Survey of time preference, delay discounting models

John R. Doyle*

Abstract

The paper surveys over twenty models of *delay discounting (*also known as *temporal discounting*, *time preference, time discounting*), that psychologists and economists have put forward to explain the way people actually trade off time and money. Using little more than the basic algebra of powers and logarithms, I show how the models are derived, what assumptions they are based upon, and how different models relate to each other. Rather than concentrate only on *discount functions* themselves, I show how discount functions may be manipulated to isolate *rate parameters* for each model. This approach, consistently applied, helps focus attention on the three main components in any discounting model: subjectively perceived money; subjectively perceived time; and how these elements are combined. We group models by the number of parameters that have to be estimated, which means our exposition follows a trajectory of increasing complexity to the models. However, as the story unfolds it becomes clear that most models fall into a smaller number of families. We also show how new models may be constructed by combining elements of different models.

The surveyed models are: Exponential; Hyperbolic; Arithmetic; Hyperboloid (Green & Myerson, Rachlin); Loewenstein and Prelec Generalized Hyperboloid; quasi-Hyperbolic (also known as β-δ discounting); Benhabib et al's fixed cost; Benhabib et al's Exponential / Hyperbolic / quasi-Hyperbolic; Read's discounting fractions; Roelofsma's exponential time; Scholten and Read's discounting-by-intervals (DBI); Ebert and Prelec's constant sensitivity (CS); Bleichrodt et al.'s constant absolute decreasing impatience (CADI); Bleichrodt et al.'s constant relative decreasing impatience (CRDI); Green, Myerson, and Macaux's hyperboloid over intervals models; Killeen's additive utility; size-sensitive additive utility; Yi, Landes, and Bickel's memory trace models; McClure et al.'s two exponentials; and Scholten and Read's trade-off model.

For a convenient overview, a single "cheat sheet" table captures the notation and essential mathematics behind all but one of the models.

Keywords:

1 Introduction

1.1 Delay discounting, time preference, and what this survey is and is not.

People are constantly making decisions that involve whether they take gains (also losses) now or at some later time(s). How people actually achieve this, either as individuals or collectively as a segment of an economy, is an active area of research that straddles the boundaries of psychology, economics, marketing, decision analysis, and more recently neuroscience. It goes by several names, some generally preferred by psychologists (*delay discounting)*, others by economists (*time preference)*. Although widely applicable, this survey will limit itself to delay discounting for simple financial decisions, not just because these are an important class of problems that people face, but also for clarity of presentation. The same models may be used in non-financial applications.

The following are examples that illustrate this claim in the context of consumer research. Estle, Green, Myerson, and Holt (2007), for instance, compared monetary with consumable rewards (candy, soda, beer), finding that the latter were discounted more steeply. Chapman and Elstein (1995), used formal models of delay discounting where rewards were presented as days on vacation (and years of health in another experiment). Others have focused on impulsivity as a personality trait, a theme particularly researched in the area of out-of-control consumption behaviors such as drug addiction, smoking, gambling, alcoholism, and so on (Madden & Bickel, 2009); but also in discounting's more prosaic connection with numerical ability (Peters, Slovic, Västfjäll, & Mertz, 2008). Focusing on the situation (i.e., the kind of commodity being discounted) rather than the person, other researchers have examined the extent to which discounting is domain specific in normal (i.e., non-addicted) people who nonetheless also experience the temptations of consumption (Hardisty & Weber, 2009; Tsukayama & Duckworth, 2010; Weatherly, Terrell, & Derenne, 2010). Yet others have examined factors that lead people to

*Cardiff Business School, Aberconway Building, Colum Drive, Cardiff CF10 3EU, U.K. Email: DoyleJR@cardiff.ac.uk.

behave more impulsively across disconnected domains: e.g., Wilson and Daly (2004) found that seeing pictures of the faces of attractive women induced men to discount money more steeply than if the faces were unattractive. Van den Bergh, Dewitte, and Warlop (2008) reached similar conclusions for men who were asked to handle bikinis. However, the precision with which these diverse examples may be understood depends on the accuracy of the models that are used to measure delay discounting. Because these models lie scattered in many different literatures, even researchers who are active in the field may be unaware of interesting new developments, which is one value of this survey.

The most general principle in delay discounting is that people prefer to be given money now rather than later, so that in subjectively valuing timed rewards, future payments are described as being *discounted* (by the fact of being delayed). This paper is a survey of delay discounting models. It would have been a brief survey, in that none of the models is examine in great depth, were it not for the large number of models that exist in the literature. Although presented from the perspective of gains, each model may also be applied to losses. To keep the survey slim I have generally not amplified where and how the models have been used, nor with what success. Nor have I surveyed empirical comparisons between models, because apart from comparisons between Hyperbolic and Exponential, and between quasi-Hyperbolic and Exponential, the literature contains few face-to-face comparisons.

