Our methodology was modeled after one of the main paradigms used in the literature on delay discounting. In doing so, we introduced safeguards to reduce the risk of selection bias and carryover effects arising from the methodologies used previously. Specifically, the delay to the larger reward was varied within-session; the operant responses required to make a choice changed spatial locations randomly across blocks of trials; a center nose-poke ensured that the animal was equidistant from both levers before making a choice; and finally, rats were trained on the paradigm extensively until their behavior was stable.

Using this paradigm, we found that MPH accelerated the rate of reward devaluation with delay. The speed at which rats learned the task and their preference for the larger reward when it was not delayed did not differ with their drug history. Accelerated devaluation may then stem from one or more of the following: a nonlinear change in the valuation of reward such that the difference between one and four pellets was no longer the same; an increase in the perception of elapsed time; or a change in the process through which reward value and elapsed time are combined. At present, there is evidence that adolescent exposure to MPH results in long-term desensitization to natural rewards (Bolaños et al., 2003; Carlezon et al., 2003) and to drugs of abuse (Andersen et al., 2002; Crowley et al., 2014; Vendruscolo et al., 2008), suggesting that a change in reward valuation is responsible. However, the enduring effects of adolescent exposure

to MPH on time perception, timed performance and its combination with reward value also deserve to be evaluated in future studies.

This study also provides experimental evidence of a nonlinear effect of MPH dose on impulsivity. Whereas doses of 2 and 4 mg/kg of MPH increased impulsivity, a lower (1 mg/kg) and a higher (8 mg/kg) dose showed no effect on the behavior. Nonlinear effects at these doses are also visible in the prefrontal cortex, an area known to play an important role in delay discounting (Winstanley, 2010). Doses of MPH shown effective in this study also increase firing rates in the prefrontal cortex (Devilbiss & Berridge, 2008), facilitate long-term potentiation in vivo (Burgos et al., 2015) and increase glutamate signaling and surface expression of several subtypes of N-methyl-D-aspartate receptors (Cheng et al., 2014). In contrast, higher doses of MPH (greater than 5 mg/kg) have the opposite effects in all four cases (Burgos et al., 2015; Cheng et al., 2014; Devilbiss & Berridge, 2008). The long-term behavioral effects of MPH we describe may then have a neural basis in the prefrontal cortex.

The impulsivity-promoting effect of MPH is specific to exposure during adolescence, a period of continued neural development in the prefrontal cortex (Giedd et al., 1999; Gogtay et al., 2004; Winstanley, 2010) as well as in the connected mesocortical dopamine system (Benes et al., 1996; Manitt et al., 2011; Naneix et al., 2012; Reynolds et al., 2015). Conspicuously, dopamine connectivity has been linked to the computation of temporal influence in the subjective valuation of reward (Pine et al., 2010; Winstanley, 2010), and MPH exposure during adolescence produces an array of persistent effects on the midbrain dopamine system. Among them, dopamine neural activity in the VTA is decreased (Brandon et al., 2003), as is the density of the dopamine transporter in the striatum. This latter effect may be due to a change in connectivity and morphology of the dopamine axons in the striatum and prefrontal cortex (Moll et al., 2001). Such downregulation in the dopaminergic system is associated with impulsive choice (Hernandez et al., 2014; Kheramin et al., 2004; Zeeb et al., 2010). Notably, the aforementioned changes in connectivity and neurotransmission in the prefrontal cortex and mesocortical dopamine system are not observed when MPH exposure occurs during adulthood (Brandon et al., 2003; Crowley et al., 2014; Moll et al., 2001; Somkuwar et al., 2013; van der Marel et al., 2014). Likewise, our findings show that although adolescent exposure to moderate doses of MPH increased impulsivity, the same did not hold true when exposure occurred during adulthood. In our adult-exposure sample, no effect of the drug was observed and if anything, a trend in the opposite direction was detected where rats that had been exposed to certain doses (2 and 4 mg/kg) seemed to exert more self-control. In light of these correlations, it is possible that the stage-specific effect of MPH on delay discounting is mediated by alterations in the development of the mesocortical dopaminergic pathway.

In sum, adolescence may constitute a critical period in the development of the system underlying delay discounting— a period in which the system is acutely susceptible to

environmental influences such as psychostimulant exposure, enabling lasting changes in impulsive behavior.

Chapter 3
Self-stimulating rats do not combine subjective reward magnitude and subjective reward delay multiplicatively.

The previous chapter showed that drug exposure interacts with the development of the rat brain during adolescence, resulting in persistent changes in the way that delay curbs reward value. It is unclear whether the interacting circuitry is responsible for the computation of the reward magnitude, the computation of time elapsed, or how those two values are integrated. These options are not necessarily mutually exclusive. In fact, it is unclear whether it is possible for them to be mutually exclusive. While it is frequently assumed that reward magnitude and reward delay are computed independently and integrated afterwards to guide behaviour, I show in this chapter that this assumption does not stand experimental testing.

3.1. Abstract

Delay discounting describes the finding that immediate rewards are often more valuable to animals than larger, delayed rewards. A common assumption has been that delay discounting results from a multiplicative interaction between ratios of reward magnitude and ratios of delay. Here, we show that this assumption is not accurate. We measured the time allocated to the pursuit of brain stimulation reward—made available immediately or after a 2- or 4-s delay—as a function of both the strength of the rewarding stimulation and its cost. We analyzed the results within the framework of the reward-mountain model. The model enables us to distinguish between effects of delay resulting from changes in the sensitivity of the directly activated neurons responsible for brain stimulation reward (or their integration) and those due to a change in the output of the integrator as predicted by the assumption of multiplicative combination of all factors associated with reward pursuit. Surprisingly, we found that delay affected both outcomes. Our findings suggest that delay engages more than one part of the brain's reward circuit, and therefore the role of delay in reward valuation is more complex than was previously understood.

3.2. Introduction

"The longer the wait, the sweeter the kiss" is a false platitude. We know that waiting for a reward decreases its value. Animals of all species consistently choose a more immediate reward over a later one and are willing to work less for a reward if they must wait for it. When interpreted literally, the phrase also implies that it is not only the value of the kiss but the intensity of it that changes. That distinction is important. The prevalent assumption amongst psychologists, neuroscientists, and economists has been that waiting for a sugary snack decreases its value; it is not the sweetness that decreases, but how much the snack is worth. Sweetness is only one variable in the calculation of the worth of the reward. In contrast to the strong experimental support that we have for delay decreasing reward value, we have little evidence for

the assumption that delay changes reward value without changing the more basic characteristics of the reward, like the sensitivity to its sweetness.

The way researchers present choices to animals determines the behavioural parameters they can measure and the psychophysical function that they can determine. Consider that an animal is given a choice between two behaviours (e.g., pecking key A or key B), and the animal is placed on a concurrent FR1-FR1 schedule. This means that a single peck on either key will result in its associated reward. Thus, if key A resulted in a small, immediate reward and key B resulted in a larger but delayed reward, the researcher could determine which reward the animal preferred simply by observing which key the animal chose to peck. The researcher could then try different combinations of delays and reward magnitudes to specify functions to which the animal's behaviour could be mapped. In these situations, animals often show an exclusive preference for the worthier outcome (Herrnstein & Loveland, 1975). Consider that now the animal is given a choice between the same two outcomes on a concurrent VI-VI schedule. The animal will only receive rewards if it pecks the key after the assigned interval. The animal cannot predict that interval, so multiple responses are necessary to check whether the interval has elapsed for the animal to receive a reward. The benefit of requiring multiple responses from the animal is that researchers can not only observe which of the two choices the animal prefers but can also determine the degree of preference for that choice. Herrnstein (1961) used this benefit to conduct an extensive examination of animal decision-making with varying rates of rewards (concurrent VI-VI schedules), which led him to pose the matching law. In these situations, animals divide their time between the behaviours in proportion to their relative worth. The matching law was further refined to the generalized matching law which can be represented as

$$\frac{B_1}{B_2} = b\left(\frac{R_1}{R_2}\right)^s = \frac{V_1}{V_2} \tag{3.01}$$

where B_1 and B_2 represent two different behaviours, R_1 and R_2 represent the rates of the reward obtained by the animal for each of those behaviours, b refers to bias, and the exponent s represents the sensitivity of the behavioural measure to variations in the reinforcement measure. V_1 and V_2 represent the values that the animal assigns to each behaviour.

Baum (1979) reviewed over a hundred studies on the matching law and found that Equation 3.01 accounted for 90.5% of the variance in behavioural allocation in those studies. The generalized matching law has also been successfully used to explain and predict various behaviours that seem quite complicated at face value. Behaviour analysts have put the law in practice by, for instance, observing and predicting the distribution of problem behaviours by children with developmental disabilities, or the substitutability of different reinforcers in treatment (Borrero et al., 2010; Reed & Kaplan, 2011).

The concatenated matching law states that different parameters of a reward are combined as follows:

$$\frac{B_1}{B_2} = b\left(\frac{R_1}{R_2}\right)^{s_1}\left(\frac{M_1}{M_2}\right)^{s_2}\left(\frac{I_1}{I_2}\right)^{s_3} = \frac{V_1}{V_2}, \quad (3.02)$$

where *B*s represent rates of behaviour, *R*s represent rates of reinforcement, *M*s represent magnitudes of the reinforcer, *I*s represent immediacies (reciprocals of delays to reinforcement), and *V*s represent values. The subscripts 1 and 2 refer to the two alternate choices of the parameters of the reinforcement and their associated behaviours. The overall bias towards one alternative is represented by b and the exponents s_1, s_2, and s_3 represent the sensitivities of the behavioural measure to the variation in each parameter of the reinforcement.

Given that each parameter scales the effects of the other, each parameter thus affects the output of the function that determines the other parameter. As such, the parameters *R*, *M,* and *I* are assumed to be independent in Equation 3.02. In other words, reward magnitude is computed and then combined with the rate of the reward or the immediacy with which it is delivered. Returning to our initial analogy, according to Equation 3.01, the introduction of a delay does not affect the sweetness of the kiss; it changes how much it is worth. Leon and Gallistel (1998) showed that the ratios of rate and the magnitude of the reward combine in a scalar (multiplicative) manner as predicted in the equation. Procedural difficulties have prevented conclusive determination of whether the same can be said about the way the ratios of delay to a reward and the magnitude of the reward are combined.

Researchers use a procedure called the concurrent-chains procedure to determine how the delay to a reward and the magnitude of the reward are combined while simultaneously maintaining the ability to measure the degree of preference between the two choices. Here, the animal makes choices between two concurrent VI schedules that lead to an FI schedule instead of a reward. The FI schedule serves to impose a delay before reward delivery. Concurrent VI schedules expose the degree of preference for the magnitude of the reward at the end of the VI in a concurrent VI procedure. The researchers argued that, similarly, the concurrent VI schedules could show the degree of preference for the delayed reward at the end of the VI interval in a concurrent-chains procedure. However, it seems that animals do not treat the chain of schedules as separate, and this reasoning thus falls apart. It becomes difficult to isolate from animal behaviour whether changes are due to the price imposed in the initial schedule or the delay in the terminal schedule.

Thus, to test whether delay and reward magnitude combine multiplicatively, would need a procedure that would allow the animal's degree of preference to be measured while isolating the effects of varying the delay to a reward on the magnitude of the reward. The ubiquity of time in all that we do makes this a difficult task. One way to combat the problem would be to mathematically isolate the effect of delay to a reward on the magnitude of a reward. A rich enough understanding of the dimensions of reward value and how they combine allowed Arvanitogiannis and Shizgal (2008) to derive a unidimensional scale of reward intensity as an axis of the multidimensional

construct of reward value. This model has since been refined and is called the *reward-mountain* model (for a full derivation, see Trujillo-Pisanty et al., 2020). They took advantage of the fact that researchers could control the input with which the unidimensional symbol for reward intensity is derived and determine the underlying form of the function that led to that translation (Gallistel et al., 1981; Gallistel & Leon, 1991). The way manipulations may affect reward intensity (before or after it has been computed) could then be observed through whether their introduction shifted or scaled the function. They used the findings by Leon and Gallistel (1998) to model how animals combine the rate of reward and BSR intensity. By doing so, they made observing the shifts or scaling of the reward-growth function computationally accessible to us through the animal's behaviour while simultaneously allowing for the observation of the animal's degree of preference. The reward-mountain model has been repeatedly validated. The model has been used to make predictions and can also be used experimentally to learn about the reward circuitry (Arvanitogiannis & Shizgal, 2008; Breton et al., 2013, 2014; Hernandez et al., 2010, 2012; Solomon et al., 2015; Trujillo-Pisanty et al., 2011).

This model can derive the reward-growth function from animal behaviour that shows degrees of animals' preference, isolating it from the effects of the price imposed to obtain the reward (VI or opportunity cost). In doing so, we can observe whether the introduction of a delay would combine with the reward-growth function in a multiplicative manner. If ratios of delay and reward magnitude combine multiplicatively, keeping with the concatenated generalized matching law, the reward-mountain model would predict that delaying the delivery of BSR would shift the mountain along the price axis (scaling the output of the reward-growth function). In contrast, if the mountain were to shift along the frequency axis, the implication would be that ratios of delay and magnitude do not combine multiplicatively. Instead, the effect of delay occurs prior to the peak detection of reward intensity (shifting the reward-growth function).

3.3. Methods

3.3.1. Subjects

Data presented here was collected from three male Long-Evans rats (Charles River, St. Constant, QC), although we started the study with eight. The rats were housed in clear plastic cages (44.5 x 25.8 x 21.7 cm) containing beta-chip bedding. The rats weighed about 375 g on arrival. Cages were kept in the animal care facility (ACF) under reverse 12-hour light/dark conditions (lights off at 8 a.m.) at a temperature of approximately 21◦C. The rats were handled daily, and enrichment was provided by adding shredded paper to the animals' cages. Food and water were provided ad libitum. The rats were housed in pairs until they underwent surgery, at which point they were housed individually for the remainder of the experiment. The rats were treated in accordance with the guidelines of the Canadian Council on Animal Care and as approved by the Concordia University Animal Research Ethics Committee.

3.3.2. Surgery

In preparation for the surgery, rats weighing between 400–600 g received subcutaneous injections of atropine sulfate (0.05 ml) to reduce bronchial secretions and of ketoprofen (0.15 ml) for its painkilling and anti-anxiolytic effects. They received a dose of ketamine-xylazine (100mg/kg) intraperitoneally to induce anesthesia 10 mins later. The rats' paws were pinched to ensure that the depth of anesthesia was adequate, after which their heads were shaved, and they were mounted into stereotaxic frames. The rats were maintained on isoflurane for the rest of the surgery. Monopolar stimulation electrodes (0.25 mm in diameter, Plastics One, Roanoke, VA, USA) were aimed unilaterally at the lateral hypothalamus: 2.8 mm posterior to bregma, 1.7 mm lateral to the midline and 7.8 mm ventral to the surface of the dura mater. An uninsulated wire was wrapped around three to four stainless-steel jewellers' screws embedded in the frontal and parietal bones to provide the path for current to return. The electrodes were lowered into the brain using standard stereotaxic manipulators and secured with dental acrylic anchored by four or five jeweller's screws. After surgery, the rats received ketoprofen (0.2 ml) to further reduce pain, saline (0.2 ml) for rehydration and were placed on top of a heat source until recovery from anesthesia was complete. The rats were given a one-week period to recover before testing.

3.3.3. Apparatus

Behavioural training and testing took place in operant conditioning chambers (34 cm x 23 cm x 60 cm) with four plexiglass walls, a hinged plexiglass front door and a mesh floor constructed in-house by David Munro. Each chamber was equipped with two retractable levers (1.5 cm x 1.5 cm; ENV–112B, MED Associates, St. Albans, Vermont) located in the center of the right and left walls, 10 cm above the floor, and cue lights located 2 cm above each lever. Only one of the levers and corresponding cue light was used for this experiment. Responses on the lever activated the cue light above it. Each chamber also contained a house light on the back wall, 35 cm above the floor, that would flash to signal the start of a new trial. A mating connector terminated the monopolar stimulation leads (0.2 mm in diameter, Plastics 1, Roanoke, VA, USA), which were attached to a slip-ring assembly to allow the rat to circle the cage without twisting the leads. A second cable linked the slip-ring assembly to the output of the constant-current stimulator. Equipment was interfaced to a computer for experimental programming and data collection using software created in-house by Steve Cabilio (PREF3, Concordia University, Montreal, QC, Canada). Stimulation parameters were monitored on an oscilloscope. The rats were placed in the chambers at approximately the same time every day, during the dark phase of the light-dark cycle and were returned to the ACF upon completion of the sessions.

Unlike the computer-controlled setup described above, each rat was screened for self-stimulation before actual testing via a handset integrated circuit pulse generator. This allowed for quick variation of the currents and frequencies to visually test the rat's optimal stimulation parameters.

3.3.4. Experimental Procedure

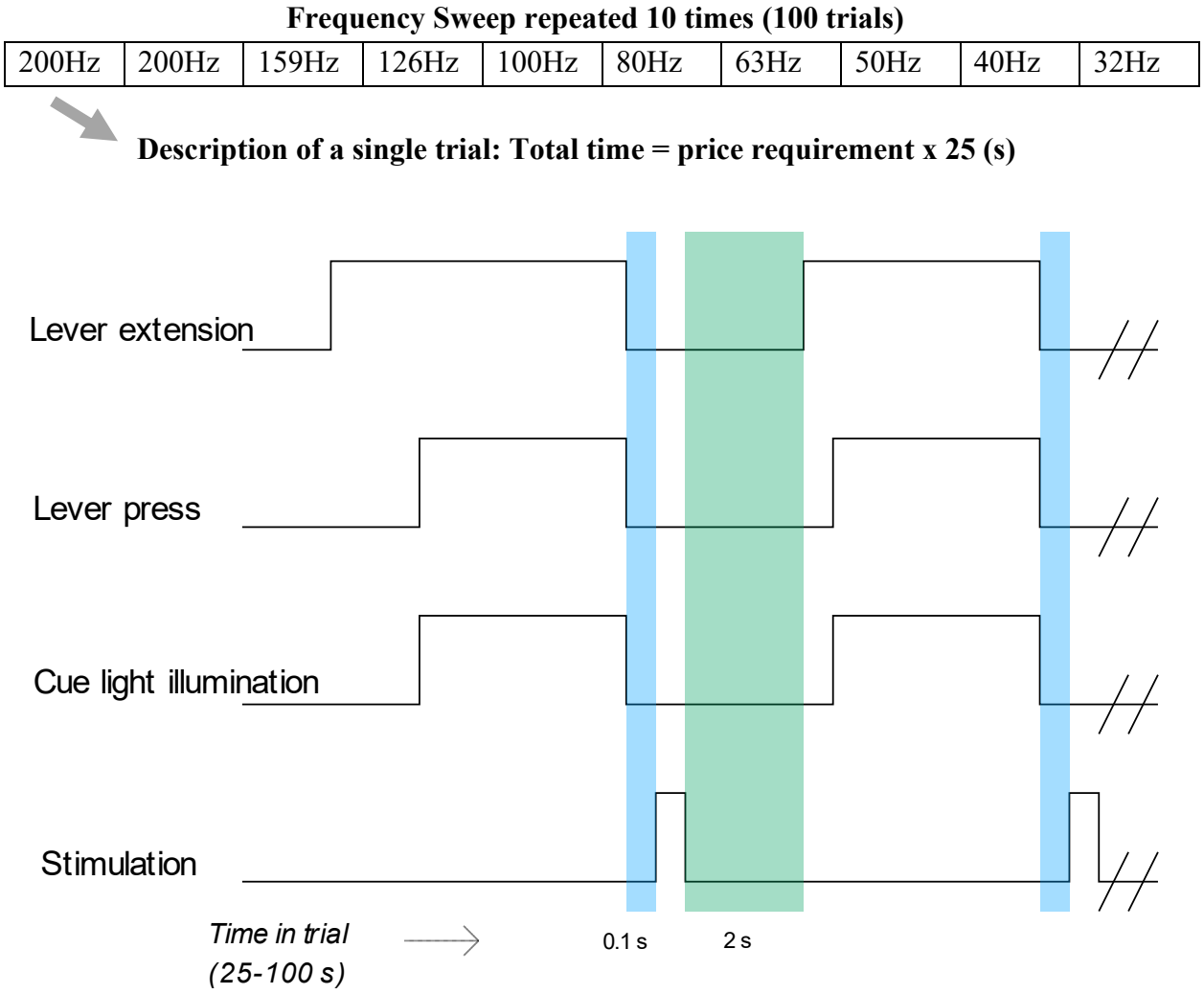
3.3.4.1. Initial Screening. The rats were first tested to determine whether the stimulation electrode was implanted in the correct location. Implantation in the correct location was assumed when rats displayed motivation to obtain a stimulation without significant motor effects at a range of currents and frequencies. Surgery was deemed unsuccessful if rats failed to approach the lever, exhibited signs of aversion, or had significant motor effects. Stimulation consisted of 0.5 s trains of cathodal, constant-current pulses that were square-wave in shape and 0.1 ms in duration. The rats were connected to the stimulator by a cable and could move freely around the operant chamber. Using manually operated stimulators, rats were initially given a low-intensity pulse train of electrical stimulation (at low current intensity and pulse frequency). The rats were trained to press the lever for stimulation using standard shaping techniques. The criterion for training was set at a minimum of 60 lever responses in 10 min. Once that criterion was reached, the electric current intensity and frequencies were gradually increased to determine the parameters that supported the maximal response rate for that rat. These parameters were used to train the rats in the computer-controlled operant chambers, which were used until the end of the experiment.

3.3.4.2. Training on basic experimental structure. The training session began with the extension of the lever on the left side of the chamber. Responses on the lever resulted in the illumination of the cue light associated with that lever. The rats were trained to press a lever at a fixed-cumulative-handling-time (FCHT) schedule for reinforcement (Breton et al., 2009; Hernandez et al., 2010). The lever had to be held down for a specific cumulative amount of experimenter-set time to obtain the stimulation at the electric current intensity and frequency previously determined to support the rats' maximal response rate. The schedule is considered cumulative because the rat can take breaks and return to the lever to finish the time requirement without penalty. The time set for the lever to be held down is considered the price the rat must pay to obtain the reward. The trials lasted varying times depending on the experimenter-set price, such that if rats were to spend all their time pressing the lever for the stimulation, they would be able to obtain 25 rewards in each trial. During this initial training, the price was set at 1 s, and so each trial lasted 25 s. Completing the price resulted in the blackout delay: extinguishing of the cue light above the lever and retraction of the lever for 2 to 4 s (depending on the rat). Rats received the stimulation 0.1 s after the beginning of the blackout delay. The blackout delay did not count in the total trial time. At the end of the blackout delay, the lever once again extended to allow the rat to obtain another reward. Once the trial time ended, the lever retracted, and the house light flashed to signal the beginning of a new trial, and for later stages of the training and experiment, different stimulation frequencies and/or prices. The time spent pressing the lever out of the total time available in a trial was calculated for all trials in a session, yielding the ratio of time allocation (TA) for each trial.

3.3.4.3. Training: Frequency Sweeps. The first phase of the training consisted of a frequency sweep (a descending sequence of experimenter-set frequencies). Figure 3.1 visually

Figure 3. 1.

Schematic of a Frequency Sweep Training Session



Note. Each cell in this figure represents one trial. Pulse frequencies were set in accordance with the rat's behaviour. For one training session the entire sweep of pulse frequencies is repeated 10 times. The price remains constant for a session but was increased across sessions until a sweep at a 4 s price requirement was obtained. The description of a single trial shows the sequence of events that occurs in a trial after the flashing of a house light. The first and second row illustrate the illumination of the cue light and the presentation of the lever, respectively. The blue shaded region depicts the delay before the delivery of the stimulation, assuming that the price requirement has been satisfied. The third row depicts the train of stimulation. The green shaded region illustrates the blackout delay after the train of stimulation has been received and before the cue light is illuminated again and the lever is extended are armed. This cycle continues such that if the animal spent all their time working for the reward, they would be able to get 25 rewards.

depicts the structure of the trials in the frequency sweep. In each frequency sweep, the price remained fixed while the frequencies decreased in logarithmic step size. Each session consisted of 10 frequency sweeps, while each frequency sweep consisted of 10 trials. The first trial served as a warmup trial and was excluded during analysis; it consisted of the highest frequency of stimulation that the rats would receive, equivalent in value to the stimulation in the second trial. After the second trial, the subsequent eight trials consisted of stimulation with decreasing frequencies. Time allocation is known to vary in a sigmoidal fashion as a function of pulse frequency, and we expected the rats to allocate most of their time pressing for the stimulation when it was at its highest. Thus, the range of the pulse frequencies tested were adjusted to fully capture the rat's behaviour across the relevant parameters on the psychometric curve and to be able to compute an accurate position parameter for the obtained curve. We aimed to get time allocation values at both asymptotes and on the slope of the psychometric curve.

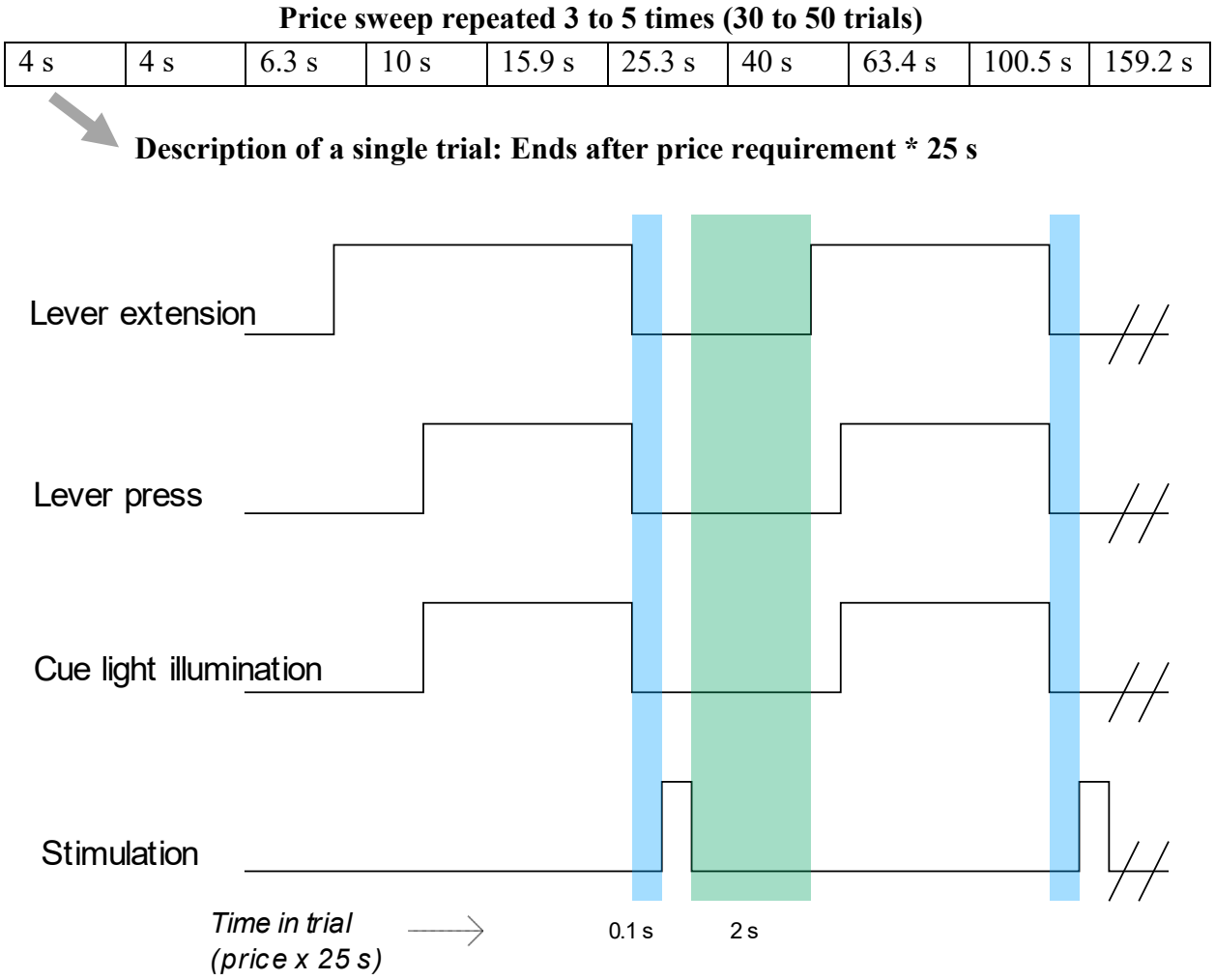
While the intensity of the electric current and the price required to obtain it within any one session of the frequency sweep training remained constant, the price was increased across sessions. The price was set to 1 s at first, and increased by 1 s increments until a final price of 4 s was reached. The price was increased after data points were obtained for each part of the frequency sweep (the asymptotes and the slope of the curve), and the rat spent maximal time allocation of at least 0.8 s. The next part of the training was introduced only once these criteria were reached.

3.3.4.4. Training: Price Sweeps. Once a frequency sweep at a price of 4 s that fully captured a rat's behaviour was obtained, training for price sweeps commenced. A price sweep consisted of trials where the current and frequency of the stimulation remained the same (the maximal parameters determined for that rat), but the price required to obtain that stimulation decreased across trials in logarithmic steps. Each session consisted of three to five sweeps. This varied depending on the rat, as with higher prices the sessions would run too long for the rat's comfort.

The rest of the experimental structure mirrored that during the frequency sweep training. Each sweep consisted of 10 trials, where the first trial was omitted during analysis and consisted of the same price requirement as the second trial. Price requirements were increased in logarithmic steps for the remaining eight trials. Time allocation also varied in a sigmoidal shape as a function of price. Thus, our criteria for a price sweep were the same as that for a frequency sweep. Behaviour was obtained from all three parts of the psychometric curve, the upper and lower asymptote, and the slope of the curve. We expected time allocation to be at least 80 percent at the lowest price. Prices were adjusted until these criteria were met to obtain an accurate position parameter for that curve, after which the rats were moved to the next stage of training. Figure 3.2 illustrates the structure of the trials in the price sweep.

Figure 3. 2.

Schematic of a Price Sweep Training Session



Note. Each cell in this figure represents one trial. For one training session the entire sweep of prices is repeated 3 to 5 times depending on the length of the highest price requirement, and therefore overall session, for the rat. The frequency throughout training is selected to be the highest frequency that the rat could handle. The description of a single trial shows the sequence of events that occurs in a trial after the flashing of a house light. The first and second row illustrate the illumination of the cue light and the presentation of the lever, respectively. The blue shaded region depicts the delay before the delivery of the stimulation, assuming that the price requirement has been met. The third row depicts the train of stimulation. The green shaded region illustrates the blackout delay after the train of stimulation has been received and before the cue light is illuminated again and the lever is extended are armed. This cycle continues such that if the animal spent all their time working for the reward, they would be able to get 25 rewards.

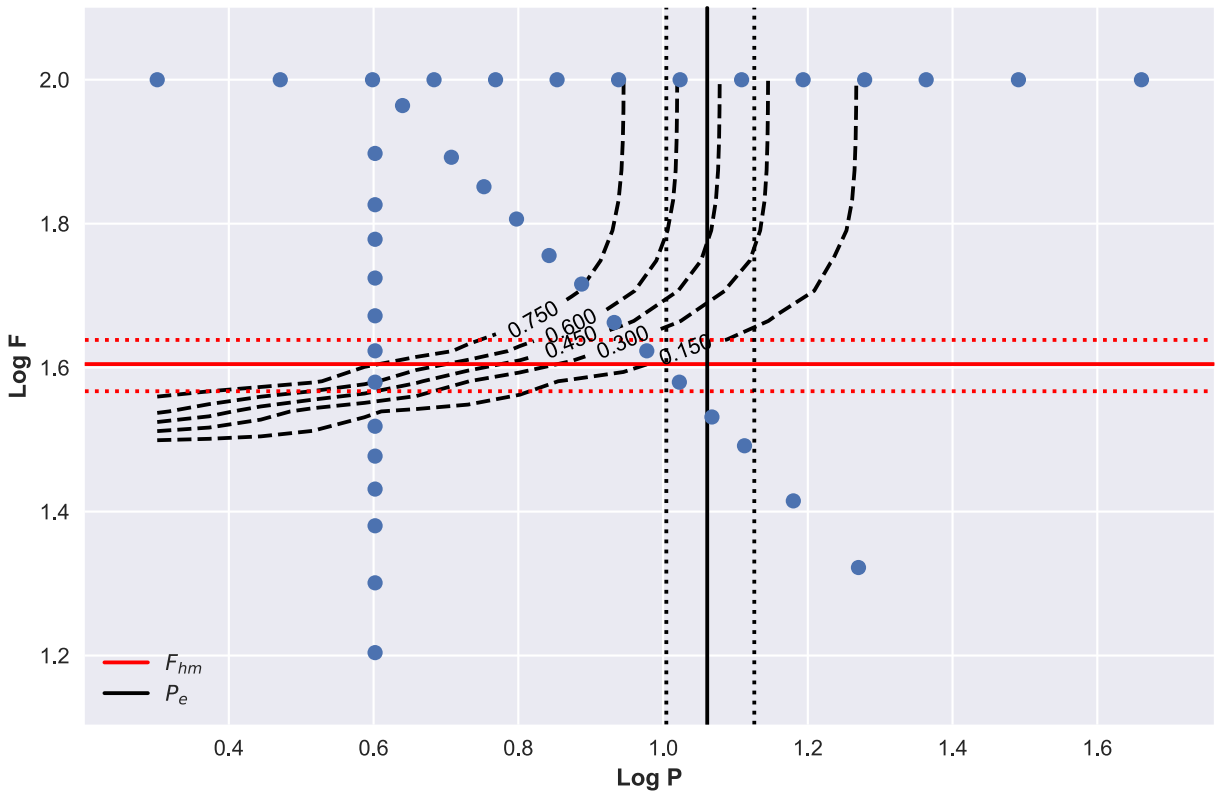
3.3.4.5. Training: Mountain.

Radial sweep. The 3-D mountain consists of the frequency sweep and price sweep described above and an additional radial sweep. Each individual trial was largely the same as that depicted in Figures 3.1 and 3.2, with some key changes. Instead of 10 frequency and price points as those used in the training, each sweep consisted of 14 points. The middle ten points were spaced equally in logarithmic steps, whereas the remaining 4 points meant to be on the extremes of the psychometric curves were twice that interval. In the radial sweep, while the intensity of the electric current remained the same, both the frequency of the stimulation and price requirement were varied in logarithmic steps. The range of the experimental parameters (pulse frequency and price) for the radial sweep was established using data obtained from training on the frequency sweep and price sweep through a simulator developed by Yannick Breton in MATLAB (The Mathworks, Natick, MA). The parameters for frequency and price were chosen so that the trajectory of the radial sweep passed close to or through the intersection of the fitted position parameters, F_{hm} and P_e (see Figure 3.3). This ensured that sufficient data would be collected to capture the full range of the rat's behaviour around the relevant parameter values, thus, increasing the likelihood of obtaining an accurate fit of the reward-mountain model.

Randomization and bracketing trials. Once all parameters for the frequency sweep at a 4 s price, the price sweep at a high frequency and the radial sweep (14 points each) had been determined, the 42 parameter combinations were randomly selected using a random number generator without replacement (Python, v. 32) to make up a session with 42 test trials. This was done so that the rat would have no knowledge of the stimulation frequency, or the price requirement in a trial without first sampling it. It seems that rats need only one encounter with the stimulation to learn the values of the reward parameters in a given trial and that they work quickly to obtain that information (Breton, 2013). For the remainder of the experiment, the ratio, TA, was consequently amended to be the time spent pressing the lever out of the total time of the trial, starting from the delivery of the first reward. It has also been observed that some rats lift their paws from the lever for a second or less only to press again. These taps seem to be involuntary and present the same opportunity cost during the brief release time as when rats are working for the stimulation since no alternate activities can be engaged in during that time (Breton et al., 2009). Consequently, any releases of the lever that were a second or less were part of our final dependent variable, TA. The intensity of the stimulation's electric current remained the same throughout the trials. The structure of the trial also remained the same as in previous training tasks. In each trial, trial time was set such that rats could get a maximum of 25 rewards if they spent their entire time pressing the lever. Cue lights would be illuminated upon responses on a lever, the lever would retract upon completion of the price requirement, and the reward would be delivered 0.1ms later. Flashing of the house light would signal the start of a new trial.

Figure 3. 3.

Positioning of the Radial Sweep



Note. The datapoints superimposed on this contour plot depict the values chosen for the frequency sweep (vertical data points), price sweep (horizontal data points), and the radial sweep (diagonal data points). The black vertical line shows the position of the mountain on the price axis (P_e), whereas the red horizontal line depicts the value of F_{hm}. We altered values of frequency and price for the radial sweep such that it crosses through the intersection of F_{hm} and P_e

Test trials, however, were not presented to the rat consecutively. Each test trial was bracketed between two other trials, one was highly desirable, and the other had low value (Breton et al., 2009). Thus, the total number of trials in the mountain task was 126, 42 test trials bracketed by two other sets of 42 trials. The leading trial was always presented before the test trial and consisted of the highest pulse frequency that the rat would encounter in the experiment, at a price requirement of 1 s. The trailing trial always followed the test trial and consisted of the lowest pulse frequency the rat would encounter in the experiment, with a price requirement of 1 sec. There are known cognitive biases that alter value-based decision-making depending on the reference points that the decision is surrounded by (e.g. Constantinople et al., 2019). Thus, these trials were introduced to provide a constant reference point of stimulation for the rats and allow for more consistent behaviour across the test trials. Once behaviour seemed stable over five days upon visual inspection of the fits, the testing phase of the experiment commenced.

3.3.5. Testing

Once rats were trained on the mountain, delay to the delivery of the reward was introduced to observe the change in position of the mountain on the two axes. This was achieved by adding six additional sweeps to each survey. Frequency, price, and radial sweeps with the same combination of frequency and price parameters as in the mountain condition were presented to rats, but these sweeps differed in that the reward was delivered either 0.1, 2, or 4 s after the price requirement was fulfilled. Blackout delays were changed to 6 s for all test trials, but remained at 2 s for the bracketing trials. The trials in each session consisted of a combination of parameters from the nine resulting sweeps that were randomly selected without replacement. Thus, each survey consisted of 378 trials (126 test trials surrounded by 252 bracketing trials). Since the number of trials exceeded the number of hours in a rat's workday, the survey was divided into three daily sessions of 126 trials each. The data was divided back into three mountains at the end of the survey, and the mountain model was fit to each subset of the data to observe shifts in the position parameters with varying delays to reward delivery.

3.4. Statistical Analysis

TA (as described in Section 3.3.4.5) was calculated for all trials in each session. The final analyses are based on data from three rats R03, R06 and R08. The first four surveys were not considered in the final analysis as they served as training sessions. Until the test session, rats had only experienced immediate delivery of a reward upon completion of the price requirement. Since this would be a new concept for them to learn, and judging by the long learning period before stable behaviour on other tasks (see Section 2.3), it was decided that training trials were necessary. After a visual inspection of the data, disorderly responses when the delay was first introduced and fits of the reward-mountain equation that failed to fit to the data, four training sessions were chosen. This was first decided with R03, and the criterion of four training sessions was maintained across all rats throughout the study. Thus, analyses are based on test data starting from Survey 5.

3.4.1. Model Fitting and Selection

For each rat, the following equation, as described by Breton et al., (2013) and Solomon et al., (2015) was fit to the TA data:

$$TA = \left[(TA_{max} - TA_{min}) \times \frac{\left(\frac{F^g}{F^g + F_{hm}^g}\right)^a}{\left(\frac{F^g}{F^g + F_{hm}^g}\right)^a + \left(\frac{P}{P_e}\right)^a} \right] + TA_{min} \quad (3.03)$$

where

a = the payoff sensitivity exponent

g = the reward-growth exponent

F_{hm} = the induced frequency of firing that produces a rewarding effect of half-maximal intensity

P_e = the subjective price at which time allocation for a maximally intense reward falls hallway between TA_{max} and TA_{min}

F = induced frequency of firing in the directly stimulated neurons and is given by

$$F = F_{bend} \times \left[\ln\left(1 + e^{\frac{F_\rho}{F_{bend}}}\right) - \ln\left(1 + e^{\frac{F_\rho - Frequency}{F_{bend}}}\right) \right] \quad (3.04)$$

where

F_{bend} = parameter governing the abruptness of the transition between the rising and flat segments of the function, and was set to 21 pulses per s (pps)

F_{rho} = the midpoint of the transitional region, and was set to 362 pps

Frequency = the pulse frequency

P = the subjective price of the stimulation train and it given by:

$$P = P_{min} + P_{bend} \times \ln\left(1 + e^{\frac{Price - P_{min}}{P_{bend}}}\right) \quad (3.05)$$

where

P_{min} = minimum subjective price and was set to be equal to 1.82 s

P_{bend} = parameter determining the abruptness of the transition between P_{min} and the rising portion of the subjective price function and was set to be equal to 0.5 s

Price = the price of the stimulation train

Parameters were set to specific values based on results from previous research (Breton et al., 2013; Solomon et al., 2015). Trujillo-Pisanty et al., 2020 included an additional term for the normalized maximum reward intensity which scales the subjective-price ratio (P/P_e). However, the term was introduced to account for cases where the frequency gets too elevated, and it becomes difficult to drive reward intensity to its maximum value. Because our F_{hm} values were low, we could treat that factor as being equal to 1.

The nonlinear least-squares method was used to fit equation 3.05 to the three sets of mountains in each survey in order to estimate the two position parameters (F_{hm} and P_e). Fits were obtained for each survey rather than to an averaged time allocation across surveys to defend the location-parameter estimates against the bias introduced by within-condition shifts of the mountain (Hernandez et al., 2010). While we had six variable parameters (TA_{max}, TA_{min}, a, F_{hm}, g, and P_e), the fits of the reward-mountain equation failed to converge for data from some sessions. How well a parameter's value can be estimated decreases with an increase in the number of parameters that are fit to a dataset. Therefore, common values of some parameters were forced. Five different models were run, each with different forced parameters (a combination of TA_{max}, a, g, and TA_{min}), while the two position parameters were, of course, always allowed to vary. The Akaike information criterion (AIC) was then used to identify the best-fitting model for each rat. The AIC assigns values to the amount of variation in the data explained by the model penalized by the number of parameters being used to explain the information; more parameters provide a heavier penalty. A smaller AIC reflects a better fit of a model. That is, the model explains more variation using the fewest possible independent variables. AIC was calculated as follows,

$$\mathrm{AIC} = -2.\log_{10}\left(\frac{SS_{res}}{length}\right) + 2k\left(\frac{length}{length - k - 1}\right) \tag{3.06}$$

where, SS_{res} = Sum of squares of the residuals

length = length of dataset, and:

$$k = n_f(n_s + n_c) \tag{3.07}$$

where, n_f = number of free parameters

n_s = number of surveys

n_c = number of common parameters)

The model that best fit the data was chosen using a combination of AIC weights and a visual analysis of the fits when AIC weights for two models were significantly similar.

3.4.2. Bootstrapping

To avoid making unnecessary and potentially incorrect assumptions about the datasets, the bootstrapping method was used to obtain the variance around our position parameters across the test surveys. The data were resampled by survey: for the analysis of the first 10 surveys, 1000 datasets were resampled, each dataset consisting of 10 surveys. For example, one dataset might consist of data from surveys [5,7,7,7,8,10,10,12,14,14], and another might consist of data from [5,6, 8,9,10,11,12,12,13,14]. Inferential statistics and graphs presented here were based on the surfaces defined by the Tukey mean parameter estimates and 95% confidence intervals, derived from the resampling procedure for the model and fitting method deemed best by the AIC values. Graphs were plotted using Python (*v. 3.7*).

3.5. Results

3.5.1. Preliminary Analysis

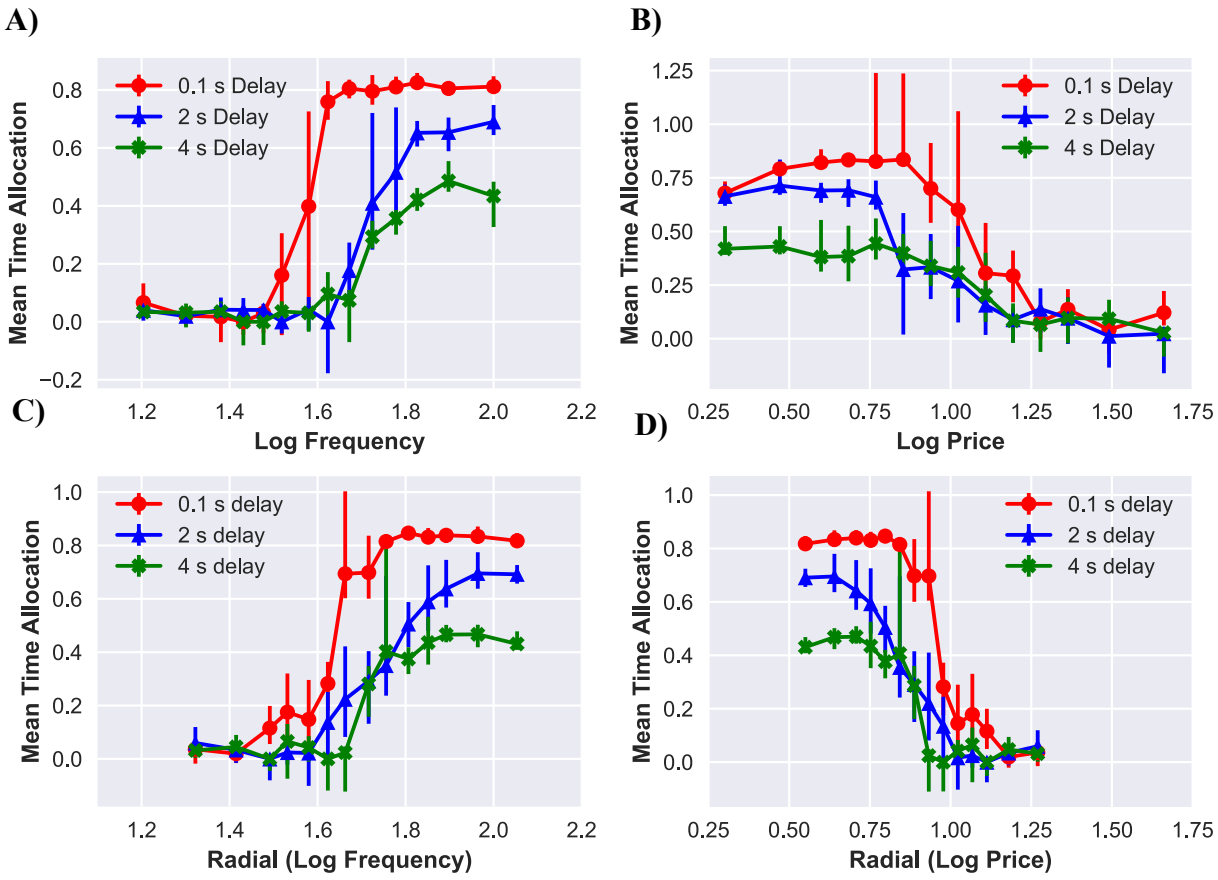
Out of eight rats that were used in the experiment, six rats were successfully implanted with electrodes in their medial forebrain bundle (MFB) at the level of the lateral hypothalamus (LH). Implantation was considered successful when rats showed a high level of pursuit of electrical stimulation with little or no motor effects in response to a wide range of intensities of the stimulation and showed no obvious aversive responses. Two of the six rats had to be removed mid-experiment as they developed severe motor effects in response to the stimulation, and one rat died from unknown causes. Data from the remaining three rats is summarized here.