Furthermore, whether a model is considered viable often depends on the discipline of the researcher (economists are much more likely to adhere to the Exponential model than psychologists). Viability may also depend on any of the following: individual differences—whether correlates of those difference are identifiable, such as the level of numeracy of the person, or not ; whether losses or gains are being considered; the method of elicitation—sometimes the basis of economists' arguments against non-exponential models (e.g., Andersen, Harrison, Lau, & Rutström, 2011); the culture in which the problem is posed; the nature of the choices being offered (money, bottles of beer, years of health); experimental manipulations—for instance that might make differences versus ratios salient.

Other interesting and important themes that I largely ignore are: the cognitive processes that people would need to achieve in order to discount in a particular way; the efficacy of using multi-item choice questionnaires (Kirby & Marakovic, 1996) compared with simple matching (Smith & Hantula, 2008), a hybrid of the two (Rachlin, Raineri & Cross, 1991), or even reaction times (Chabris, Laibson, Morris, Schuldt & Taubinsky, 2008); the design of stimuli to improve statistical power

to detect differences between models (Doyle, Chen & Savani, 2011); measurement of model goodness-of-fit and impulsivity (Myerson, Green & Warusawitharana, 2001; Wileyto, Audrain-McGovern, Epstein & Lerman, 2004; Doyle & Chen, 2012); how models should be compared (Killeen, 2009; Doyle & Chen, 2012); the parallels between risky choice and intertemporal choice (Prelec & Loewenstein, 1991; Green & Myerson, 2004; Rachlin, 2006). The paper escapes being a mere catalogue, in that I have aimed to make the structure of the models transparent, comparing them through their mathematics and what implications each has if taken seriously as a model of human behavior. Having disassembled the models into their component parts, I emphasize their role as building blocks by showing how it is possible to re-assemble the components to create novel, speculative models: a process that future researchers may continue.

The paper has the following structure. The introduction argues for a new focus in delay discounting in terms of rate parameter (Section 1.2); describes power laws, log laws, and related mathematical issues (Section 1.3); and the conventions of notation to be adopted (Section 1.4). Sections 2 through 5 describe delay discounting models with zero through three free parameters, respectively. Section 6 describes models not easily classified elsewhere. Section 7 provides a retrospective grouping of models into families, and shows how speculative models may be generated. Conclusions are in Section 8.

1.2 Orientation: discount functions or rate parameter?

There are two strategies for presenting discounting models. In one, a choice between P (present choice) now and F (future choice) later, it is presumed that the F is discounted to its net present value (NPV) and compared with P, and people choose F if its discounted value is greater than P. This is the usual view, from which "delay discounting" "time discounting" or "temporal discounting" take their name. It focuses attention on the discount function that maps an F into a P. In simple models F is multiplied by a *discount factor*, which is a particular realization of the discount function for a given time delay, to derive the notionally equivalent P that should make a decision maker indifferent between F and P.

An alternative strategy is to rearrange the discounting equation and focus on the rate parameter implied by a given discounting model (e.g., in the exponential model, the rate parameter is the continuously compounded interest rate that would make a P now into an F at time T). This survey adopts the rate parameter view. The advantages are that the components of the discounting model are now quite transparent: how the model treats money, how the model treats time, and how the treated time and

money are composed into the final formula. Discount factors increase monotonically with rate parameters, so there is no conflict between the two views in terms of the preference ordering they compute for a set of choices. In the rate parameter view, people are presumed to have, at least during testing, a stable internally held rate that acts as a decision criterion: people choose F if, relative to P, F implies a higher rate than that criterion rate. Seen in this way, rate parameters are also decision parameters. People who are impulsive are identified by having high internal criterion rates. One discounting model is preferred over another as an account of behavior if it more correctly estimates the preference order of the choices on offer. There seems no prior reason to prefer the discounting view over the rate parameter view, or vice versa. Therefore, one task for future research is to determine whether people adopt a discounting view or rate parameter view of intertemporal problems, or indeed another view. Although the views are mathematically equivalent, in bringing certain aspects of the problem to the fore while pushing other aspects into the background, different views may not be psychologically equivalent in terms of the kinds of mental operations required to run a model, the heuristics that might be used to short-cut full computation, and therefore the typical kinds of deviations that people would produce from their true intentions[1].