Raw data across test sessions was averaged for each delay condition and is displayed for a sample rat, R03, in Figure 3.4. When the reward was delivered immediately after the price requirement was met, all three rats spent more than 80% of their time working for the stimulation at their response asymptotes which consisted of the highest frequencies and the lowest costs. On the other end when the stimulation was at its lowest frequency and the price requirement was high, rats spent less than 20% of their time working for the stimulation. An exception was seen at very low costs of 0.25 log units (less than 0.25 s), where a slight dip below 80% of time allocation is observed for R03. Given that time allocation reliably goes above 80% of time allocation at higher prices, the dip in time allocation is unlikely to result from the value of the opportunity cost and may be due to a combination of the high-intensity reward and the motor effects associated with it. Analysis conducted after removing that price did not markedly change the results reported here.

Figure 3.4 shows, through a two-dimensional lens, that the proportion of their time the rats allocated to working for the stimulation increased with increasing frequencies, decreasing opportunity costs, and decreasing delays to reward delivery. Time-allocation curves obtained through frequency sweeps seem to shift as the delay to reward delivery is increased, such that a higher frequency is needed for the rat to allocate the same proportion of time to the pursuit of the reward. Time-allocation curves obtained through price sweeps also shift as a function of delay to reward delivery, such that the rat works less for the same price requirement when a reward is

Figure 3. 4.

A 2-D Perspective on the Effect of Delay on Time Allocation



Notes. The figure shows the effect of delayed reward delivery on time allocation as a function of reward strengths and costs. Time-allocation data presented here was resampled 1000 times to obtain Tukey means and middle 95th percentile. Data points show the Tukey mean proportion of time allocated to working for the reward by a sample rat (R03) in response to varying frequencies of BSR and opportunity costs; frequency sweeps (A), price sweeps (B) and radial sweeps (C and D). Data points in red (circles) and their corresponding curves show time allocation values during trials in which the reward was delivered immediately upon completion of the price requirement. The blue data points (triangles) show the shift in time allocation when the delay to the delivery of a reward was increased by 2 s. The green data points (crosses) and curves show a further shift when the delay was increased by another 2 s (4 s from baseline).

delivered after a delay. The same result is obtained when both frequency and price are varied together in the radial sweeps. Rats allocate less time to working for the stimulation as a function of decreasing frequencies and increasing opportunity costs. The figure shows that the rat allocates less time to the pursuit of stimulation as the delay to the delivery of the stimulation is increased. The curves seem to shift rightwards along the frequency axis and leftwards along the price axis.

As discussed earlier, the decrease in time allocation induced by delaying reward delivery could result from the effect of delay on several variables that affect the overall value of the reward. A similar delay-induced shift of the time-allocation curve obtained from frequency sweeps would be observed whether delay changed the sensitivity to the stimulation, the cost associated with the stimulation, the value of other alternate activities, or whether it combined with the output of the reward growth function in a scalar manner. In other words, shifts on the frequency axis when viewed through a two-dimensional perspective as in Figure 3.4 would arise from whether delay affected the input or the output of the reward-growth function, an effect that cannot be disambiguated through that two-dimensional analysis. Instead, that information must be computationally extracted from the rats' reward seeking behaviour. The reward-mountain model was fit to the data obtained from each rat to computationally extract the processing stage affected by delay and, thus, determine whether delay combines with reward intensity in a scalar manner.

3.5.2. Model fitting and selection

Equation 3.03 describes the reward-mountain model and consists of six parameters that are allowed to vary during the fitting process:

- F_{hm} = pulse frequency at which reward intensity as seen in the reward-growth function is midway between the minimum and maximum intensity values
- P_e = price at which time allocation is midway between the rat's maximum and minimum values
- TA_{max} = maximum time allocated to pursuit of the stimulation
- TA_{min} = minimum time allocated to pursuit of the stimulation
- a = exponent for price sensitivity that sets the slope of the mountain surface along both the frequency and price axes
- g = reward-growth exponent that sets the slope of the mountain along the frequency axis

When all six parameters were left free to vary, the model failed to converge with data from some sessions. Thus, to limit the flexibility of the fitting, and because delay is not expected to change more than the position parameters, five different variants of the reward-mountain model, that differed only in the parameters left free to vary, were fit to the data. Specifically, the following parameters were left free to vary in each variant of the model,

- **Variant 1:** All six parameters; TA_{max}, TA_{min}, a, g, F_{hm}, P_e

- **Variant 2:** a, g, F_{hm}, P_e
- **Variant 3:** a, F_{hm}, P_e
- **Variant 4:** g, F_{hm}, P_e
- **Variant 5:** Only F_{hm} and P_e

Sessions that failed to converge when a variant of the model was fit to the data were removed from the analysis for that variant. The resulting fits from the remaining sessions were resampled 1,000 times with n = 10 to obtain estimates of parameter values. The estimates obtained from analysis of Variant 1 of the model were selected as parameter values to force for each rat in Variants 2–5 of the model. For most sessions, fits for Variants 3 and 5 converged, in stark contrast to fits for Variants 1, 2 and 4 of the reward-mountain model (see Table 3.1). AIC values were then calculated to determine the explanatory power of each variant of the model after imposing a penalty for each additional parameter that could vary. Lower AIC values signal a better model according to those criteria. Fitting Variants 3 and 5 of the reward-mountain model to the data produced a similar proportion of test sessions that converged. Variant 5 converged to data from all sessions and Variant 3 converged to data from nine out of 10 sessions. However, Variant 3 had the lowest AIC value, and Variant 5 had the highest AIC value. For all rats, Akaike values for the models where either a or g were left free to vary (Variants 3 and 4) were lower than when both or neither could vary (Variants 1, 2 and 5). Tables 3.2, 3.3 and 3.4 show AIC values for each variant of the model for R03, R06, and R08, respectively. Results were consistent across all variants of the model in all rats. Among Variants 3 and 4, AIC values for the baseline condition were similar, but Variant 3 had consistently lower AIC values in all rats for the test conditions. More sessions failed to converge for Variant 4, indicating that that the model did not fit the data well for those sessions. A visual analysis of the fits also showed consistently better fits for Variant 3 across all rats (see Figure 3.5). Results suggest that along with changing the position of the mountain, delay to reward delivery also led to a marked change in the slope of the mountain along the frequency and price axes (parameter a). Variant 3 of the reward-mountain model fit time-allocation data in response to the varying parameters well. Adjusted R^2 values ranged from .96 to .97 in the 0.1 s-delay condition, .95 to .97 in the 2 s delay condition and equalled .97 in the 4 s delay condition. Table 3.5 details adjusted R^2, and root mean square error (RMSE) values obtained by fits of Variant 3 of the reward-mountain model as a function of delay for each rat. Thus, all further analyses were conducted using the variant of the reward-mountain model with three free parameters – a and the two positions parameters F_{hm} and P_e. No significant systematic trends were observed in residual correlations.

Table 3. 1.

Proportion of Test Sessions That Failed to Converge for a Sample Rat, R03

Model's free parameters	Baseline	2 s Delay	4 s Delay
F_{hm}, P_e, TA_{max}, TA_{min}, *a, g*	0.5	0.5	1
F_{hm}, P_e, *a*, *g*	0.6	0.5	0.6
F_{hm}, P_e, *a*	0.1	0	0
F_{hm}, P_e, *g*	0.1	0.4	0.3
F_{hm}, P_e	0	0	0

Note. Highlighted row represents the variant of the model chosen to fit to data.

Table 3. 2.

AIC Values for R03 for the Five Variants of the Reward-Mountain Model

Model's free parameters	Baseline	2 s Delay	4 s Delay
F_{hm}, P_e, TA_{max}, TA_{min}, *a, g*	-172.17	-172.20	-171.37
F_{hm}, P_e, *a*, *g*	-311.59	-311.76	-311.45
F_{hm}, P_e, *a*	-851.81	-851.95	-851.59
F_{hm}, P_e, *g*	-851.87	-312.18	-313.31
F_{hm}, P_e	552.00	551.67	550.67

Note. Highlighted row represents the variant of the model chosen to fit to data.

Table 3. 3.

AIC Values for R08 for the Five Variants of the Reward-Mountain Model

Model's free parameters	Baseline	2 s Delay	4 s Delay
F_{hm}, P_e, TA_{max}, TA_{min}, *a, g*	-167.36	-168.33	-167.07
F_{hm}, P_e, *a*, *g*	-299.60	-301.72	-301.31
F_{hm}, P_e, *a*	-799.15	-804.67	-804.25
F_{hm}, P_e, *g*	-799.07	-302.43	-303.06
F_{hm}, P_e	497.07	741.87	741.31

Note. Highlighted row represents the variant of the model chosen to fit to data.

Table 3. 4.

AIC Values for R06 for the Five Variants of the Reward-Mountain Model

Model's free parameters	Baseline	2 s Delay	4 s Delay
F_{hm}, P_e, TA_{max}, TA_{min}, *a, g*	-172.16	-172.90	-172.13
F_{hm}, P_e, *a*, *g*	-312.02	-312.45	-311.87
F_{hm}, P_e, *a*	-852.44	-852.77	-851.97
F_{hm}, P_e, *g*	-852.23	-312.73	-312.97
F_{hm}, P_e	551.31	550.80	550.91

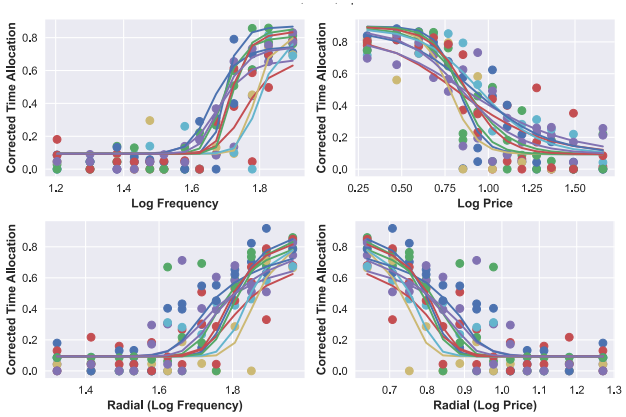
Note. Highlighted row represents the variant of the model chosen to fit to data

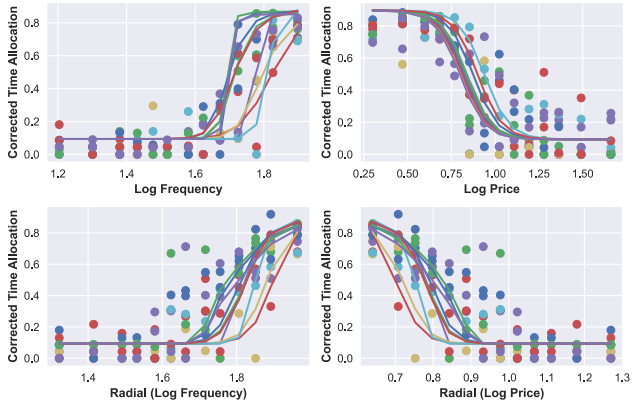
Figure 3. 5.

Comparison Between Fits of Variant 3 and Variant 4

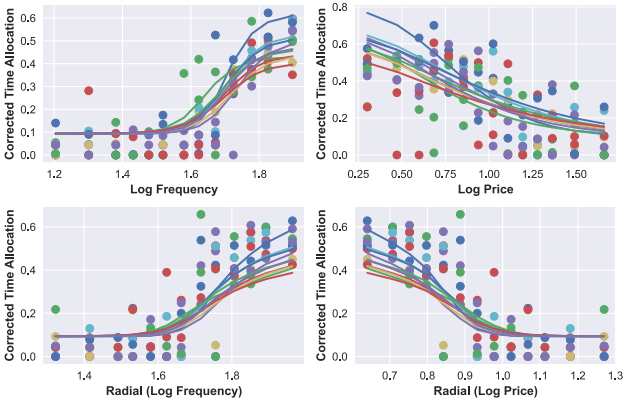
Variant 3 **Variant 4**

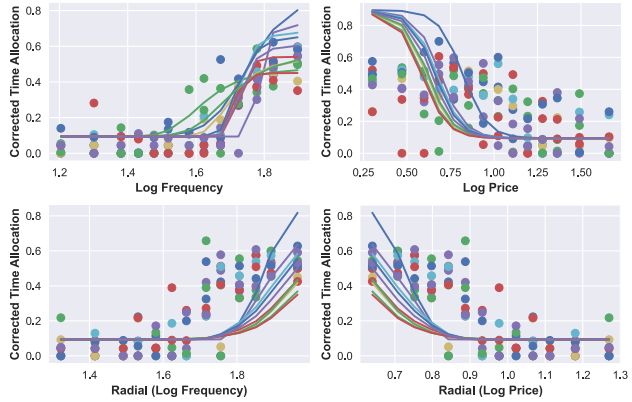
2 s delay





4 s delay





Note. The figure shows behavioural data and two-dimensional projections of the fitted surface of the reward-mountain along the pulse-frequency, price, and radial sweeps for two variants of the model. The data depicted here was obtained from R03 across ten test sessions. Each curve represents the fit to the data from one session. The two graphs on the left show fits for each session to the variant of the reward-mountain model in which the parameters *a*, F_{hm}, and P_e could vary (Variant 3). The two graphs on the right, in contrast, show fits for the variant of the reward-mountain model where the value for *a* was forced, but *g* could vary (Variant 4). The top two graphs use data from the condition where reward was delivered 2 s after the price requirement was met. The bottom two graphs use data from the condition where reward was delivered 4 s after the price requirement was met. It is especially clear in the 4 s delay that Variant 3 of the reward-mountain model fits the data better.

Table 3. 5.

R^2 and RMSE Values for Fits of Variant 3 of the Reward-Mountain Model

Rat	Baseline		2s Delay		4s Delay	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
R03	.96	.14	.97	.14	.97	.12
R06	.96	.15	.95	.17	.97	.14
R08	.97	.13	.97	.14	.97	.12

3.5.3. Changes in position of the mountain across conditions

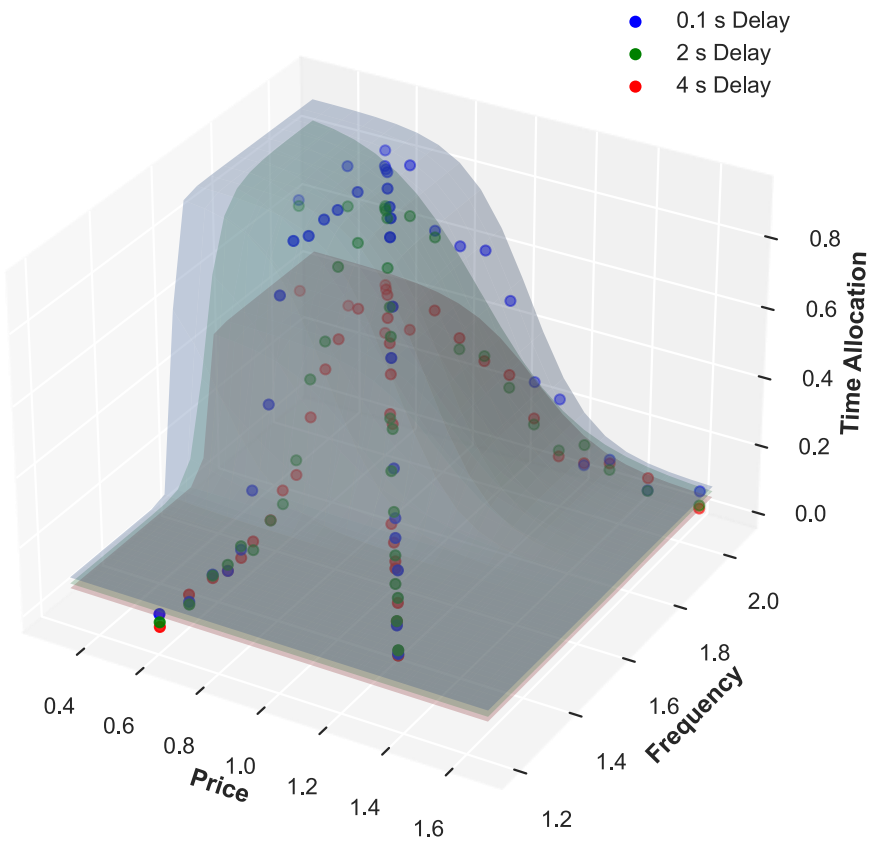
Figure 3.6 shows an example of the time-allocation data and the resulting surface plot upon fitting the reward-mountain model to data from all three test conditions. Figure 3.7 depicts the contour plots of those fits for each rat, making the shifts of the mountain along each axis clearer. Figure 3.8 illustrates the difference between fitted position parameters of the mountain when the reward was delayed and the baseline condition in the form of bar graphs. Shifts in the position of the mountain on each axis induced by the delay to reward delivery are illustrated here for each rat. Table 3.6 details the values of these differences in log units along with their confidence intervals.

The mountain shifted diagonally for all three rats when the reward was delayed by 4 s; this shift was statistically reliable. While a statistically reliable diagonal shift was clearly observed for R03 when the reward was delivered with a 2 s delay, that was not the case for R06 or R08. For R06, when the reward was delayed by 2 s, a statistically reliable shift on the price axis but not the frequency axis was observed. For R08, a statistically reliable shift on the frequency axis was observed, but the variance in the shifts on the price axis when the reward was delayed by 2 s was quite large and, while statistically reliable, the lower percentile lay very close to zero. Compared to the condition in which the reward was delivered immediately, the frequency that produced half-maximal intensity was 1.179 times higher, averaged across all rats, when the reward was delayed by 2 s and 1.283 times higher, averaged across all rats, when the reward was delayed by 4 s. Compared to the condition in which the reward was delivered immediately, the price at which rats allocate half of their time to pursue the reward was 1.705 times higher, on average, when the reward was delayed by 2 s and 3.434 times higher, on average, when the reward was delayed by 4 s.

3.5.4. Behavioural stability and training

Each trial in a session was bracketed by a leading trial with a very appealing reward of high frequency and low cost and a trailing trial with an unappealing reward of low frequency and low cost. The test trials in between were randomly chosen from selected values in frequency sweeps, price sweeps and radial sweeps from each of the three conditions such that the animal could not predict the parameters of the stimulation without receiving a reward in that trial. After the first reward was delivered, the animal would have full information about the parameters: the frequency of stimulation, the delay in receiving it, and the price to obtain that stimulation. The structure here allowed us to measure whether the animal had been sufficiently trained on the task by observing how quickly the animal started to allocate its time to working for the stimulation during the bracketing trials. Figure 3.9 shows that each animal could predict whether the trial was a leading or a trailing trial from the first test session when delay was introduced. As shown in the figure, R03 consistently started to work for the reward within the first second during leading trials. This occurred before any information about the reward was made available during that trial. While the difference between test trials and trailing trials was not as stark as the

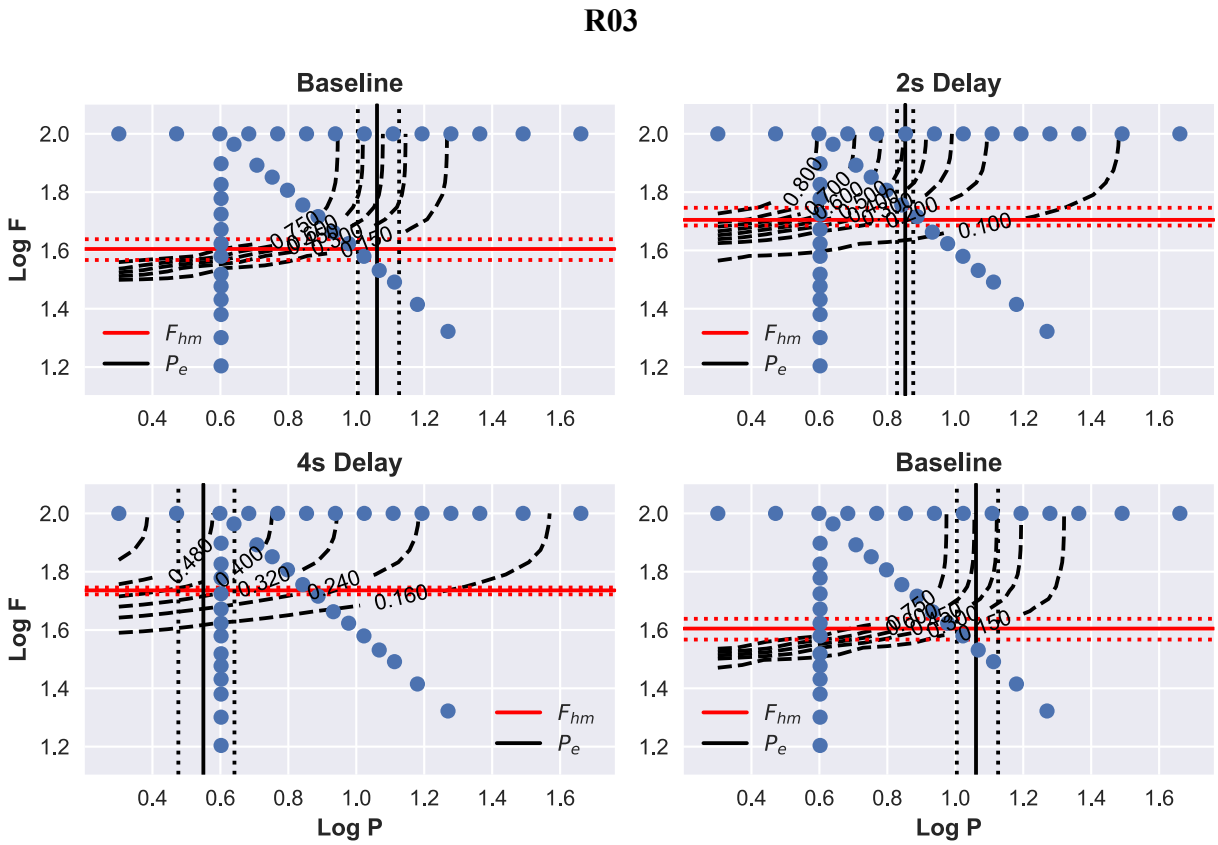
Figure 3. 6.
Fitted Surface Plots at Different Delays



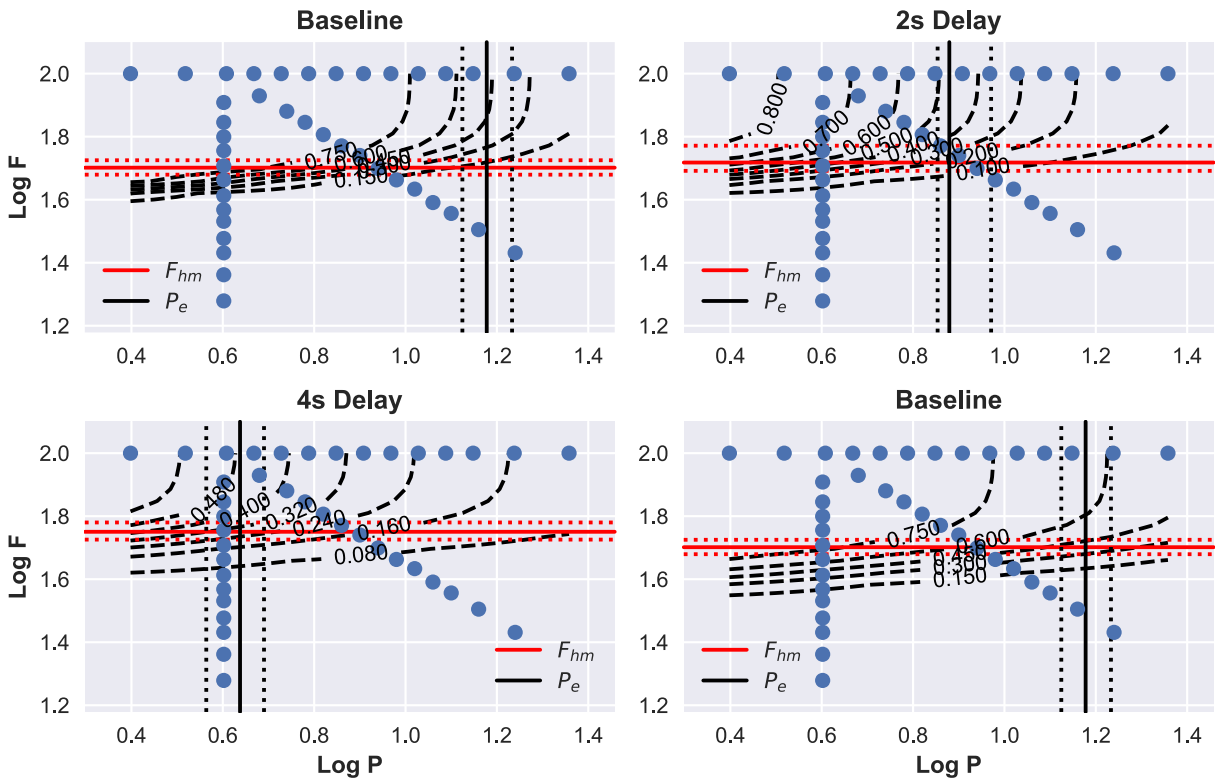
Note. Three fitted surfaces of the reward mountain model and the corresponding average time allocation values at the pulse-frequency, price, and radial sweeps for R03 are depicted here. The first surface uses data obtained from the baseline condition in which reward is delivered immediately (light blue surface, and dark blue data points). The other two surface plots use data from the test conditions in which the reward is either delivered 2 s after completion of the price requirement (light green surface, green data points) or is delivered 4 s after completion of the price requirement (light red/grey surface, red data points). The mountain shifts diagonally inwards as delay to reward delivery is increased, such that both the reward-growth function shifts rightwards, and the price function shifts leftwards.

Figure 3. 7.

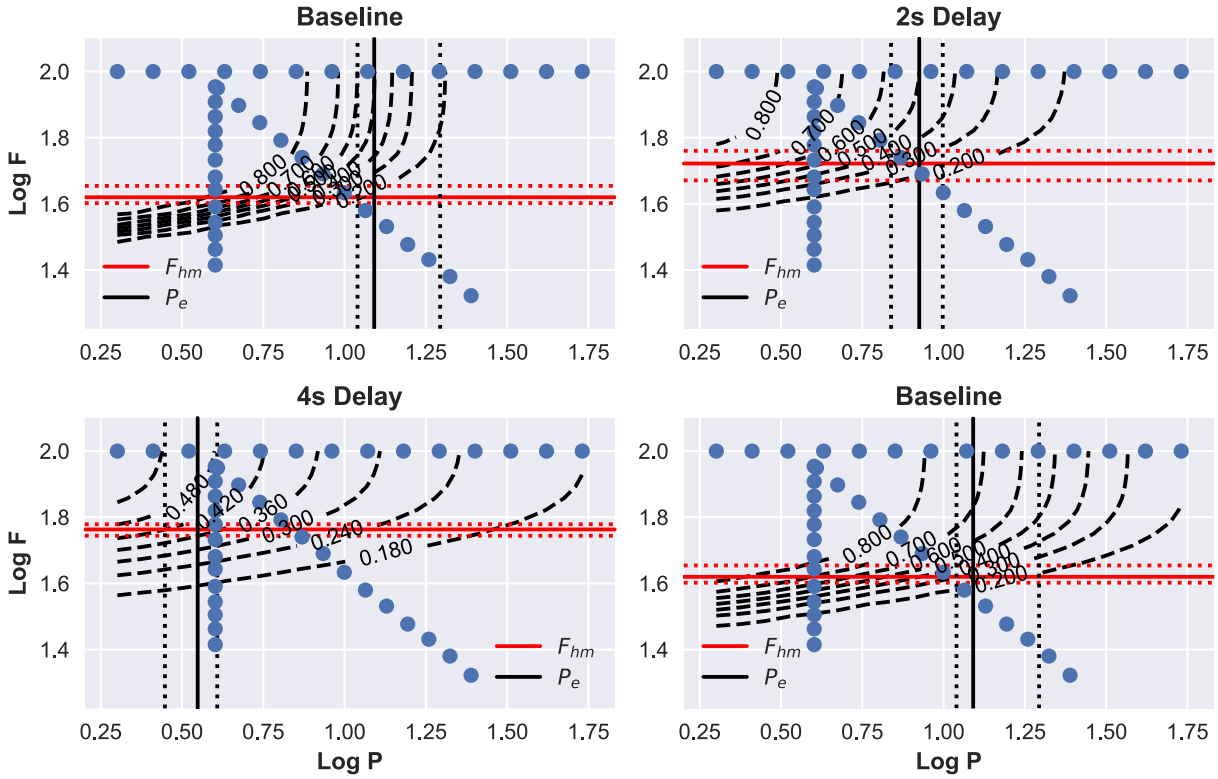
Contour Plots of the Reward-Mountain Surfaces for Each Rat



R06


Baseline
2s Delay
4s Delay
Baseline
Log F
Log P
F_{hm}
P_e
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.150
0.480
0.400
0.320
0.240
0.160
0.080
0.750
0.600
0.450
0.300
0.150

R08

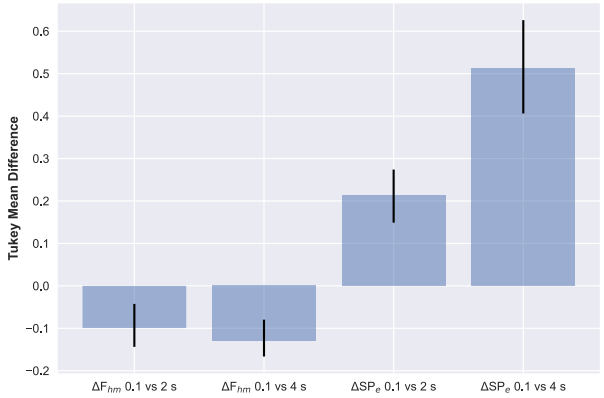

Baseline
2s Delay
4s Delay
Baseline
Log F
Log P
F_{hm}
P_e
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.480
0.420
0.360
0.300
0.240
0.180

Note. This figure shows contour plots of the reward-mountain surfaces fit to TA values obtained from each delay condition for each rat. Sets of graphs constructed from data obtained from rats R03, R06 and R08 represent the cross-sections of a horizontally sliced reward-mountain. In each set of graphs, the baseline condition where the reward was delivered immediately after the price requirement was met is represented in both the upper left and bottom right quadrants to compare the delay-induced shift in the position parameters more easily. The horizontal red line represents the position parameter F_{hm} along the pulse-frequency axis; the pulse frequency is halfway between the maximum and minimum reward intensities. The vertical black line represents the position parameter P_e along the price axis; the price is halfway between the maximum and minimum opportunity costs. Dotted lines around both position parameters represent 95% confidence intervals. The blue circles represent the levels of the independent variables (frequency and price) along the three sweeps on which the rats were tested.

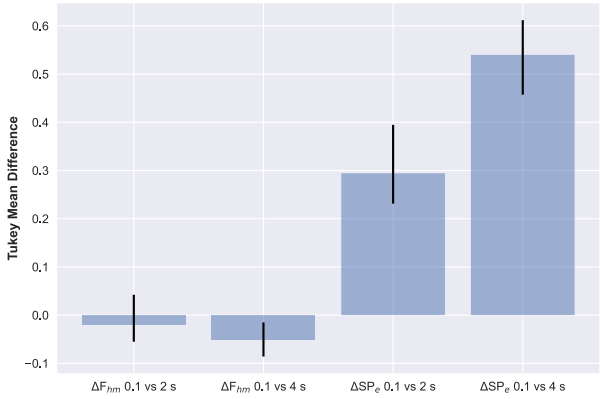
Figure 3. 8.

Delay-Induced Shift in the Position Parameters for Each Rat

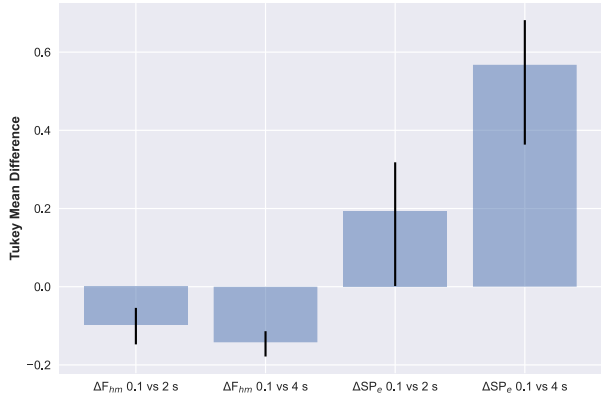
R03



R06



R08



Note. The bar graphs contrast the delay-induced shift in the two position parameters of the reward-mountain. Each graph is constructed from data obtained from one of the rats R03, R06, and R08. Error bars represent the middle 95th percentile of the bootstrapped estimates of the position parameters from the ten test sessions. A statistically reliable shift of the mountain along each axis is assumed when zero lies outside of the confidence intervals of the mean difference between position parameters.

Table 3.6.

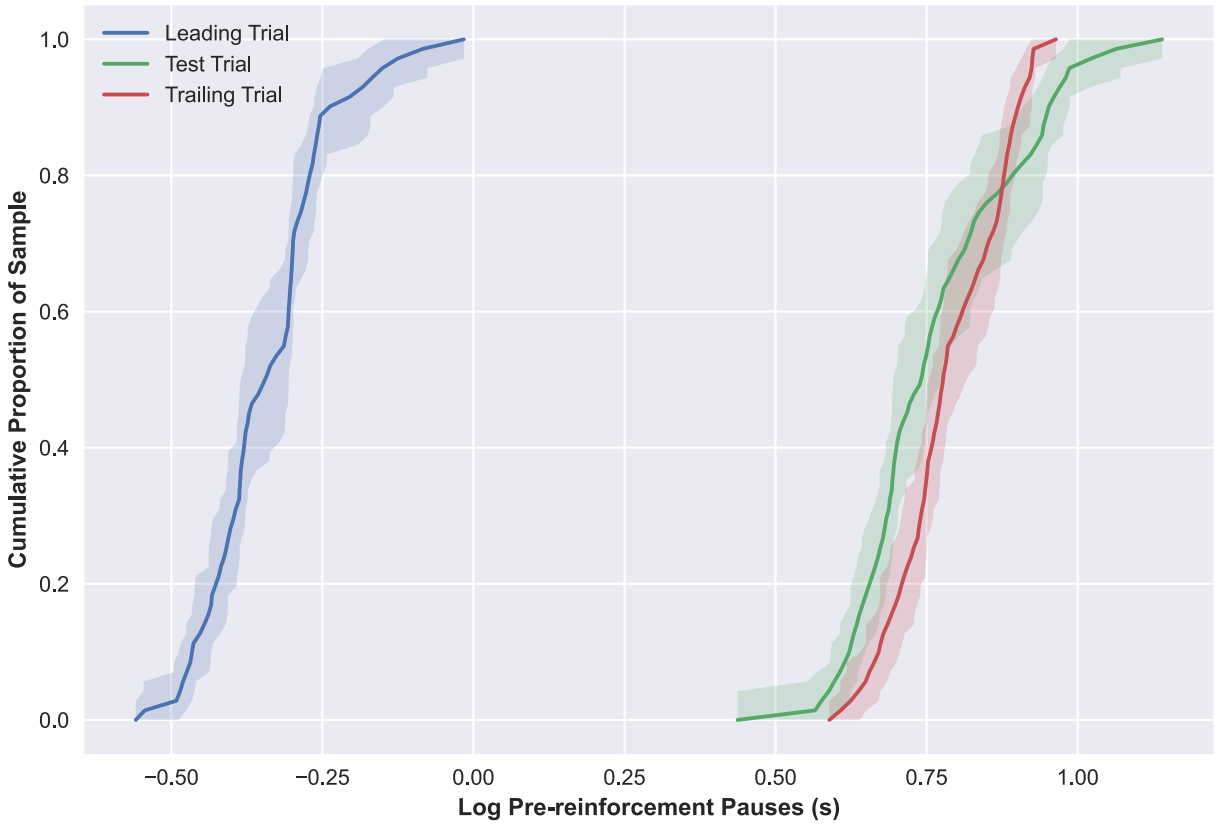
Tukey Mean Differences in Position Parameters of the Reward-Mountain after a 2-s and a 4-s Delay in Reward Delivery.

Parameter	R03			R06			R08		
Condition	CI_{low}	Shift	CI_{high}	CI_{low}	Shift	CI_{high}	CI_{low}	Shift	CI_{high}
$\log_{10} F_{hm}$ 0.1s vs 2s	-0.153	-0.099*	-0.058	-0.089	-0.018	0.017	-0.143	-0.098*	-0.039
$\log_{10} F_{hm}$ 0.1s vs 4s	-0.183	-0.131*	-0.094	-0.086	-0.050*	-0.019	-0.170	-0.144*	-0.109
$\log_{10} P_e$ 0.1s vs 2s	0.145	0.209*	0.284	0.198	0.293*	0.353	0.057	0.193*	0.383
$\log_{10} P_e$ 0.1s vs 4s	0.400	0.511*	0.614	0.463	0.540*	0.621	0.434	0.557*	0.743

** Middle 95th percentile of bootstrapped parameter estimates do not include zero.*

Figure 3. 9.

Pre-reinforcement Pauses on Test and Bracketing Trials



Note. The x-axis represents the length of pauses at the beginning of trials before R03 started to work for a reward in log units. The y-axis represents the cumulative proportion of samples across sessions. Data is divided by session, where three sessions make up one survey. The first fifteen sessions are included here, including the first four training trials. Given that no objective information was provided at this point during the trial, it seemed that the rat had learned the structure of the session, where sampling was no longer required for the rat to consistently anticipate the strong reward during the leading trials. 95% bootstrapped confidence intervals surround the cumulative proportions.

comparison with leading trials, the rat's behaviour clearly shows that enough training had occurred for the rat to be able to predict the triadic structure of the session. Furthermore, a within subjects ANOVA showed no statistically significant differences from session to session on either of the position parameters for any of the delay conditions (p>.05). Thus, the behaviour of all rats was considered stable during the test sessions.

3.6. Discussion

The self-stimulation phenomenon has allowed researchers to dissect the reward circuitry and quantify animal behaviour at each dissection. By merging these quantifications into a unified framework, the reward-mountain model describes how the direct electrical stimulation of neurons in the MFB is integrated into a symbol of subjective reward magnitude and is later translated into behaviour. The model, at its core, rests on the reward-growth function, which describes the integration of the electrical stimulation in the MFB and stores the resulting magnitude as an engram for further computation. Direct physical access to and knowledge of the function that does that translation has allowed the mathematical observation of whether a manipulation affects the input of the reward-growth function (observed as a horizontal shift on the x-axis) or whether it affects its output (observed as a scaling of all values along the y-axis). Thus, manipulating the input of the function by altering the current or train duration should move the mountain along the frequency axis, and manipulating variables that ought to affect the output of the function, or combine with the output of the function to determine payoff, such as rate of reinforcement or opportunity cost, should move the mountain along the price axis. Since its formation in the 90s, the model has been validated on numerous occasions (Arvanitogiannis & Shizgal, 2008; Breton et al., 2013). Upon successful validation, the reward-mountain model is now being used to explore the reward circuitry by observing what stages of it are affected by various manipulations. For instance, the delivery of cocaine, dopamine transporter blockers and cannabinoid agonists all move the mountain along the price axis (Hernandez et al., 2010, 2012; Trujillo-Pisanty et al., 2011, 2013). The reward-mountain model fits our data very well.

If delay to reward delivery combines with reward magnitude the same way that other costs have shown to, we would expect a shift of the mountain on the price axis and not on the frequency axis. That expectation is consistent with the concatenated matching law, which assumes that ratios of delays would combine with ratios of reward magnitude in a simple scalar manner. The results observed in our study, however, present a different story. While a large shift on the price axis was observed, a large, statistically reliable shift on the frequency axis was also apparent. Introducing a delay to reward delivery for rats in our study seemed to affect both the output of the reward-growth function and its input. Unlike the combination of ratios of reward magnitude and rate of reinforcement, this suggests that the combination of the ratios of reward magnitude and reward delay cannot be described by a simple multiplication. The shift on the price axis indicates that delay affects or combines with the output of the reward growth function to change reward value. The shift on the frequency axis indicates that delay to reward delivery

affects either the signal within the directly stimulated neurons or the way that signal is integrated. Although there may be overlap in how costs (rate, opportunity cost, and delay to reward delivery) are combined with reward magnitude to form value, results from this study show that there is also a divergence; delay to reward delivery affects the reward circuitry at multiple stages.

This result is perplexing when considering that the concept of rate, opportunity costs and delay to reward delivery are all interrelated. After all, the rate of reinforcement, by definition, is the reciprocal of the expected interval between reinforcers. It seems that the animal uses different computational processes for the combination of magnitude and delay experienced in a VI schedule compared to experiences where a reward is delivered at a known interval after the price requirement is met. Two potential differences that may lead to different computational processes come to mind. First, the delay calculated in a VI schedule is an average delay over time, whereas the delay experienced by the animal before reward delivery can be measured immediately after one exposure. Second, and relatedly, the experience of delay during a VI schedule is uncertain. Any level of certainty arises only after multiple exposures and on a molar scale, whereas, in the latter scenario, the animal is certain of the next delay after a single exposure to it. It seems from our results that the latter computation, at least in part, occurs simultaneously with the integration of the stimulation of the neurons in the MFB, whereas the former computation does not.

Our results suggest that there may be more than one mechanism with which delay to a reward is encoded; there certainly seems to be multiple mechanisms with which the two are combined. For the two magnitudes to combine arithmetically, they must share common properties (apples and oranges cannot be multiplied meaningfully). The output of the reward-growth function is thought to integrate the signal from the stimulation of neurons and takes a form with which it can combine with other values that share those neural properties, e.g., with a signal for the rate of reinforcement. If delay to reward delivery simultaneously combines with the input of the function, to be able to do so it must take on a form that is different than that of the output. Thus, the possibility that time is treated and encoded differently by the brain in different contexts is glaring. Both rate of reinforcement and delay to reward delivery share the effect of shifting the reward mountain along the price axis. Whether that is an overlap in the information processing by the brain or whether distinct mechanisms cause the same shift along the price axis cannot be ascertained through this model. Several variables with different mechanisms could result in scaling the output of the reward-growth function and in moving the mountain along the price axis.

Our results suggest that ratios of subjective reward delay and subjective reward magnitude do not combine multiplicatively. A hallmark of multiplicative combination is that the factor with which one variable is changed results in an equal change in the product of the two factors at all values of the variables. Thus, if the ratios of subjective reward magnitude and subjective reward delay were to combine multiplicatively, doubling the delay to reward delivery

should reduce the value of a large reward by the same factor as a small reward. Thus, the delay to reward delivery should have the same effect on the product of delay and magnitude, independent of the reward magnitude. Results from studies using concurrent variable-interval schedules have been interpreted to suggest that reward magnitude and delay do not have independent effects on value (e.g., Green & Snyderman, 1980; Ito & Asaki, 1982). These differences in results have been assumed, not unreasonably, to result from differences in procedure (Leon & Gallistel, 1998). Rodriguez and Logue (1986) similarly reported that only one pigeon out of eight seemed to have a significant interaction between reward magnitude and reward delay. The methodology used in our study and specifically, our choice of electrical stimulation rather than food reward allowed us to significantly reduce the variability associated with behaviour and allowed us to test the effect of delay on the full spectrum of animal behaviour toward reward magnitude and reward costs. As such, and combined with evidence from our results that the output of the reward-growth function is also scaled with delay to reward delivery, it is possible that a model with scalar multiplication of reward magnitude and reward delay would have fit the data obtained by Rodriguez and Logue (1986). However, they did not have the resolution to detect the additional interaction between reward magnitude and reward delay. The other studies that we are aware of that have studied the effect of delay on the value of BSR were not designed to look for interactions between the two (Fouriezos & Randall, 1997; Mazur et al., 1987). The data from those studies support our data. Fouriezos and Randall, (1997) found a 0.1 log unit/s change in reward magnitude. Both hyperbolic functions (which assume scalar multiplication) and exponential functions (which do not) fit their data equally well and no statistically significant differences between the models were observed. Mazur et al. (1987), showed that a hyperbolic function fit their data better than an exponential one, but under the assumption that ratios of reward delay and reward magnitude combine multiplicatively. That is, they assumed that the discount rate would be the same for all ratios of magnitudes. Mapping the effect of a wide range of delays on the frequency and price sweep would be the next step in being able to model the way that the two parameters interact to determine value.

In humans, it is readily acknowledged that the factor with which delay discounts the value of reward depends on reward magnitude, such that smaller rewards are discounted much more steeply than larger rewards (e.g., Estle et al., 2006; Green et al., 1997; Raineri & Rachlin, 1993). Since Rodriguez and Logue (1986), many other researchers have unsuccessfully attempted to replicate the interaction found in humans in other animals (R. Grace, 1999; Green et al., 2004; Ong & White, 2004; Richards et al., 1997). This difference in the findings has been boiled down to a species-difference (Winstanley, C., 2010), although it may also, in come cases at least, be due to a difference in methodology (see Sonnenschein et al., 2003). However, results from our study suggest that decision-making in rats about delayed rewards may have more in common with the observations of decision-making in humans than has often been assumed. The reason why the dependence of delay's effect on subjective reward magnitude is more readily observed in

humans than animals needs to be further investigated, with more consistent methodology and by looking at behaviour through a multidimensional perspective. If that difference remains, one possible explanation is that studies using humans often use hypothetical scenarios, whereas animals are exposed to real rewards. Moreover, animals are repeatedly presented with the same scenario to train them on a task that may become habitual while that repetition is unnecessary in procedures used to test delay discounting in humans.