In this survey I classify models somewhat superficially by the number of additional free parameters that an experimental design would need to estimate over and above the rate parameter. I build up the complexity of the models. In the end, however, I am able to show that most models belong to a few families, from which simpler models are special cases. I also speculate on future models.

1.3   Stevens, Weber, powers, logs, ratios, differences, and started time

Most of the models described in this survey can be classified by their treatment of three main issues: (i) (how) is the subjective perception of money to be treated; (ii) (how) is the subjective perception of time to be treated; and (iii) (how) is money to be traded-off against time? In deciding these questions, I will be making repeated use of a few psychological principles and related mathematical relationships. Though they appear throughout the paper, it is worth collecting them together before I begin.

---

[1]Analogously, although $8! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$, if someone adopts the former description (ascending sequence) they tend to estimate smaller numbers than if they adopt the mathematically equivalent descending sequence description (Tversky & Kahneman, 1974). Hence, although delay discounting and rate parameter descriptions are mathematically equivalent, they may not be equivalent in their consequences.

1.3.1   Stevens versus Weber

A long-standing debate in psychology has been whether basic psychophysical phenomena such as the subjective brightness of light, heaviness, sound, pain are best modeled by using logs or powers. The Weber-Fechner "law" dates back to the 18th century and promotes a log treatment. The law assumes that the subjective change in a stimulus (ΔS) has a constant relation with the change in objective energy (ΔE) of a stimulus relative to the existing energy (E).

$$\Delta S = k \, \Delta E \, / \, E \qquad (1)$$

As an example, many candles will be needed at noon on a sunny day to give the same subjective increase in brightness experienced by lighting a single candle at twilight. Integrating, we have

$$S = k \, log(E) \qquad (2)$$

which is the Weber-Fechner law: Weber associated with form (1), Fechner with form (2). (Log base e is assumed throughout this paper.) Its main challenger is the power "law", which instead suggests that:

$$S = K \, E^c \qquad (3)$$

where c depends on the context. For instance, Stevens (1957; 1961) documented different power laws for different psychophysical phenomena. Subjective loudness is proportional to (objective sound pressure)$^{.67}$, whereas subjective pain is proportional to (electrical shock to finger)$^{3.5}$, and so on. It is worth noting that in Wikipedia's entry for *Stevens' Power Law*, there is an approximately even split between exponents > 1, and exponents < 1. The models surveyed in this paper assume that the abstract concepts of time and money (and even time intervals) may be treated in the same way as physical stimuli, by using either Weber's Law or Stevens' Power Law. Eisler's (1976) comprehensive review of studies spanning a hundred years settled on a Stevens power exponent for time perception of $c \simeq .9$. Classical economics, for instance, expects a decreasing marginal utility of money and therefore that the exponent should be < 1. Surveying his own and others' work Stewart (2009) noted a stable exponent for money that was just under .5, i.e. approximately a square root law.[2]

1.3.2   Powers and logs (Stevens $\supseteq$ Weber-Fechner?)

Undoubtedly Stevens' power law is more flexible than the Weber-Fechner log law, though at the expense of having

---

[2]Power law and log utility functions are but special cases of Hyperbolic Absolute Risk Aversion (HARA). Nonetheless, it is they rather than the general HARA that have spawned most behavioral models in intertemporal choice.

to estimate an additional parameter. If c = 1 we are treating time or money as if objective. As c decreases towards zero, E^c becomes more curved, in fact more logarithmic-like. This follows from a mathematical result that is important for this paper, which is that if x > 0:

$$(x^c - 1)/c \rightarrow log(x), \ as \ c \rightarrow 0 \qquad (4)$$

In many applications that compute relative differences (including here) constant scaling factors do not matter just as long as they are applied to all observations. Consequently, the c in the denominator can usually be ignored. Therefore for small c > 0, x^c becomes log-like around 1, rather than around 0, as is normal. Therefore, by subtracting 1 from x^c, we can morph from linear through fractional power to log, under the control of the parameter c. This is very useful.

The equation y = $(x^c - 1)$ / c is none other than the Box-Cox transformation, much used in statistics to find suitable functional forms for the relationship between x and y (Box & Cox, 1964). As Wakker (2008) has noted in his useful tutorial on power functions, researchers often neglect to explore the space of c < 0, which gives rise to reciprocal powers. For instance, c = –1 gives the functional form: y = 1 – 1/x. In his Section 2.1 he gives a cautionary example of failing to analyse c < 0.