3.7. Conclusion

The underlying assumption in research conducted on decision-making with delayed reinforcement has largely been that the ratios of subjective reward magnitude and those of subjective delay to reinforcement combine in a multiplicative manner. This assumption has dictated many of the computational models proposed to describe the combinatorial process for the two numbers and those proposed to describe how reward and time are individually encoded. Results from our study provide evidence that reward magnitude and reward delay are not subjectively combined in a simple scalar manner. As such, a reassessment and reinterpretation of the literature from a very large area in this field of study is required to accommodate a non-scalar combination of subjective reward magnitude and subjective reward delay.

Chapter 4
General Discussion

While conducting the work presented in this dissertation, I was reminded of a valuable, but embarrassingly basic, lesson. To accurately interpret data and to design experiments in a way that gives us more meaningful answers, we must first diligently test the assumptions that underlie our interpretations. For me, that meant that we must understand the fundamental ways with which animals compute different aspects of their world to guide their behaviour before we can make sense of how any mediating variables may affect that computation.

I showed in Chapter 2 that MPH could disrupt the development of animals' decision-making systems. Animals who were exposed to methylphenidate during adolescence were, on average, more impulsive than their sober counterparts, even long after the drug intake had ended. Specifically, the animals were more impulsive in that they devalued delayed rewards significantly faster than the control group. This effect was not observed when the exposure to methylphenidate occurred during adulthood. Of course, whether these results apply to humans in general, to humans who are typically prescribed MPH, or to adolescent humans who take MPH recreationally remains to be investigated. However, the results indicate that the system underlying such decision-making develops over time. They indicate that the development of the decision-making process can be disrupted, and they identify the neurotransmitters that are possibly responsible for mediating that disruption. I initially reasoned that the neurotransmitters involved mediate that decision-making through changing either

- the way they experienced a reward,
- the way they experienced time, or
- the way they combined those two entities.

In considering reward and time as two separate entities, I clearly considered the generalized matching law and its extensions to be laws meant not to be broken. In my defence, the matching law has been supported in thousands of experiments across species and procedures. The incorrect assumption I initially made was to generalize the same replicability and reliability to its extensions.

The concatenated matching law, an extension of the generalized matching law, treats reward magnitude and the delay to obtain the reward as two independent entities, the product of which makes up reward value. Results detailed in Chapter 3, however, show support against that assumption. Using the reward-mountain model, we could test whether the delay to reward delivery combines with reward intensity after reward intensity has been computed (as predicted by the assumption of scalar combination in the concatenated matching law), or whether the delay changes the sensitivity of the neurons responsible for that rewarding effect during the

computation of reward intensity. In this latter scenario, ratios of reward delay and reward magnitude do not combine multiplicatively. We found in Chapter 3 that delay changed both. The surprising result was that delayed delivery of the reward changed the sensitivity to rewarding stimulation in our rats, prior to the peak detection of reward intensity. This shows that delay had access to the input of the function that transforms the stimulation into reward intensity, and that the product of the ratios of reward delay and reward magnitude is not a simple multiplication. As expected, delay also moved the mountain along the price axis. Thus, delay also caused its effect at a stage of processing later in the reward circuitry. However, the reward-mountain model does not identify where the effect later in the reward-circuitry occurs. The effect could have occurred through changes in the reward gain, the opportunity cost, or the value of alternate activities. Thus, most of this discussion focuses on the implications of the more unexpected finding that delay moved the mountain prior to the peak detection of reward intensity.

It is important to note here that other time-dependent variables, such as the rate of reward delivery, have shown to combine multiplicatively with reward magnitude in line with the concatenated matching law (Leon & Gallistel, 1998). When the rate of a reward is increased, the average delay to reward delivery is also increased. Within our task, the time to delivery of a reward also inherently changes as one of our other independent variables changes. Specifically, as the experimenter increases the opportunity cost, they also increase the time to reward delivery from the time the animal begins to press the lever. The reward mountain model which assumes multiplicative combination of ratios of opportunity cost and reward intensity has previously been successfully validated. Thus, the interval to the reward delivery after the price requirement is met, seems to be computationally different from those other timing processes because unlike those other timing processes, it also affects reward sensitivity. It seems then that interval timing may not take the same neural form uniformly across the brain. Instead, the brain may be able to incorporate different forms of timing for different purposes. Hereafter, unless specified otherwise, reward delays in this discussion refer to the delay to reward delivery that rats experience after the satisfaction of the price requirement as that experienced by animals in our paradigm.

4.1. Effect of MPH on the Reward Circuitry

According to data presented in Chapter 3, MPH could cause its effect on the discounting of delayed rewards through several mechanisms:

- a nonlinear change in the valuation of the reward, such that the difference between the values of a smaller and larger reward is no longer the same.
- an increase in the perceived elapsed time prior to the peak detection of reward intensity.
- an increase in the perceived elapsed time after the peak detection of reward intensity.
- a change in the process through which reward intensity and elapsed time are combined, during the computation of reward intensity.

- a change in the process through which reward intensity and elapsed time are combined, after the computation of reward intensity.

Some insight into which of the above modalities stimulant drugs may adopt to affect delay discounting can be gained by further use of the reward-mountain model. Another condition could be added to the experimental design used in Chapter 3, where half of the rats are exposed to stimulants, and the other half are not. To ensure that long-term effects of manipulations can be observed through this experimental procedure, we continued to test rats for many months after sufficient data had been collected. The last test session collected for R03 was five months after the first test session. Data for rats R06 and R08 was continuously collected for four months and three months, respectively, after which point testing had to be abruptly stopped due to the COVID-19 global pandemic. Nonetheless, we believe this to be a sufficient time interval to test whether the effect of delay on the mountain remains stable over time. Studies conducted to observe the long-term effects of stimulant drugs have waited as little as 10 days to detect statistically significant differences in delay discounting (Hernandez et al., 2014), and in our study detailed in Chapter 2, we waited 26 days after the last injection of MPH to begin training the animals. Our data showed no statistically reliable differences between the position parameters observed during the first 10 and those during the last 10 sessions for any of the three rats. The 95% confidence intervals around the position parameter between the blocks of sessions overlapped with each other. However, it should be noted that for R06 in the last ten sessions, the 95% confidence intervals for the delay induced shift in the F_{hm} parameters when comparing both delays with baseline, included zero. Similarly, for R08, the confidence intervals around the delay-induced shift in the F_{hm} parameter for the smaller delay of 2 s compared to baseline also included zero during the last 10 sessions (see Table 4.1). Whether that difference between the first 10 sessions and the last 10 sessions in those two rats is an artifact of higher variability in those animals, or whether the shift in F_{hm} slowly decreases with time in some rats requires further research with a larger sample size. It is possible that as the task becomes more habitual, the animal changes the computational processing with which that task is conducted. Even if that is the case, the change in shifts was small and with an adequate control group, any long-term effects of manipulations would still be able to be observed along each axis.

That said, it would not be possible to observe what long-term effects a drug may have during the animal's development using this experimental procedure. Even if we were able to implant electrodes in adolescent rat brains, the position of those electrodes would move as their brains develop and become larger, confounding the interpretation of any effects that may be observed. However, long-term effects of stronger stimulants, such as cocaine, on impulsive choice have been observed even in adult rats (Hernandez et al., 2014; Logue et al., 1992; Paine et al., 2003; Simon et al., 2007). Thus, where in the stages of reward circuitry those drugs act to alter the effect of delay on reward value can indeed be observed using the reward-mountain

Table 4. 1.

Tukey Mean differences in Position Parameters: First vs Last Ten Sessions

Sessions	R03			R06			R08		
	CI_{low}	Shift	CI_{high}	CI_{low}	Shift	CI_{high}	CI_{low}	Shift	CI_{high}
$\log_{10} F_{hm}$: 0.1s vs 2s									
First 10	-0.153	-0.099	-0.058	-0.089	-0.018	0.017	-0.143	-0.098	-0.039
Last 10	-0.084	-0.059	-0.026	-0.008	0.027	0.111	-0.090	-0.023	0.152
$\log_{10} F_{hm}$: 0.1s vs 4s									
First 10	-0.183	-0.131	-0.094	-0.086	-0.050	-0.019	-0.170	-0.144	-0.109
Last 10	-0.126	-0.094	-0.054	-0.053	-0.023	0.0168	-0.151	-0.083	-0.065
$\log_{10} P_e$: 0.1s vs 2s									
First 10	0.145	0.209	0.284	0.198	0.293	0.353	0.057	0.193	0.383
Last 10	0.142	0.229	0.279	0.204	0.277	0.333	0.153	0.232	0.337
$\log_{10} P_e$: 0.1s vs 4s									
First 10	0.400	0.511	0.614	0.463	0.540	0.621	0.434	0.557	0.743
Last 10	0.481	0.568	0.637	0.349	0.436	0.545	0.649	0.769	0.885

Note. middle 95th percentile of bootstrapped parameter estimates do not overlap in any of the conditions.

model. Moreover, the reward-mountain model is not specific only to BSR rewards. In theory, the model could be used with other rewards as well. If the reward-mountain model is validated with food rewards, the developmental effects of drugs on the stage of reward circuitry in which delay discounting occurs can also be usefully investigated.

4.2 Non-Scalar Combination of Reward Magnitude and Delay and the Delay Discounting Function

Animals across species devalue future rewards in a similar manner. While the exact form of the mathematical function underlying that behaviour is still under debate, the consensus seems to be that the function is hyperbolic. In Chapter 1, I detailed some reasons why I believe that consensus came to be. The main reason is that a hyperbolic function can predict preference reversals despite a constant discount rate. This is because in a hyperbolic function, the rate of change is not constant at all points of the curve. This can result in a reversal of preference as the absolute delays to both rewards are increased by the same length of time. To be able to explain preference reversals with an exponential function, in which the curve loses the same constant proportion of height for every unit of time, it must be assumed that that the constant which determines the discount rate varies with the magnitude of the reward (Ainslie & Herrnstein, 1981) and thus that the ratio of reward magnitude cannot combine with the ratio of delays in a simple scalar manner. This would be inconsistent with the concatenated matching law. Results from Chapter 3 imply that the exponential function should not be ruled out as an accurate descriptor of animal behaviour on the assumption of the concatenated matching law alone. Efforts to describe a more valid normative account of decisions made about delayed rewards are needed and require further testing. The form of the mathematical functions that discount reward value prior to the peak detection of reward intensity (shifts on the frequency axis) and after the peak detection of reward intensity (shifts on the price axis) cannot be interpreted through the data collected from our study because the confidence intervals surrounding the position parameters are too wide (see Table 4.1) and the rats were only tested on three different delays to reward delivery. A psychophysical study therefore also needs to be conducted using a wider range of delays to describe the shifts in the reward mountain model and quantify how the range of delays affects reward value on each axis.

4.3. Neural mechanisms for interval timing and the reward-mountain model

Given the large variety of models that have been proposed to describe the way mammalian brains assess time intervals, there are several ways to categorize the models. One of those categories relevant to our discussion includes models that assume a dedicated stopwatch-like neural timer (or multiple such timers) compared to models that use other non-dedicated mechanisms for timing. The most popular and widely cited theory of interval timing is the scalar expectancy theory which assumes a centralized and dedicated stopwatch-like timer. I identify several reasons, in Chapter 1, to favour models that do not assume such timers. Similar to learning, which is considered to be a task carried out in a non-dedicated, non-centralized manner

in the brain, some researchers suggest that all neurons are capable of interval timing (Buonomano, 2017; Ivry & Schlerf, 2008). In fact, Ward and colleagues (2013) hypothesized that while learning, mammalian brains do not associate a conditioned stimulus with an unconditioned stimulus but instead obtain information about the interval between the two stimuli. Evidence outlined in Section 1.6 of single neurons that can learn time intervals supports this idea.

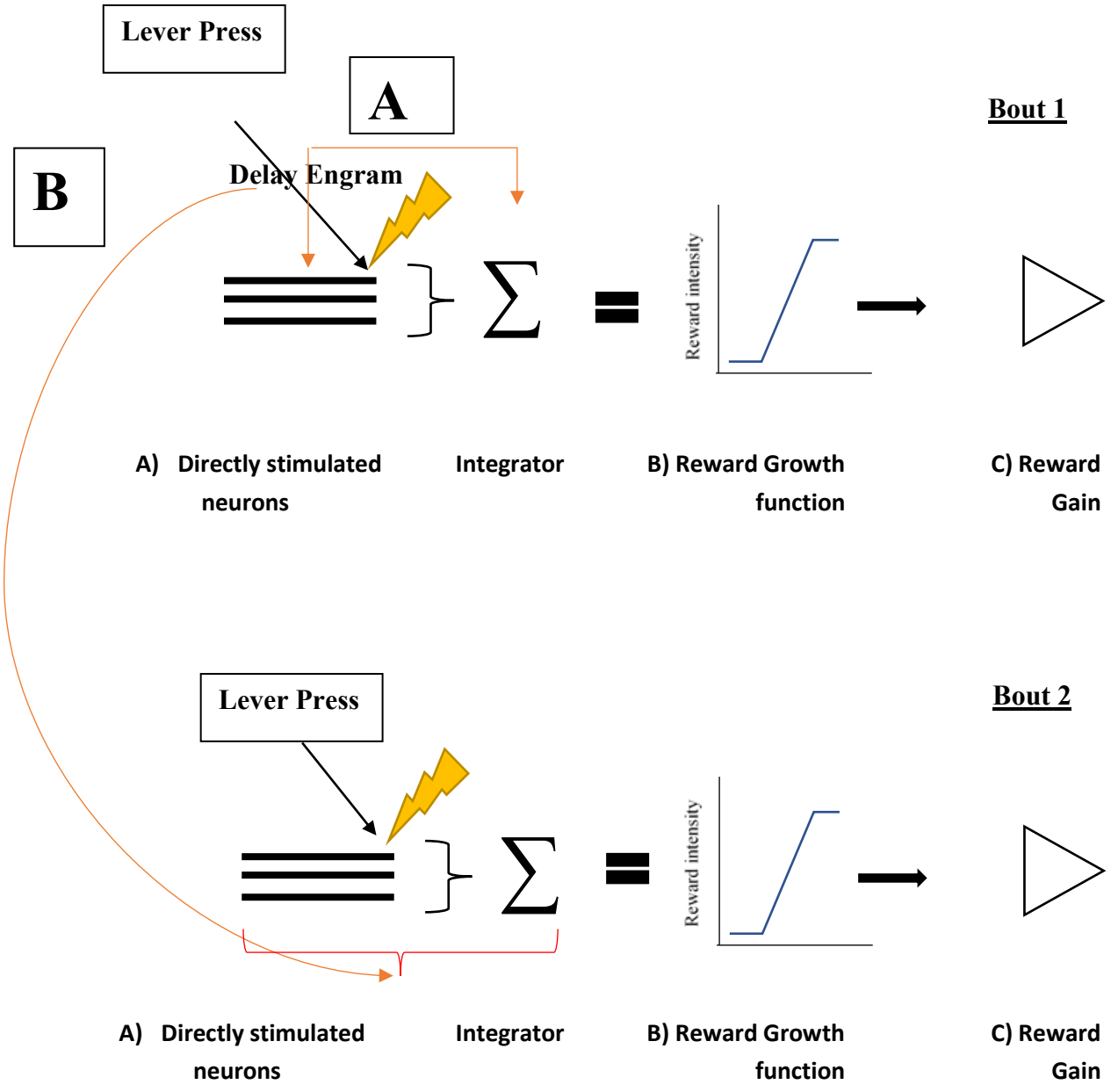
Our study shows that one of the ways delay to reward delivery affects reward value during ICSS is by changing the direct effect of the stimulation itself, or by changing how the resulting signal is integrated. While these results do not give definitive proof that interval timing may be ubiquitously processed in the brain, they certainly allow for this possibility. We could not have allowed for this possibility if we assumed that ratios of reward magnitude and reward delay combine multiplicatively. In that scenario the encoding of the reward delay and the reward magnitude would have had to be separate from each other. The possible ways that delay could cause a shift of the mountain along the frequency axis are, thus, worth discussion.

Figure 4.1 zooms in on the first stage neurons and their integration in the reward mountain model to clarify that discussion. A description of a trial and the chronology of the information that the animal receives is also pertinent here. When an animal presses the lever at the beginning of the trial, the animal has no idea what the parameters of the reward are going to be. As time elapses and the animal continues to press the lever, the animal gets more information about the price, and that information becomes established over time in its brain. Once the response requirement has been satisfied the lever retracts, and the cue light turns off signalling the price requirement to obtain the reward to the animal. It then begins to obtain information about the delay to reward delivery. That information is not complete prior to the stimulation of the directly stimulated neurons. The train of stimulation is delivered either 0.1 s, 2 s or 4 s later. The interval between the animal's recognition that the lever has retracted and the cue light has turned off, and the delivery of the train stimulation, then, according to our results, affects F_{hm} during subsequent rewards in that trial. There seem to be two ways that might explain how that interval could shift F_{hm} and result in reward intensity that has been discounted by the delay (see Figure 4.2):

A. The duration of the interval affects the signal for reward intensity immediately before the reward intensity is computed. The first train of stimulation that the animal receives in this scenario would be delay discounted. This could result from different mechanisms:
    1. The simplest possibility would be that as the interval elapses, a neural process desensitizes the neurons that produce the rewarding effect. This possibility would not require an engram, and the delay interval would affect the directly stimulated neurons or their integration in real-time. In this possibility, delay discounting would be a *regular* problem, in that delay to reward delivery could shift F_{hm} without being written to memory.

Figure 4. 1.

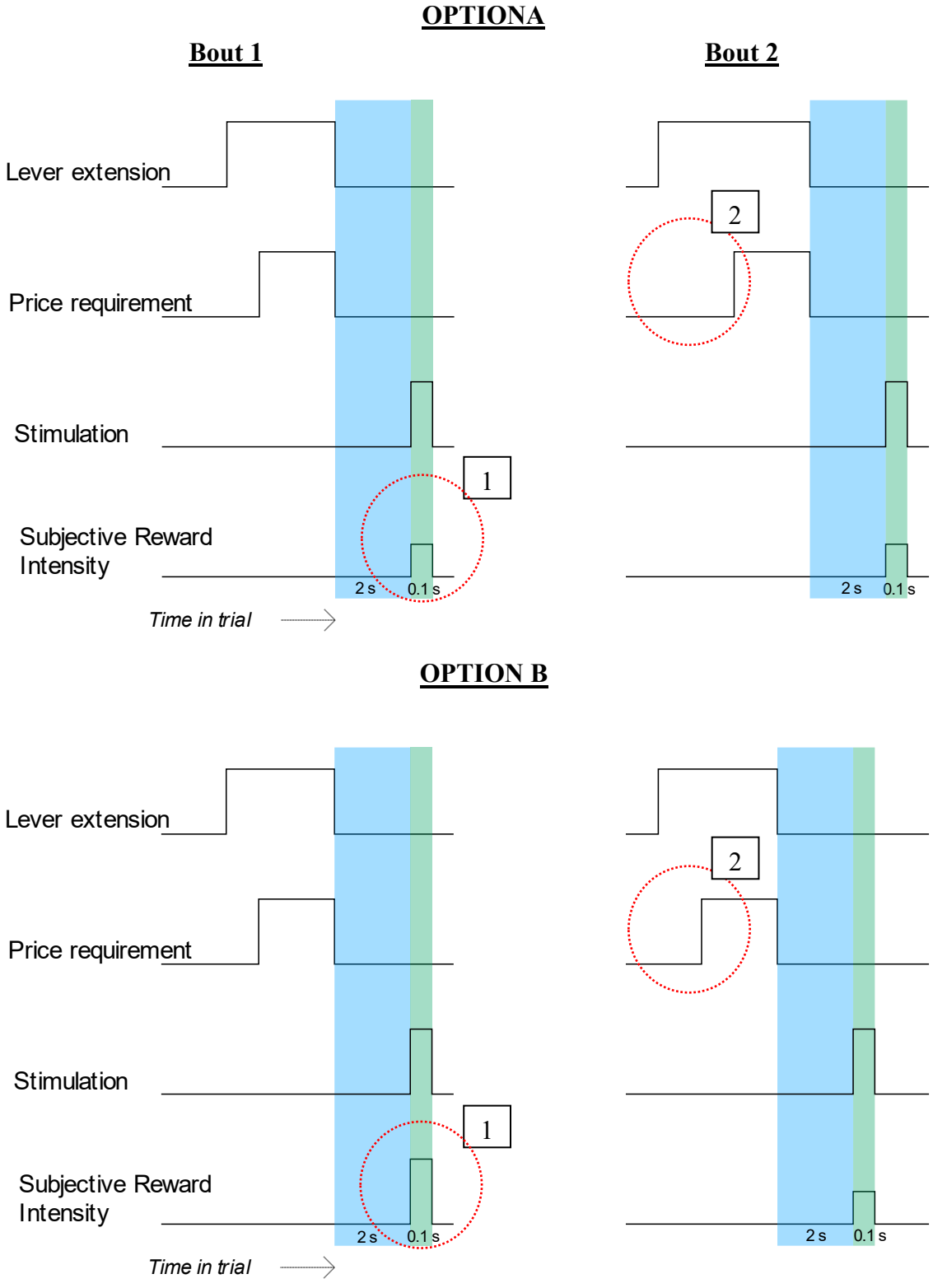
Simplified Schematic of the Reward-Mountain Model



Note. This figure zooms into the simplified schematic to focus on the first part of the reward-mountain model that translates electrical stimulation in the directly stimulated neurons into the reward growth function. I have supplemented the schematic with two options A and B which point at the stage in reward processing where delay could be discounting reward intensity at its input to the reward-growth function.

Figure 4. 2.

Ways delay may shift F_{hm} to return a delay-discounted reward intensity



Note. The diagram shows two conditions. The graphs on the left-hand side illustrate the sequence of events during the first bout in a trial. At this point, the rat is unaware of the delay at which reward will be delivered. The graphs on the right side illustrate the sequence of events in the second bout. This is the bout immediately after the first reward is delivered and the rat knows the parameters of the trial. The top two quadrants illustrate predictions arising from Option A, while the bottom two quadrants show predictions arising from Option B. Trial time runs from left to right. The first row in each quadrant depicts the presentation of the cue light and lever. The second row depicts the duration that the lever is pressed. This is variable depending on the rat, but the illustration depicts the scenario where the lever is pressed until the price requirement set by the experimenter is satisfied. The blue shaded region depicts that satisfaction of the price requirement and the onset of the experimenter-set delay to reward delivery. The third row depicts the delivery of the stimulation, and the fourth row represents the subjective reward intensity resulting from that stimulation. The green shaded region represents the time during which reward is delivered. According to Option A, 1) subjective reward intensity is immediately discounted and can thus 2) affect behaviour immediately after the first reward is delivered. Under Option B, the animal receives 1) a full strength of reward intensity at first so 2) behaviour immediately after the first reward cannot be affected by delay. It is possible, however, for the reward intensity to be discounted in the second bout. Behaviour in this scenario can only be affected in the third bout (not pictured here).

2. During the interval, a time-dependent neural process makes note of the states of the two events, and estimates the time interval that has elapsed between them. Upon stimulation, that interval is stored into an engram. The input of the stimulation (the aggregate spike rate) is also recorded as an engram. These two values are then combined, the combination would feed into the reward growth function and determine the reward intensity. Although this possibility is more complicated, there is benefit to the animal in storing the aggregate spike rate as the ability to shift the reward intensity function allows much greater control over the variable.

B. During the interval, a time-dependent neural process makes note of the states of the two events and estimates the time interval that has elapsed between them. Upon stimulation, that interval is stored into an engram. The aggregate spike rate is not recorded or affected in real-time. Instead, since only the output of the reward growth function (reward intensity) is stored into an engram, rather than the input, the engram cannot affect reward sensitivity the first time that the reward is experienced. Sensitivity of neurons before the reward intensity is computed can only be altered after the reward intensity has been computed at least once. Thus, at first, the reward intensity experienced is of full strength. The engram for the interval then affects the sensitivity of the directly stimulated neurons for subsequent rewards.

While the results detailed in Chapter 3 cannot decisively distinguish between these possibilities, and this discussion is somewhat speculative at this point, it is a useful exercise so that we can impose limitations on models of delay discounting that are testable. For instance, possibility A can be distinguished from possibility B simply by observing how quickly delay shifts the mountain along the frequency axis.

4.4. How quickly does the delay induced shift in F_{hm} occur?

Under possibility A, the value that is stored for reward intensity is already discounted by delay on the first train of stimulation that the animal receives. This is because in either scenario, the elapsing delay is responsible for the desensitization of reward intensity in real-time. Under possibility B, on the other hand, the animal individually writes down the delay interval and writes down only the output of the reward intensity, rather than the input. For delay be written to memory, the neurons would first have to be stimulated, and their signal would feed into the reward growth function without being discounted. The representation of delay would then not be able to affect the input to the reward growth function until the next reward that the animal receives. Thus, the animal would receive a full-strength reward intensity signal in response to the first train, but the subsequent trains would be delay discounted. This can be observed behaviourally.

Under possibility B, if delay to reward delivery were to change the firing of the directly stimulated neurons or their integration, there would be no immediate shift on the frequency axis. A shift would only be possible starting from the third bout in the trial, after the reward intensity for the first reward has been calculated. Delay could not change the computation of reward intensity during the second bout (behaviour immediately following delivery of the first reward). We analyzed rat behaviour by frequency, price, and radial sweeps limited to the second bout in the trials to observe whether any trends stood out. Figure 4.3 shows how long the animal paused before working for the second stimulation (hereafter called post-reinforcement pauses) averaged across sessions as functions of frequency, price, and delays for R03. Figure 4.4 shows the proportion of sessions the rat decided not to work for further rewards in the trial after gaining information about the parameters of the first reward, hereafter called post-reinforcement discontinuances. Although data from an exemplary rat is shown, data was similar for the other two rats as well.

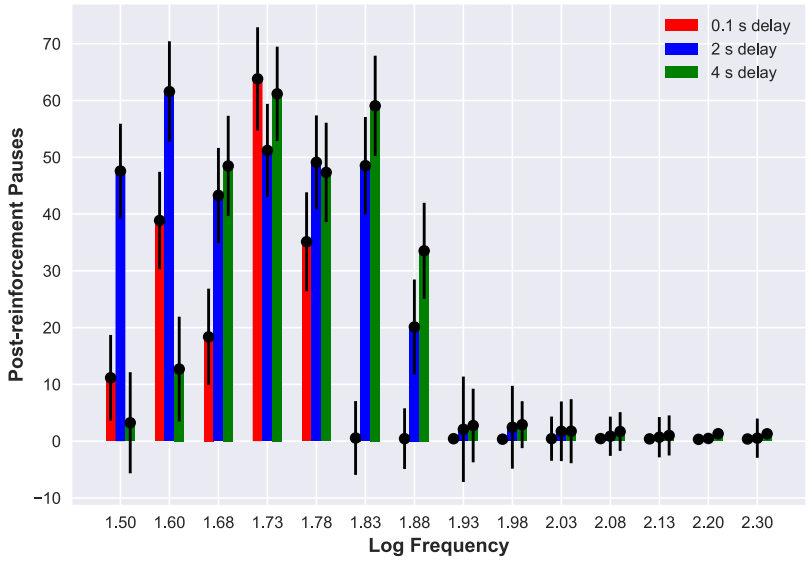
Clear, orderly trends were observed in the frequency sweeps; both post-reinforcement pauses, and post-reinforcement discontinuances clearly shifted with increasing delays. This same trend was not observed across the price sweeps, despite the much larger shifts in P_e compared to those in F_{hm} that were observed in Chapter 3. Thus, the effect of delay on value seems, after a single exposure, to immediately interact with varying frequencies, as predicted by possibility A. To further differentiate between the two possibilities, we analyzed whether an additional interaction was observed with varying frequencies and prices on additional bouts of stimulation throughout the duration of the trial. An analysis of this trend across bouts showed the same effect getting larger in the frequency sweep over the trial duration, and an effect of delay on the price sweep emerge as the trial progressed. We observed a systematic effect of opportunity cost on delay discounting on the price sweep, noticeable after the third reward rather than the first. Specifically, as the trial progressed rats stopped pressing at lower prices across sessions when the reward was delayed. If the animal showed that delay after having received one reward was immediately discouraging, that implies possibility A, and not B. The effect on the frequency sweep was immediate and got marginally bigger over time during the trial duration. Figures 4.5 and 4.6 illustrate this through averaged data points for frequency and price after the first reward on bout 2, and after the third reward on bout 4. A delay induced shift in the frequency curve immediately after delivery of the first reward, but not of the price curve, if substantiated, would suggest that a change in reward sensitivity (the input to the reward growth function) results in the immediate change in behaviour. Thus, these results indicate that we may be able to eliminate option B from our possible scenarios above, leaving only A as the likely explanation.

While the results of this analysis led to interesting observations, the experimental procedure that we used was not designed to observe the effects of delay on behaviour specific to the second bout of behaviour. Analysis of time allocation limited to the second bout, given our current procedure, discarded up to 24/25 of the data, leaving us with little remaining power for

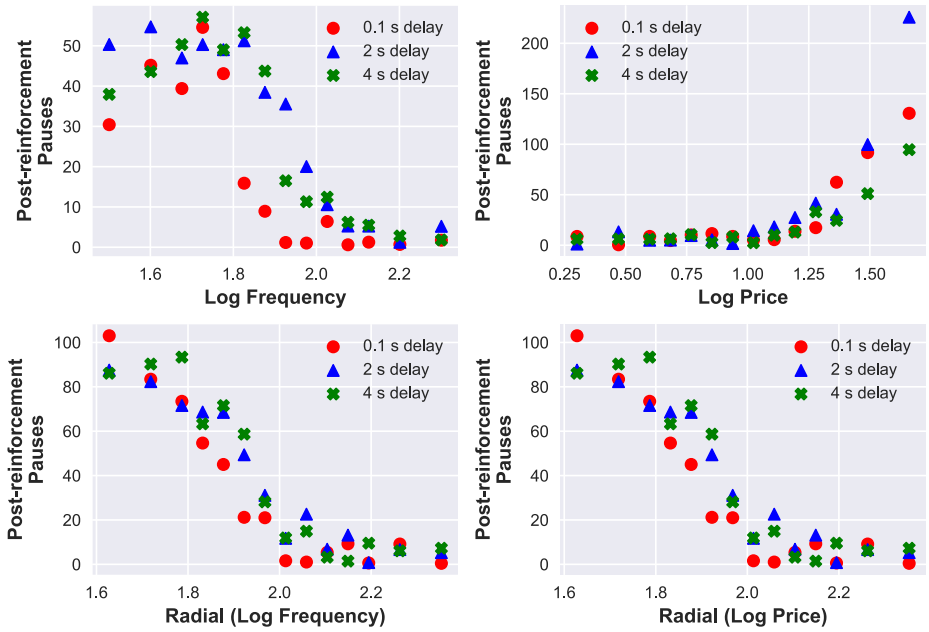
Figure 4. 3.

Post-Reinforcement Pauses as a Function of Frequency at Separate Delays

A)



B)

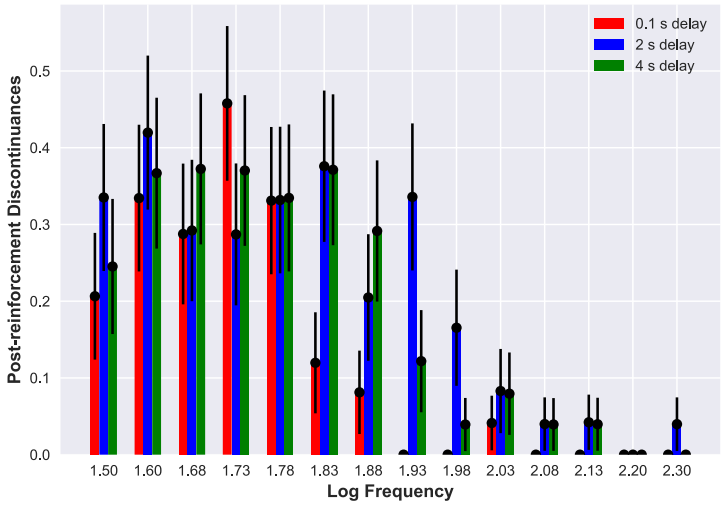


Note. A) Bars represent the Tukey mean of bootstrapped post-reinforcement pauses for R03. Error bars represent bootstrapped standard errors. B) Post reinforcement pauses separated into frequency sweeps, price sweeps and radial sweeps for all three conditions. An immediate shift on the frequency axis, but not the price axis supports the idea that the duration of the interval affects the signal for reward intensity immediately before or during the computation of reward intensity.

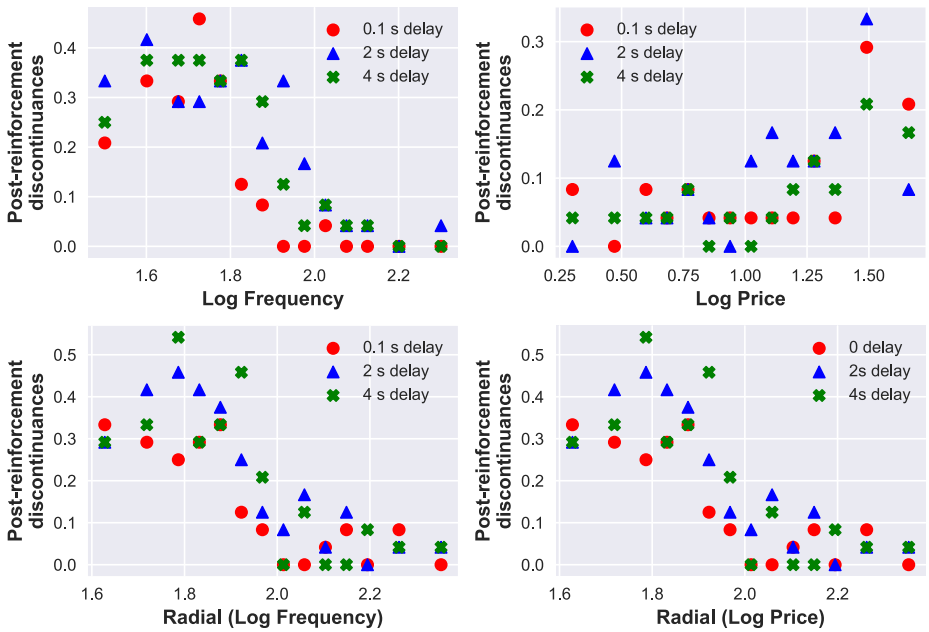
Figure 4. 4.

Post-Reinforcement Discontinuances as a Function of Delay

A)



B)



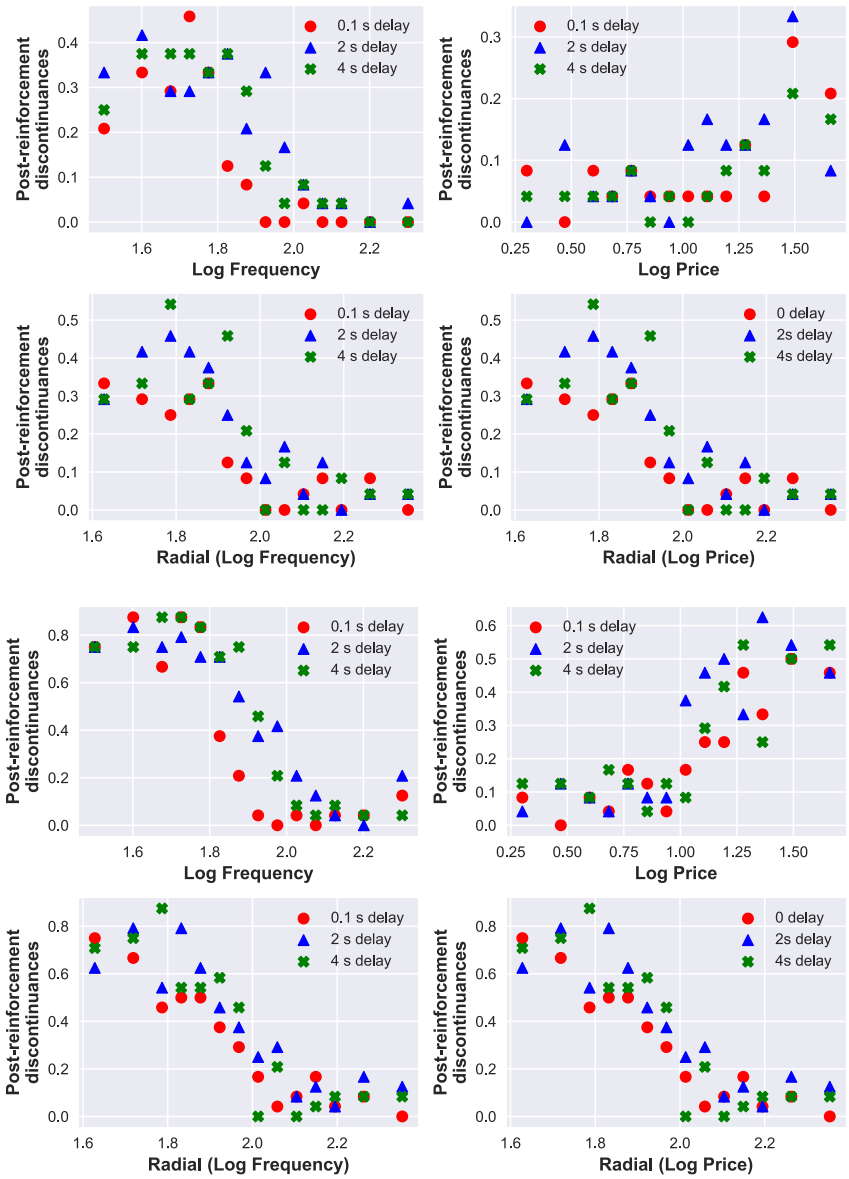
Note. A) Bars represent the Tukey mean of bootstrapped post-reinforcement discontinuances for R03. This is the proportion of times in the sessions that the animal stopped working for the reward immediately after obtaining the first reward. Error bars represent bootstrapped standard errors. B) Post reinforcement discontinuances separated into frequency sweeps, price sweeps and radial sweeps for all three conditions. An immediate shift on the frequency axis, but not the price axis supports the idea that the duration of the interval affects the signal for reward intensity immediately before or during the computation of reward intensity.

Figure 4. 5.

Post-Reinforcement Discontinuances on Bout 2 and Bout 4.

A) Bout 2

B) Bout 4

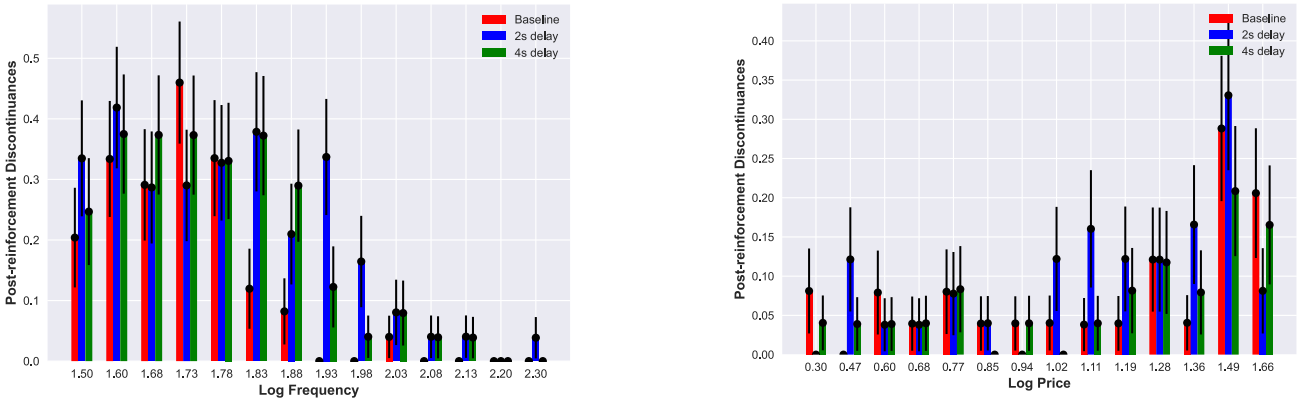


Note. Data presented here was obtained from R03. The y-axis represents the proportion of trials that R03 discontinued pressing the lever after delivery of the A) first reward, and B) the third reward. Datapoints show bootstrapped averages across the sessions. An effect of delay on frequency is apparent immediately after the first reward, whereas the effect of delay as a function of price is only apparent on bout 4 at the 2 second delay. This suggests that a change in reward sensitivity (the input to the reward growth function) results in the immediate change in the devaluation of delayed rewards.

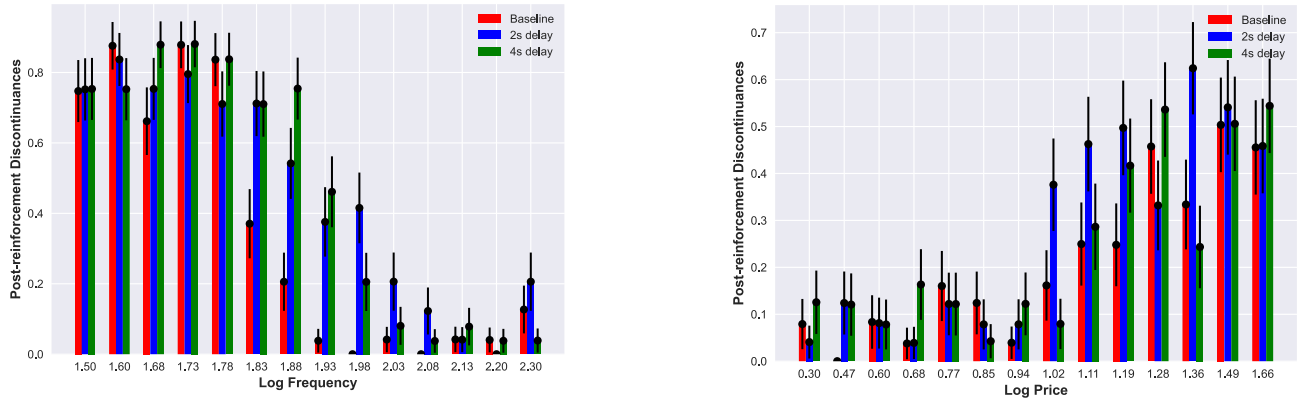
Figure 4. 6.

Bar Graph of Post-Reinforcement Discontinuances on Bout 2, and Bout 4.

A) Bout 2



B) Bout 4



Note. Data presented here was obtained from R03. The y-axis represents the proportion of trials that R03 discontinued pressing the lever after delivery of the A) first reward, and B) the third reward. Datapoints show bootstrapped averages across the sessions. An effect of delay on frequency is apparent immediately after the first reward, whereas the effect of delay as a function of price is only apparent on bout 4 at the 2 second delay. This suggests that a change in reward sensitivity (the input to the reward growth function) results in the immediate change in the devaluation of delayed rewards. Error bars represent bootstrapped SEM.

adequate analysis. The high variability in the data on the price sweep may have obscured any delay-induced shifts. To validate that the shift in post-reinforcement pauses and discontinuances as a function of frequency arises from effects on the reward sensitivity and not after reward intensity has been computed, the structure of the procedure to obtain the reward mountain would need to be changed. Suppose the animal is only able to get three rewards per trial, one to inform the animal of the parameters, the second to observe time allocation on the second bout, and the third to give the animal enough time to get that second reward, and each trial is randomly repeated 25 times. In that case, we can obtain a much larger amount of data that is limited to the point in decision-making immediately after the rat obtains information about the reward. Our results show that rats only need a single exposure to change their behaviour allowing us to split each trial into many parts to examine what leads to that immediate change. If a frequency shift disappears with this procedure, it would suggest that the effect of delay on the computation of reward intensity at its input is not immediate, and a model with slower updating of information would be required. On the other hand, if a frequency shift is observed, then possibility A from our scenario above can be eliminated. The change in behaviour immediately after the first exposure to the reward can then be attributed to the effect of the delay on the input of the function and not the output. The simplest way to account for this data is that an ongoing desensitization of neurons produces the reward devaluing effect as the delay elapses. Ongoing research on the neural mechanisms underlying interval timing can present clues on how that could happen at a neurological level.

4.5. What can we learn from neural mechanisms of circadian timing?

The search for a neural basis for interval timing has centered around a dedicated stopwatch similar to the master clock found in circadian timing. However, unlike circadian timing, which uses the sun as a reference point for synchronization throughout the body, in interval timing the reference points vary greatly (any event can be timed). A dedicated, centralised system to time intervals seems less useful when a dedicated, centralised reference point does not exist. In that vein, some researchers (Ivry & Schlerf, 2008) have argued for a non-dedicated mechanism for interval timing. As discussed earlier, a variety of models have been proposed for interval timing in a non-dedicated manner, on a scale of single neurons to populations of neurons that spatially and temporally work together to act as a clock. For example, a decay model could fit the data, such that the ability of neurons to fire decreases as time elapses after the response requirement. Instead, perhaps the integrator simultaneously counts the action potentials, computes the time that has elapsed since the response requirement had been met, and combines the two using phase-locked oscillatory activity. Both these examples imply that the signal for reward magnitude can depend on, and change as the time interval to its delivery elapses.

Many of these proposed mechanisms are centered around the electrical and synaptic properties of neurons. Much neural evidence has shown that neuronal firing is timed (see Section

1.6.3). However, how those neurons know when to fire is a lot less clear. Information according to mechanisms that rely on electrical or synaptic properties could be modelled by, for example, synfire chains. Neurons in this model can be compared to falling dominoes. They fire in a chain where each neuron's firing is responsible for a sequence of timed firing in other neurons. We would know the time elapsed depending on which neurons in the sequence were active or which domino had fallen (Hahnloser et al., 2002; Long & Fee, 2008). The problem with these proposed mechanisms is that neurons or synapses must then be assigned to each time interval. If memory of delays between all possible events and all possible rewards (both anticipatory and experienced) are stored in synapses or network activities, an infinite number of such combinations would need to be stored. While the brain has an incredibly large number of neurons, the number is not infinite. Thus, these mechanisms seem implausible. The way that circadian timing deals with this issue is that each cell in our body has an approximately 24-hour timer that can then be synchronised with the sun through a master clock. Thus, all neurons have the ability to time, but they synchronize with each other to be on the same page. Perhaps, what we can generalize from the information learned about neural circadian timing is not the master clock that does the synchronising, but the cell-intrinsic biochemical timing mechanisms that allow circadian timing to occur so ubiquitously throughout the body. Perhaps the directly stimulated neurons or the integrator are like the Purkinje cells I described in section 1.6.3, and each neuron can learn when to expect the stimulation, altering their behaviour to the stimulation at the time that it is delivered. This implies that intercellular processes in the neurons signalling, or computing reward magnitude could simultaneously compute the delay to their stimulation, and construct that combination into one signal.

The neural evidence for the ubiquity of time encoding for durations is compiling quickly. In the last year, two pivotal studies have been published that show evidence, for the first time, of a molecular cell-intrinsic neural mechanism for interval timing through studies conducted on fruit flies (Thornquist et al., 2020, 2021). The mechanism shown exists within four neurons of the drosophila ventral nervous system that can time 7 mins between the beginning of copulation to the time of ejaculation. At the completion of that time interval, the four neurons fire, the fly ejaculates, and becomes less motivated to continue mating. In these two papers, the authors show that this timing process exists within each of those neurons and is controlled by CAMKII. CAMKII is phosphorylated when copulation begins, and slowly dephosphorylates as the 7 mins progress. The four neurons which are recurrently connected communicate, around the six-minute mark, then synchronize with each other to fire at the same time. This change occurs through a combination of electrical activity and molecular machinery, where the speed at which ATP is converted to CAMP is slowed or hastened to decrease or increase the membrane potential of the neuron, so that the neurons can fire together. If timing in neurons in a basic circuit in the relatively simple nervous system of fruit flies is capable of being computed simultaneously with other functions, it seems reasonable to propose that the neural circuitry in higher mammals would

also have that capability. It also seems reasonable to propose that perhaps timing is an inherent part of all neural processing. Importantly, if all neurons can compute time, then the space for a pre-existing memory of an infinite number of durations need not be allocated to synapses and neural networks. All that would be needed are memories of existing events, through which neurons can calculate the duration that has elapsed (Gallistel & King, 2009).

4.6. Multidimensionality of reward intensity

If one prescribes to Immanuel Kant's theory of time and space, time and space arise from the mind's operations and precede all experiences, such that nothing can be experienced outside of time and space. This is undoubtedly intuitively appealing. If we apply this theory to our discussion here, reward delay may not be able to be separated from the reward magnitude itself. Similarly, the time it takes to make the effort to get a reward could be imbued in the computation of the effort itself, and the average time to reward in a reward rate could be imbued in the computation of the reward rate.

It seems from our results that the measuring system for reward magnitude cannot be independent of confounding influences but that brains encode a magnitude in its temporal context. In all the possible ways magnitude and delay may combine subjectively to alter the input of the reward growth function, whether reward delay is stored separately from reward magnitude, and whether the effect of delay occurs immediately or gradually, delay affects reward magnitude during its computation. In other words, the computation of reward magnitude seems to be interwoven with the delay to its delivery.

The interdependence between reward magnitude and reward delay is clear if the directly stimulated neurons themselves can change their behaviour with time (option A1). Suppose we allow for option B as a possible mechanism. The change in sensitivity in this possibility occurs during the computation of reward magnitude, and therefore, once delay is incorporated into that computation, the output of the reward-growth function consists of the (partial) effect of delay on reward value. In other words, the output of the reward growth function, the symbol for reward intensity incorporates reward delay. Before delay is incorporated into the computation, reward magnitude has not yet been placed in its temporal context, and the animal behaves as such, acting as though the reward magnitude they receive is of full strength. If a symbol incorporates information about two dimensions, it implies that the symbol itself is at least two-dimensional or that it cannot be decoded into the two separate dimensions once it is formed. The challenge now is to discover the properties of that symbol. It could range anywhere from molecular machinery within the DNA to reverberating activity in neural networks, a combined effort by the many neuronal networks of the brain. Finally, consider the scenario in which option A2 underlies the way delay changes reward sensitivity. In this case, an engram of the input of the reward growth function is stored to be combined with the engram of the delay to reward delivery, resulting in a delay-discounted output of the reward growth function which can be combined with other costs.

We must ask ourselves whether this added step so early in reward processing is a computational consequence of the way time is encoded and assessed, or is it merely a coincidence that delay acts at this early stage of reward processing unlike all other costs tested so far? Given the plethora of neural substrates demonstrated to be related to time perception, a coincidence seems to me to be an unlikely answer to that question.

In any case, the stimulation that animals receive in ICSS has previously been described as a signal for a unidimensional encoding integrated to form an engram for reward intensity. An aggregate of spike count makes up the output of that signal which is considered the symbol for maximum reward intensity produced by the stimulation. Our results, on the other hand, suggest that the output of the reward growth function features not only the aggregate spike rate but also the delay to the delivery of the reward. If the output of the reward-growth function is indeed the symbol for the previously experienced subjective reward magnitude, then in line with Kant's theory about space and time, reward magnitude as a separate entity on its own cannot exist outside its temporal context.

4.7. Conclusion

We found that MPH can cause long-term changes in the devaluation of delayed rewards when intake of the drug occurs during development. We then found that delay itself affects the computation of reward value at multiple stages in the reward circuitry. Assessments of when a reward is delivered seem to be intertwined with the assessments of reward magnitude. Since we present at least two stages in the reward circuitry where delay might devalue rewards, a single hyperbolic delay discounting function does not fully explain the underlying computation. Psychophysical studies must be conducted to determine the forms of the functions that discount reward value at multiple stages. This could then allow for a more accurate sense of how and where manipulations such as drug exposure may alter the devaluation of reward by delay.

References

Abbas, Z., Sweet, A., Hernandez, G., & Arvanitogiannis, A. (2017). Adolescent Exposure to Methylphenidate Increases Impulsive Choice Later in Life. *Frontiers in Behavioral Neuroscience*, *11*, 214. https://doi.org/10.3389/fnbeh.2017.00214

Adriani, W., Canese, R., Podo, F., & Laviola, G. (2007). 1H MRS-detectable metabolic brain changes and reduced impulsive behavior in adult rats exposed to methylphenidate during adolescence. *Neurotoxicology and Teratology*, *29*(1), 116–125. https://doi.org/10.1016/j.ntt.2006.11.010

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, *82*(4), 463–496. https://doi.org/10.1037/h0076860

Ainslie, G. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person* (pp. xvi, 440). Cambridge University Press.

Ainslie, G., & Herrnstein, R. J. (1981). Preference reversal and delayed reinforcement. *Animal Learning & Behavior*, *9*(4), 476–482. https://doi.org/10.3758/BF03209777

Alsop, B., & Davison, M. (1986). Preference for multiple versus mixed schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *45*(1), 33–45. https://doi.org/10.1901/jeab.1986.45-33

Andersen, S. L., Arvanitogiannis, A., Pliakas, A. M., LeBlanc, C., & Carlezon, W. A. (2002). Altered responsiveness to cocaine in rats exposed to methylphenidate during development. *Nature Neuroscience*, *5*(1), 13–14. https://doi.org/10.1038/nn777

Arvanitogiannis, A., & Shizgal, P. (2008). The reinforcement mountain: Allocation of behavior as a function of the rate and intensity of rewarding brain stimulation. *Behavioral Neuroscience*, *122*(5), 1126–1138. https://doi.org/10.1037/a0012679

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior*, *32*(2), 269–281. https://doi.org/10.1901/jeab.1979.32-269

Baum, W. M., & Rachlin, H. C. (1969). Choice as time allocation. *Journal of the Experimental Analysis of Behavior*, *12*(6), 861–874. https://doi.org/10.1901/jeab.1969.12-861

Bavarian, N., Flay, B. R., Ketcham, P. L., & Smit, E. (2015). The Illicit Use of Prescription Stimulants on College Campuses: A Theory-Guided Systematic Review. *Health Education & Behavior*, *42*(6), 719–729. https://doi.org/10.1177/1090198115580576

Benes, F. M., Vincent, S. L., Molloy, R., & Khan, Y. (1996). Increased interaction of dopamine-immunoreactive varicosities with GABA neurons of rat medial prefrontal cortex occurs during the postweanling period. *Synapse*, *23*(4), 237–245. https://doi.org/10.1002/(SICI)1098-2396(199608)23:4<237::AID-SYN1>3.0.CO;2-8

Berg, M. E., & Grace, R. C. (2004). Independence of terminal-link entry rate and immediacy in concurrent chains. *Journal of the Experimental Analysis of Behavior*, *82*(3), 235–251. https://doi.org/10.1901/jeab.2004.82-235

BGM channel. (2015, April 9). *Howard Rachlin, "Matching" SQAB*. https://www.youtube.com/watch?v=1DtoEv7iLws

Bolaños, C. A., Barrot, M., Berton, O., Wallace-Black, D., & Nestler, E. J. (2003). Methylphenidate treatment during pre- and periadolescence alters behavioral responses to emotional stimuli at adulthood. *Biological Psychiatry*, *54*(12), 1317–1329. https://doi.org/10.1016/S0006-3223(03)00570-5

Borrero, C. S. W., Vollmer, T. R., Borrero, J. C., Bourret, J. C., Sloman, K. N., Samaha, A. L., & Dallery, J. (2010). Concurrent Reinforcement Schedules for Problem Behavior and Appropriate Behavior: Experimental Applications of the Matching Law. *Journal of the*

Experimental Analysis of Behavior, *93*(3), 455–469. https://doi.org/10.1901/jeab.2010.93-455

Brandon, C. L., Marinelli, M., & White, F. J. (2003). Adolescent exposure to methylphenidate alters the activity of rat midbrain dopamine neurons. *Biological Psychiatry*, *54*(12), 1338–1344. https://doi.org/10.1016/S0006-3223(03)00787-X

Breton, Y.-A. (2013). *Molar and Molecular Models of Performance for Rewarding Brain Stimulation* [Phd, Concordia University]. https://spectrum.library.concordia.ca/977979/

Breton, Y.-A., Conover, K., & Shizgal, P. (2014). The effect of probability discounting on reward seeking: A three-dimensional perspective. *Frontiers in Behavioral Neuroscience*, *8*. https://doi.org/10.3389/fnbeh.2014.00284

Breton, Y.-A., Marcus, J. C., & Shizgal, P. (2009). Rattus Psychologicus: Construction of preferences by self-stimulating rats. *Behavioural Brain Research*, *202*(1), 77–91.

Breton, Y.-A., Mullett, A., Conover, K., & Shizgal, P. (2013). Validation and extension of the reward-mountain model. *Frontiers in Behavioral Neuroscience*, *7*. https://doi.org/10.3389/fnbeh.2013.00125

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, *6*(10), 755–765. https://doi.org/10.1038/nrn1764

Buonomano, D. V. (2017). *Your brain is a time machine: The neuroscience of physics and time* (First). W.W. Norton & Company, Inc.

Burgos, H., Cofré, C., Hernández, A., Sáez-Briones, P., Agurto, R., Castillo, A., Morales, B., & Zeise, M. L. (2015). Methylphenidate has long-lasting metaplastic effects in the prefrontal cortex of adolescent rats. *Behavioural Brain Research*, *291*, 112–117. https://doi.org/10.1016/j.bbr.2015.05.009

Carlezon, W. A., Mague, S. D., & Andersen, S. L. (2003). Enduring behavioral effects of early exposure to methylphenidate in rats. *Biological Psychiatry*, *54*(12), 1330–1337. https://doi.org/10.1016/j.biopsych.2003.08.020

Catania, A. C. (1963). Concurrent performances: A baseline for the study of reinforcement magnitude. *Journal of the Experimental Analysis of Behavior*, *6*(2), 299–300. https://doi.org/10.1901/jeab.1963.6-299

Cheng, J., Xiong, Z., Duffney, L. J., Wei, J., Liu, A., Liu, S., Chen, G.-J., & Yan, Z. (2014). Methylphenidate Exerts Dose-Dependent Effects on Glutamate Receptors and Behaviors. *Biological Psychiatry*, *76*(12), 953–962. https://doi.org/10.1016/j.biopsych.2014.04.003

Chung, S.-H., & Herrnstein, R. J. (1967). Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior*, *10*(1), 67–74. https://doi.org/10.1901/jeab.1967.10-67

Church, R. M. (1984a). Properties of the internal clock. *Annals of the New York Academy of Sciences*, *423*, 566–582. https://doi.org/10.1111/j.1749-6632.1984.tb23459.x

Church, R. M. (1984b). Properties of the internal clock. *Annals of the New York Academy of Sciences*, *423*, 566–582. https://doi.org/10.1111/j.1749-6632.1984.tb23459.x

Church, R. M. (1997). Chapter 2 Timing and temporal search. In C. M. Bradshaw & E. Szabadi (Eds.), *Advances in Psychology* (Vol. 120, pp. 41–78). North-Holland. https://doi.org/10.1016/S0166-4115(97)80054-4

Church, R. M. (2002). A tribute to John Gibbon. *Behavioural Processes*, *57*(2–3), 261–274. https://doi.org/10.1016/s0376-6357(02)00018-9

Church, R. M. (2003). A concise introduction to scalar timing theory. In *Functional and neural mechanisms of interval timing* (pp. 3–22). CRC Press/Routledge/Taylor & Francis Group. https://doi.org/10.1201/9780203009574.sec1

Conover, K. L., & Shizgal, P. (19941001a). Competition and summation between rewarding effects of sucrose and lateral hypothalamic stimulation in the rat. *Behavioral Neuroscience*, *108*(3), 537. https://doi.org/10.1037/0735-7044.108.3.537

Conover, K. L., & Shizgal, P. (19941001b). Differential effects of postingestive feedback on the reward value of sucrose and lateral hypothalamic stimulation in rats. *Behavioral Neuroscience*, *108*(3), 559. https://doi.org/10.1037/0735-7044.108.3.559

Conover, K. L., Woodside, B., & Shizgal, P. (1994). Effects of sodium depletion on competition and summation between rewarding effects of salt and lateral hypothalamic stimulation in the rat. *Behavioral Neuroscience*, *108*(3), 549–558. https://doi.org/10.1037//0735-7044.108.3.549

Constantinople, C. M., Piet, A. T., & Brody, C. D. (2019). An Analysis of Decision under Risk in Rats. *Current Biology*, *29*(12), 2066-2074.e5. https://doi.org/10.1016/j.cub.2019.05.013

Cordes, S., & Gallistel, C. R. (2008). Intact Interval Timing in Circadian CLOCK Mutants. *Brain Research*, *1227*, 120–127. https://doi.org/10.1016/j.brainres.2008.06.043

Critchfield, T. S., & Kollins, S. H. (2001). Temporal discounting: Basic research and the analysis of socially important behavior. *Journal of Applied Behavior Analysis*, *34*(1), 101–122. https://doi.org/10.1901/jaba.2001.34-101

Crowley, N. A., Cody, P. A., Davis, M. I., Lovinger, D. M., & Mateo, Y. (2014). Chronic methylphenidate exposure during adolescence reduces striatal synaptic responses to ethanol. *European Journal of Neuroscience*, *39*(4), 548–556. https://doi.org/10.1111/ejn.12426

Davison, M. (1983). Bias and sensitivity to reinforcement in a concurrent-chain schedule. *Journal of the Experimental Analysis of Behavior*, *40*(1), 15–34. https://doi.org/10.1901/jeab.1983.40-15

Davison, M. (1988). Concurrent schedules: Interaction of reinforcer frequency and reinforcer duration. *Journal of the Experimental Analysis of Behavior*, *49*(3), 339–349. https://doi.org/10.1901/jeab.1988.49-339

Davison, M. C. (1976). Preference for fixed-interval schedules: Effects of unequal initial links. *Journal of the Experimental Analysis of Behavior*, *25*(3), 371–376. https://doi.org/10.1901/jeab.1976.25-371

Davison, M., & Hogsden, I. (1984). Concurrent variable-interval schedule performance: Fixed versus mixed reinforcer durations. *Journal of the Experimental Analysis of Behavior*, *41*(2), 169–182. https://doi.org/10.1901/jeab.1984.41-169

Davison, M., & McCarthy, D. (2016). *The Matching Law: A Research Review*. Routledge.

Devilbiss, D. M., & Berridge, C. W. (2008). Cognition-Enhancing Doses of Methylphenidate Preferentially Increase Prefrontal Cortex Neuronal Responsiveness. *Biological Psychiatry*, *64*(7), 626–635. https://doi.org/10.1016/j.biopsych.2008.04.037

Diesmann, M., Gewaltig, M.-O., & Aertsen, A. (1999). Stable propagation of synchronous spiking in cortical neural networks. *Nature*, *402*(6761), 529–533. https://doi.org/10.1038/990101

Durstewitz, D. (2004). Neural representation of interval time. *NeuroReport*, *15*(5), 745–749.

Elliffe, D., Davison, M., & Landon, J. (2008). Relative Reinforcer Rates and Magnitudes Do Not Control Concurrent Choice Independently. *Journal of the Experimental Analysis of Behavior*, *90*(2), 169–185. https://doi.org/10.1901/jeab.2008.90-169

Epstein, R. (1981). Amount consumed as a function of magazine-cycle duration. *Behaviour Analysis Letters*, *1*(1), 63–66.

Estle, S. J., Green, L., Myerson, J., & Holt, D. D. (2006). Differential effects of amount on temporal and probability discounting of gains and losses. *Memory & Cognition*, *34*(4), 914–928. https://doi.org/10.3758/BF03193437

Evenden, J. L., & Ryan, C. N. (1996). The pharmacology of impulsive behaviour in rats: The effects of drugs on response choice with varying delays of reinforcement. *Psychopharmacology*, *128*(2), 161–170. https://doi.org/10.1007/s002130050121

Fouriezos, G., & Randall, D. (1997). The cost of delaying rewarding brain stimulation. *Behavioural Brain Research*, *87*(1), 111–113. https://doi.org/10.1016/s0166-4328(97)02280-8

Frank, R. A., & Stutz, R. M. (1984). Self-deprivation: A review. *Psychological Bulletin*, *96*(2), 384–393. https://doi.org/10.1037/0033-2909.96.2.384

Gallistel, C. R. (1978). Self-stimulation in the rat: Quantitative characteristics of the reward pathway. *Journal of Comparative and Physiological Psychology*, *92*(6), 977. https://doi.org/10.1037/h0077513

Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience* (pp. xvi, 319). Wiley-Blackwell. https://doi.org/10.1002/9781444310498

Gallistel, C. R., & Leon, M. (1991). Measuring the subjective magnitude of brain stimulation reward by titration with rate of reward. *Behavioral Neuroscience*, *105*(6), 913–925.

Gallistel, C. R., Shizgal, P., & Yeomans, J. S. (1981). A portrait of the substrate for self-stimulation. *Psychological Review*, *88*(3), 228–273. https://doi.org/10.1037/0033-295X.88.3.228

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279

Gibbon, J., Fairhurst, S., & Goldberg, B. (1997). Chapter 8 Cooperation, conflict and compromise between circadian and interval clocks in pigeons. In C. M. Bradshaw & E. Szabadi (Eds.), *Advances in Psychology* (Vol. 120, pp. 329–384). North-Holland. https://doi.org/10.1016/S0166-4115(97)80060-X

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience*, *2*(10), 861–863. https://doi.org/10.1038/13158

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L., & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences*, *101*(21), 8174–8179.

Grace, R. (1999). The Matching Law And Amount-dependent Exponential Discounting As Accounts Of Self-control Choice. *Journal of the Experimental Analysis of Behavior*, *71*(1), 27–44. https://doi.org/10.1901/jeab.1999.71-27

Grace, R. C. (1995). Independence of reinforcement delay and magnitude in concurrent chains. *Journal of the Experimental Analysis of Behavior*, *63*(3), 255–276. https://doi.org/10.1901/jeab.1995.63-255

Grace, R. C., Bedell, M. A., & Nevin, J. A. (2002). Preference and resistance to change with constant- and variable-duration terminal links: Independence of reinforcement rate and magnitude. *Journal of the Experimental Analysis of Behavior*, *77*(3), 233–255. https://doi.org/10.1901/jeab.2002.77-233

Green, L., Myerson, J., Holt, D. D., Slevin, J. R., & Estle, S. J. (2004). Discounting of Delayed Food Rewards in Pigeons and Rats: Is There a Magnitude Effect? *Journal of the Experimental Analysis of Behavior*, *81*(1), 39–50. https://doi.org/10.1901/jeab.2004.81-39

Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, *25*, 715–723. https://doi.org/10.3758/BF03211314

Green, L., & Snyderman, M. (1980). Choice between rewards differing in amount and delay: Toward a choice model of self control. *Journal of the Experimental Analysis of Behavior*, *34*(2), 135–147. https://doi.org/10.1901/jeab.1980.34-135

Grossberg, S., & Schmajuk, N. A. (1989). Neural dynamics of adaptive timing and temporal discrimination during associative learning. *Neural Networks*, *2*(2), 79–102. https://doi.org/10.1016/0893-6080(89)90026-9

Hahnloser, R. H. R., Kozhevnikov, A. A., & Fee, M. S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature*, *419*(6902), 65–70. https://doi.org/10.1038/nature00974

Harper, D. G. C. (1982). Competitive foraging in mallards: "Ideal free' ducks. *Animal Behaviour*, *30*(2), 575–584. https://doi.org/10.1016/S0003-3472(82)80071-7

Hernandez, G., Breton, Y.-A., Conover, K., & Shizgal, P. (2010). At What Stage of Neural Processing Does Cocaine Act to Boost Pursuit of Rewards? *PLOS ONE*, *5*(11), e15081. https://doi.org/10.1371/journal.pone.0015081

Hernandez, G., Oleson, E. B., Gentry, R. N., Abbas, Z., Bernstein, D. L., Arvanitogiannis, A., & Cheer, J. F. (2014). Endocannabinoids Promote Cocaine-Induced Impulsivity and Its Rapid Dopaminergic Correlates. *Biological Psychiatry*, *75*(6), 487–498. https://doi.org/10.1016/j.biopsych.2013.09.005

Hernandez, G., Trujillo-Pisanty, I., Cossette, M.-P., Conover, K., & Shizgal, P. (2012). Role of Dopamine Tone in the Pursuit of Brain Stimulation Reward. *The Journal of Neuroscience*, *32*(32), 11032–11041. https://doi.org/10.1523/JNEUROSCI.1051-12.2012

Herrnstein, R. J. (1961). Relative and Absolute Strength of Response as a Function of Frequency of Reinforcement1,2. *Journal of the Experimental Analysis of Behavior*, *4*(3), 267–272. https://doi.org/10.1901/jeab.1961.4-267

Herrnstein, R. J., & Loveland, D. H. (1975). Maximizing and matching on concurrent ratio schedules. *Journal of the Experimental Analysis of Behavior*, *24*(1), 107–116. https://doi.org/10.1901/jeab.1975.24-107

Herzog, E. D., Aton, S. J., Numano, R., Sakaki, Y., & Tei, H. (2004). Temporal precision in the mammalian circadian system: A reliable clock from less reliable neurons. *Journal of Biological Rhythms*, *19*(1), 35–46. https://doi.org/10.1177/0748730403260776

Higa, J. J., & Staddon, J. E. R. (1997). Chapter 1 Dynamic models of rapid temporal control in animals. In C. M. Bradshaw & E. Szabadi (Eds.), *Advances in Psychology* (Vol. 120, pp. 1–40). North-Holland. https://doi.org/10.1016/S0166-4115(97)80053-2

Ito, M., & Asaki, K. (1982). Choice behavior of rats in a concurrent-chains schedule: Amount and delay of reinforcement. *Journal of the Experimental Analysis of Behavior*, *37*(3), 383–392. https://doi.org/10.1901/jeab.1982.37-383

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *12*(7), 273–280. https://doi.org/10.1016/j.tics.2008.04.002

Johansson, F., Jirenhed, D.-A., Rasmussen, A., Zucca, R., & Hesslow, G. (2014). Memory trace and timing mechanism localized to cerebellar Purkinje cells. *Proceedings of the National Academy of Sciences*, *111*(41), 14930–14934. https://doi.org/10.1073/pnas.1415371111

Johnson, H. A., Goel, A., & Buonomano, D. V. (2010). Neural dynamics of in vitro cortical networks reflects experienced temporal patterns. *Nature Neuroscience*, *13*(8), 917–919. https://doi.org/10.1038/nn.2579

Karmarkar, U. R., & Buonomano, D. V. (2007). Telling time in the absence of clocks. *Neuron*, *53*(3), 427–438. https://doi.org/10.1016/j.neuron.2007.01.006

Keller, J. V., & Gollub, L. R. (1977). Duration and rate of reinforcement as determinants of concurrent responding. *Journal of the Experimental Analysis of Behavior*, *28*(2), 145–153. https://doi.org/10.1901/jeab.1977.28-145

Kheramin, S., Body, S., Ho, M.-Y., Velázquez-Martinez, D. N., Bradshaw, C. M., Szabadi, E., Deakin, J. F. W., & Anderson, I. M. (2004). Effects of orbital prefrontal cortex dopamine depletion on inter-temporal choice: A quantitative analysis. *Psychopharmacology*, *175*(2), 206–214. https://doi.org/10.1007/s00213-004-1813-y

Killeen, P. R. (1985). Incentive theory: IV. Magnitude of reward. *Journal of the Experimental Analysis of Behavior*, *43*(3), 407–417. https://doi.org/10.1901/jeab.1985.43-407

Kitano, K., Okamoto, H., & Fukai, T. (2003). Time representing cortical activities: Two models inspired by prefrontal persistent activity. *Biological Cybernetics*, *88*(5), 387–394. https://doi.org/10.1007/s00422-002-0390-6

Koob, G. F., & Moal, M. L. (1997). Drug Abuse: Hedonic Homeostatic Dysregulation. *Science*, *278*(5335), 52–58. https://doi.org/10.1126/science.278.5335.52

Kreitzman, L., & Foster, R. (2010). *Seasons of Life: The biological rhythms that enable living things to thrive and survive*. Profile Books.

Leo, D., Adriani, W., Cavaliere, C., Cirillo, G., Marco, E. M., Romano, E., Porzio, U. D., Papa, M., Perrone-Capano, C., & Laviola, G. (2009). Methylphenidate to adolescent rats drives enduring changes of accumbal Htr7 expression: Implications for impulsive behavior and

neuronal morphology. *Genes, Brain and Behavior*, *8*(3), 356–368. https://doi.org/10.1111/j.1601-183X.2009.00486.x

Leon, M., & Gallistel, C. R. (1992). The function relating the subjective magnitude of brain stimulation reward to stimulation strength varies with site of stimulation. *Behavioural Brain Research*, *52*(2), 183–193. https://doi.org/10.1016/s0166-4328(05)80229-3

Leon, M. I., & Gallistel, C. R. (1998). Self-stimulating rats combine subjective reward magnitude and subjective reward rate multiplicatively. *Journal of Experimental Psychology. Animal Behavior Processes*, *24*(3), 265–277. https://doi.org/10.1037//0097-7403.24.3.265

Lewis, P. A., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: Evidence from neuroimaging. *Current Opinion in Neurobiology*, *13*(2), 250–255. https://doi.org/10.1016/S0959-4388(03)00036-9

Lewis, P. A., Miall, R. C., Daan, S., & Kacelnik, A. (2003). Interval timing in mice does not rely upon the circadian pacemaker. *Neuroscience Letters*, *348*(3), 131–134. https://doi.org/10.1016/s0304-3940(03)00521-4

Logue, A. W., & Chavarro, A. (1987). Effect on choice of absolute and relative values of reinforcer delay, amount, and frequency. *Journal of Experimental Psychology: Animal Behavior Processes*, *13*(3), 280–291. https://doi.org/10.1037/0097-7403.13.3.280

Logue, A. W., Tobin, H., Chelonis, J. J., Wang, R. Y., Geary, N., & Schachter, S. (1992). Cocaine decreases self-control in rats: A preliminary report. *Psychopharmacology*, *109*(1), 245–247. https://doi.org/10.1007/BF02245509

Long, M. A., & Fee, M. S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature*, *456*(7219), 189–194. https://doi.org/10.1038/nature07448

Low, K. G., & Gendaszek, A. E. (2002). Illicit use of psychostimulants among college students: A preliminary study. *Psychology, Health & Medicine*, *7*(3), 283–287. https://doi.org/10.1080/13548500220139386

Macar, F., Lejeune, H., Bonnet, M., Ferrara, A., Pouthas, V., Vidal, F., & Maquet, P. (2002). Activation of the supplementary motor area and of attentional networks during temporal processing. *Experimental Brain Research*, *142*(4), 475–485. https://doi.org/10.1007/s00221-001-0953-0

Manitt, C., Mimee, A., Eng, C., Pokinko, M., Stroh, T., Cooper, H. M., Kolb, B., & Flores, C. (2011). The Netrin Receptor DCC Is Required in the Pubertal Organization of Mesocortical Dopamine Circuitry. *Journal of Neuroscience*, *31*(23), 8381–8394.

Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: Coincidence detection of oscillatory processes. *Brain Research. Cognitive Brain Research*, *21*(2), 139–170. https://doi.org/10.1016/j.cogbrainres.2004.06.012

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In *The effect of delay and of intervening events on reinforcement value* (pp. 55–73). Lawrence Erlbaum Associates, Inc.

Mazur, J. E., Stellar, J. R., & Waraczynski, M. (1987). Self-control choice with electrical stimulation of the brain as a reinforcer. *Behavioural Processes*, *15*(2–3), 143–153. https://doi.org/10.1016/0376-6357(87)90003-9

McDowell, J. J. (2005). On The Classic And Modern Theories Of Matching. *Journal of the Experimental Analysis of Behavior*, *84*(1), 111–127. https://doi.org/10.1901/jeab.2005.59-04

McLean, A. P., & Blampied, N. M. (2001). Sensitivity to relative reinforcer rate in concurrent schedules: Independence from relative and absolute reinforcer duration. *Journal of the Experimental Analysis of Behavior*, *75*(1), 25–42. https://doi.org/10.1901/jeab.2001.75-25

Medina, J. F., Garcia, K. S., Nores, W. L., Taylor, N. M., & Mauk, M. D. (2000). Timing mechanisms in the cerebellum: Testing predictions of a large-scale computer simulation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(14), 5516–5525.

Michon, J. A. (1985). The Compleat Time Experiencer. In J. A. Michon & J. L. Jackson (Eds.), *Time, Mind, and Behavior* (pp. 20–52). Springer. https://doi.org/10.1007/978-3-642-70491-8_2

Moll, G. H., Hause, S., Rüther, E., Rothenberger, A., & Huether, G. (2001). Early Methylphenidate Administration to Young Rats Causes a Persistent Reduction in the Density of Striatal Dopamine Transporters. *Journal of Child and Adolescent Psychopharmacology*, *11*(1), 15–24. https://doi.org/10.1089/104454601750143366

Moore, J. W. (Ed.). (2002). *A Neuroscientist's Guide to Classical Conditioning*. Springer-Verlag. https://doi.org/10.1007/978-1-4419-8558-3

Naneix, F., Marchand, A. R., Di Scala, G., Pape, J.-R., & Coutureau, E. (2012). Parallel Maturation of Goal-Directed Behavior and Dopaminergic Systems during Adolescence. *Journal of Neuroscience*, *32*(46), 16223–16232.

Navarick, D. J., & Fantino, E. (1976). Self-control and general models of choice. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*(1), 75–87. https://doi.org/10.1037/0097-7403.2.1.75

Neuringer, A. J. (1967). Effects of reinforcement magnitude on choice and rate of responding. *Journal of the Experimental Analysis of Behavior*, *10*(5), 417–424. https://doi.org/10.1901/jeab.1967.10-417

Nieuwenhuys, R., Geeraedts, L. M. G., & Veening, J. G. (1982). The medial forebrain bundle of the rat. I. General introduction. *Journal of Comparative Neurology*, *206*(1), 49–81. https://doi.org/10.1002/cne.902060106

Odum, A. L., & Baumann, A. A. L. (2010). Delay discounting: State and trait variable. In *Impulsivity: The behavioral and neurological science of discounting* (pp. 39–65). American Psychological Association. https://doi.org/10.1037/12069-002

Olds, J. (20060327). Satiation effects in self-stimulation of the brain. *Journal of Comparative and Physiological Psychology*, *51*(6), 675. https://doi.org/10.1037/h0039616

Ong, E. L., & White, K. G. (2004). Amount-dependent temporal discounting? *Behavioural Processes*, *66*(3), 201–212. https://doi.org/10.1016/j.beproc.2004.03.005

Paine, T. A., Dringenberg, H. C., & Olmstead, M. C. (2003). Effects of chronic cocaine on impulsivity: Relation to cortical serotonin mechanisms. *Behavioural Brain Research*, *147*(1–2), 135–147. https://doi.org/10.1016/s0166-4328(03)00156-6

Papachristos, E. B., Jacobs, E. H., & Elgersma, Y. (2011). Interval timing is intact in arrhythmic Cry1/Cry2-deficient mice. *Journal of Biological Rhythms*, *26*(4), 305–313. https://doi.org/10.1177/0748730411410026

Pardey, M. C., Kumar, N. N., Goodchild, A. K., Clemens, K. J., Homewood, J., & Cornish, J. L. (2012). Long-Term Effects of Chronic Oral Ritalin Administration on Cognitive and Neural Development in Adolescent Wistar Kyoto Rats. *Brain Sciences*, *2*(3), 375–404. https://doi.org/10.3390/brainsci2030375

Patke, A., Young, M. W., & Axelrod, S. (2020). Molecular mechanisms and physiological importance of circadian rhythms. *Nature Reviews Molecular Cell Biology*, *21*(2), 67–84. https://doi.org/10.1038/s41580-019-0179-2

Pietras, C. J., Cherek, D. R., Lane, S. D., Tcheremissine, O. V., & Steinberg, J. L. (2003). Effects of methylphenidate on impulsive choice in adult humans. *Psychopharmacology*, *170*(4), 390–398. https://doi.org/10.1007/s00213-003-1547-2

Pine, A., Shiner, T., Seymour, B., & Dolan, R. J. (2010). Dopamine, Time, and Impulsivity in Humans. *Journal of Neuroscience*, *30*(26), 8888–8896.

Pitts, R. C., & McKinney, A. P. (2005). Effects of methylphenidate and morphine on delay-discount functions obtained within sessions. *Journal of the Experimental Analysis of Behavior*, *83*(3), 297–314. https://doi.org/10.1901/jeab.2005.47-04

Plaud, J. J. (1992). The prediction and control of behavior revisited: A review of the matching law. *Journal of Behavior Therapy and Experimental Psychiatry*, *23*(1), 25–31. https://doi.org/10.1016/0005-7916(92)90021-A

Rachlin, H. (2006). Notes on Discounting. *Journal of the Experimental Analysis of Behavior*, *85*(3), 425–435. https://doi.org/10.1901/jeab.2006.85-05

Rachlin, H., & Baum, W. M. (1972). Effects of Alternative Reinforcement: Does the Source Matter?1. *Journal of the Experimental Analysis of Behavior*, *18*(2), 231–241. https://doi.org/10.1901/jeab.1972.18-231

Rachlin, H., & Green, L. (1972). Commitment, Choice and Self-Control1. *Journal of the Experimental Analysis of Behavior*, *17*(1), 15–22. https://doi.org/10.1901/jeab.1972.17-15

Raible, F., Takekata, H., & Tessmar-Raible, K. (2017). An Overview of Monthly Rhythms and Clocks. *Frontiers in Neurology*, *8*. https://doi.org/10.3389/fneur.2017.00189

Raineri, A., & Rachlin, H. (1993). The effect of temporal constraints on the value of money and other commodities. *Journal of Behavioral Decision Making*, *6*(2), 77–94. https://doi.org/10.1002/bdm.3960060202

Rammsayer, T. H. (1999). Neuropharmacological evidence for different timing mechanisms in humans. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *52*(3), 273–286. https://doi.org/10.1080/713932708

Rammsayer, T. H., & Lima, S. D. (1991). Duration discrimination of filled and empty auditory intervals: Cognitive and perceptual factors. *Perception & Psychophysics*, *50*(6), 565–574. https://doi.org/10.3758/bf03207541

Reed, D. D., & Kaplan, B. A. (2011). The Matching Law: A Tutorial for Practitioners. *Behavior Analysis in Practice*, *4*(2), 15–24. https://doi.org/10.1007/BF03391780

Reppert, S. M., & Weaver, D. R. (2002). Coordination of circadian timing in mammals. *Nature*, *418*(6901), 935–941. https://doi.org/10.1038/nature00965

Reynolds, L. M., Makowski, C. S., Yogendran, S. V., Kiessling, S., Cermakian, N., & Flores, C. (2015). Amphetamine in Adolescence Disrupts the Development of Medial Prefrontal Cortex Dopamine Connectivity in a dcc -Dependent Manner. *Neuropsychopharmacology*, *40*(5), 1101–1112. https://doi.org/10.1038/npp.2014.287

Richards, J. B., Mitchell, S. H., de Wit, H., & Seiden, L. S. (1997). Determination of discount functions in rats with an adjusting-amount procedure. *Journal of the Experimental Analysis of Behavior*, *67*(3), 353–366. https://doi.org/10.1901/jeab.1997.67-353

Rider, D. P. (1983). Choice for aperiodic versus periodic ratio schedules: A comparison of concurrent and concurrent-chain procedures. *Journal of the Experimental Analysis of Behavior*, *40*(3), 225–237. https://doi.org/10.1901/jeab.1983.40-225

Rodriguez, M. L., & Logue, A. W. (1986). Independence of the amount and delay ratios in the generalized matching law. *Animal Learning & Behavior*, *14*(1), 29–37. https://doi.org/10.3758/BF03200034

Schibler, U., & Sassone-Corsi, P. (2002). A web of circadian pacemakers. *Cell*, *111*(7), 919–922. https://doi.org/10.1016/s0092-8674(02)01225-4

Shiels, K., Hawk, L. W., Reynolds, B., Mazzullo, R., Rhodes, J., Pelham, W. E., Waxmonsky, J. G., & Gangloff, B. P. (2009). The Effects of Methylphenidate on Discounting of Delayed Rewards in ADHD. *Experimental and Clinical Psychopharmacology*, *17*(5), 291–301. https://doi.org/10.1037/a0017259

Shizgal, P. (1997). Neural basis of utility estimation. *Current Opinion in Neurobiology*, *7*(2), 198–208. https://doi.org/10.1016/S0959-4388(97)80008-6

Shuler, M. G., & Bear, M. F. (2006). Reward Timing in the Primary Visual Cortex. *Science*, *311*(5767), 1606–1609. https://doi.org/10.1126/science.1123513

Shull, R. L., & Pliskoff, S. S. (1967). Changeover delay and concurrent schedules: Some effects on relative performance measures. *Journal of the Experimental Analysis of Behavior*, *10*(6), 517–527. https://doi.org/10.1901/jeab.1967.10-517

Simmons, J. M., & Gallistel, C. R. (1994). Saturation of subjective reward magnitude as a function of current and pulse frequency. *Behavioral Neuroscience*, *108*(1), 151–160. https://doi.org/10.1037//0735-7044.108.1.151

Simon, N. W., Mendez, I. A., & Setlow, B. (2007). Cocaine exposure causes long-term increases in impulsive choice. *Behavioral Neuroscience*, *121*(3), 543–549. https://doi.org/10.1037/0735-7044.121.3.543

Snyderman, M. (1983). Delay and amount of reward in a concurrent chain. *Journal of the Experimental Analysis of Behavior*, *39*(3), 437–447. https://doi.org/10.1901/jeab.1983.39-437

Solomon, R. B., Trujillo-Pisanty, I., Conover, K., & Shizgal, P. (2015). Psychophysical inference of frequency-following fidelity in the neural substrate for brain stimulation reward. *Behavioural Brain Research*, *292*, 327–341. https://doi.org/10.1016/j.bbr.2015.06.008

Somkuwar, S. S., Jordan, C. J., Kantak, K. M., & Dwoskin, L. P. (2013). Adolescent Atomoxetine Treatment in a Rodent Model of ADHD: Effects on Cocaine Self-Administration and Dopamine Transporters in Frontostriatal Regions. *Neuropsychopharmacology*, *38*(13), 2588–2597. https://doi.org/10.1038/npp.2013.163

Sonnenschein, B., Conover, K., & Shizgal, P. (2003). Growth of brain stimulation reward as a function of duration and stimulation strength. *Behavioral Neuroscience*, *117*(5), 978–994. https://doi.org/10.1037/0735-7044.117.5.978

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience & Biobehavioral Reviews*, *24*(4), 417–463. https://doi.org/10.1016/S0149-7634(00)00014-2

Squires, N., & Fantino, E. (1971). A model for choice in simple concurrent and concurrent-chains schedules. *Journal of the Experimental Analysis of Behavior*, *15*(1), 27–38. https://doi.org/10.1901/jeab.1971.15-27

Staddon, J. E., & Higa, J. J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior*, *71*(2), 215–251. https://doi.org/10.1901/jeab.1999.71-215

Thornquist, S. C., Langer, K., Zhang, S. X., Rogulja, D., & Crickmore, M. A. (2020). CaMKII Measures the Passage of Time to Coordinate Behavior and Motivational State. *Neuron*, *105*(2), 334-345.e9. https://doi.org/10.1016/j.neuron.2019.10.018

Thornquist, S. C., Pitsch, M. J., Auth, C. S., & Crickmore, M. A. (2021). Biochemical evidence accumulates across neurons to drive a network-level eruption. *Molecular Cell*, *81*(4), 675-690.e8. https://doi.org/10.1016/j.molcel.2020.12.029

Tieu, K. H., Keidel, A. L., McGann, J. P., Faulkner, B., & Brown, T. H. (1999). Perirhinal-amygdala circuit-level computational model of temporal encoding in fear conditioning. *Psychobiology*, *27*(1), 1–25. https://doi.org/10.3758/BF03332095

Todorov, J. C. (1973). Interaction of frequency and magnitude of reinforcement on concurrent performances. *Journal of the Experimental Analysis of Behavior*, *19*(3), 451–458. https://doi.org/10.1901/jeab.1973.19-451

Trujillo-Pisanty, I., Conover, K., & Shizgal, P. (2013). A new view of the effect of dopamine receptor antagonism on operant performance for rewarding brain stimulation in the rat. *Psychopharmacology*. https://doi.org/10.1007/s00213-013-3328-x

Trujillo-Pisanty, I., Conover, K., Solis, P., Palacios, D., & Shizgal, P. (2020). Dopamine neurons do not constitute an obligatory stage in the final common path for the evaluation and pursuit of brain stimulation reward. *PloS One*, *15*(6), e0226722. https://doi.org/10.1371/journal.pone.0226722

Trujillo-Pisanty, I., Hernandez, G., Moreau-Debord, I., Cossette, M.-P., Conover, K., Cheer, J. F., & Shizgal, P. (2011). Cannabinoid receptor blockade reduces the opportunity cost at which rats maintain operant performance for rewarding brain stimulation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(14), 5426–5435. https://doi.org/10.1523/JNEUROSCI.0079-11.2011

Urban, K. R., & Gao, W.-J. (2015). Evolution of the Study of Methylphenidate and Its Actions on the Adult Versus Juvenile Brain. *Journal of Attention Disorders*, *19*(7), 603–619. https://doi.org/10.1177/1087054712455504

van der Marel, K., Klomp, A., Meerhoff, G. F., Schipper, P., Lucassen, P. J., Homberg, J. R., Dijkhuizen, R. M., & Reneman, L. (2014). Long-Term Oral Methylphenidate Treatment in Adolescent and Adult Rats: Differential Effects on Brain Morphology and Function. *Neuropsychopharmacology*, *39*(2), 263–273. https://doi.org/10.1038/npp.2013.169

van Gaalen, M. M., van Koten, R., Schoffelmeer, A. N. M., & Vanderschuren, L. J. M. J. (2006). Critical involvement of dopaminergic neurotransmission in impulsive decision making. *Biological Psychiatry*, *60*(1), 66–73. https://doi.org/10.1016/j.biopsych.2005.06.005

van Rijn, H., Gu, B.-M., & Meck, W. H. (2014). Dedicated clock/timing-circuit theories of time perception and timed performance. *Advances in Experimental Medicine and Biology*, *829*, 75–99. https://doi.org/10.1007/978-1-4939-1782-2_5

Vanderveldt, A., Oliveira, L., & Green, L. (2016). Delay Discounting: Pigeon, Rat, Human – Does it Matter? *Journal of Experimental Psychology. Animal Learning and Cognition*, *42*(2), 141–162. https://doi.org/10.1037/xan0000097

Vendruscolo, L. F., Izídio, G. S., Takahashi, R. N., & Ramos, A. (2008). Chronic methylphenidate treatment during adolescence increases anxiety-related behaviors and ethanol drinking in adult spontaneously hypertensive rats. *Behavioural Pharmacology*, *19*(1), 21–27. https://doi.org/10.1097/FBP.0b013e3282f3cfbe

Ward, R. D., Gallistel, C. R., & Balsam, P. D. (2013). It's The Information! *Behavioural Processes*, *95*, 3–7. https://doi.org/10.1016/j.beproc.2013.01.005

Welsh, D. K., Engel, E. M., Richardson, G. S., Dement, W. C., & Engle, E. M. (1986). Precision of circadian wake and activity onset timing in the mouse. *Journal of Comparative*

Physiology. A, Sensory, Neural, and Behavioral Physiology, *158*(6), 827–834. https://doi.org/10.1007/BF01324824

White, K. G., & Pipe, M.-E. (1987). Sensitivity to reinforcer duration in a self-control procedure. *Journal of the Experimental Analysis of Behavior*, *48*(2), 235–249. https://doi.org/10.1901/jeab.1987.48-235

Winstanley, C. (2010). The neural and neurochemical basis of delay discounting. In Madden, G.J & Bickel, W.K., *Impulsivity: The behavioral and neurological science of discounting* (pp. 95–121). American Psychological Association. https://content.apa.org/record/2009-18177-004

Wit, H., & Mitchell, S. (2010). Drug effects on delay discounting. *Impulsivity: The Behavioral and Neurological Science of Discounting*, 213–241. https://doi.org/10.1037/12069-008

Yamazaki, T., & Tanaka, S. (2005). Neural modeling of an internal clock. *Neural Computation*, *17*(5), 1032–1058. https://doi.org/10.1162/0899766053491850

Zeeb, F. D., Floresco, S. B., & Winstanley, C. A. (2010). Contributions of the orbitofrontal cortex to impulsive choice: Interactions with basal levels of impulsivity, dopamine signalling, and reward-related cues. *Psychopharmacology*, *211*(1), 87–98. https://doi.org/10.1007/s00213-010-1871-2

Zosel, A., Bartelson, B. B., Bailey, E., Lowenstein, S., & Dart, R. (2013). Characterization of adolescent prescription drug abuse and misuse using the Researched Abuse Diversion and Addiction-related Surveillance (RADARS(®)) System. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*(2), 196-204.e2. https://doi.org/10.1016/j.jaac.2012.11.014