Sometimes it doesn't matter whether we subtract 1 or not. For instance, often we are interested in difference relationships such as v = $(x^c - y^c)$. In this particular case, the Stevens' Power Law nests the Weber-Fechner Log Law as a special case. This follows because we can re-write the equations as:

$$v/c = (x^c - 1)/c \ - \ (y^c - 1)/c \qquad (5)$$

then using the mathematical result,

$$v/c \rightarrow log(x) - log(y) = log(x/y), \ as \ c \rightarrow 0, \quad (6)$$

Therefore in order to build in modeling flexibility, rather than forcing a relationship to be log(x/y) and thus fixing on a Weber-Fechner law, we can let it be the more general $x^c - y^c$ and let an optimizing program determine which c best fits the empirical data. If c is close to zero, then a Weber-Fechner-like log law applies, if c is close to 1, then x and y are linear, if c = .5, then we have a square-root law, and so on.

Sometimes it does matter. For instance, when a single variable is being considered, $x^c \rightarrow 1$ as $c \rightarrow 0$, which implies that the x variable is redundant. In order to avoid this problem, and to be consistent with the log behavior of $(x^c - y^c)$, the form $(x^c - 1)/c$ is often used to express powers, instead of x^c.

1.3.3 Started values

Unfortunately, it may happen that x takes the value 0, either always or sometimes. Typically this happens when an amount of money is offered *now* (at t = 0) compared with an amount offered later. This is then a problem, because log(0) is undefined. The problem also occurs for any power law whose exponent is negative (e.g., t^{-1} = 1/t is also undefined for t=0). So, any model that wants to treat time as logarithmic, either explicitly or implicitly as the limit of a power law, throws up an anomaly for t = 0, which is actually the most frequently used value of t. One way out of this difficulty is to present t = 0 as a special (sometimes absent) case. Another is to use "started time" instead of time itself. Started counts, started reciprocals, started logs, etc. were recommended by Tukey (1977) and Mosteller and Tukey (1977) to overcome just this kind of problem when dealing with troublesome zeros. Started time translates all measurements t into t + ε, where ε is small relative to the values that t will take. Equivalently, in our calculations we can translate time from days, as may have been stated in the questionnaire, into hours (or even minutes), and just let ε = 1. Under this translation, t = 0 would become $t_* = 1$; t = 2 days would become $t_* = 49$; 7 days would become $t_* = 169$ (all started hours), and so on, with the * subscript emphasizing that we are dealing with *star*ted time. Although this tweak is motivated by avoiding a mathematical problem, started time may pragmatically reflect reality better than the nominal time appearing on the questionnaire; because when a reward is offered "now" participants would not assume they would get the reward instantaneously, rather that they would get it after the experiment was over—i.e., quite possibly about an hour later. This delay between mathematical and actual now is even more apparent if the amount is offered "today".

Ainslie and Herrnstein (1981, p.480) make a similar point, distinguishing between the nominal and effective timing of reinforcements: "It could be argued that the usual convention of measuring delay from a subject's [pigeon's] peck to the elevation of the feeder should be abandoned in favor of the interval from the peck until actual eating takes place, or even until the food enters the subject's bloodstream. The 'true' moment of reinforcement is not known and might not even exist as a distinct instant..." However, researchers have not yet taken up the challenge contained in the last sentence, which is to model the timing of events fuzzily rather than as discrete points.

Generally there should not be a parallel issue for modeling money because in any realistic choice F and P will both be non-zero. If it ever did become an issue, money could similarly be "started" by adding one cent to both P and F.

1.3.4 Treatment of time and money.

In almost all models surveyed here money and time are combined multiplicatively. A ratio is formed between the subjective value of additional money and the subjective dis-value of additional delay: that ratio is the rate parameter. We can distinguish two very general views on time and money. In one, money and time are seen as qualitatively distinct, which means that treating them differently is not an issue. For instance, money can be compounded or subject to inflation, which thus change its face value and/or its real worth: time cannot. Money is traded: time is generally not, and so on. In the other view subjective perceptions of money and subjective time are modeled as if they were standard psychophysical stimuli (pain, brightness, loudness, etc.). Now if psychophysics itself treats its stimuli via consistent functional forms, for instance by Stevens' power law (or the Weber-Fechner law), it sets a strong precedent for treating time and money via similar functional forms. Accordingly, if a model treats subjective time and money inconsistently, the burden of proof should be on the model's creator to justify why it does so.

Using the convention that F is offered at time T, and f is offered at time t, with T > t (see Section 1.4: f is known as P as the special case when t = 0), also that a, b, A and B are exponents of the model, used in different ways, the following pairs do *not* have similar functional forms:

$(F - f)^a$ $T^b - t^b$ (power on a money difference: difference of times to a power)

F/f   T – t   (money is ratio: times are differenced)

whereas the following *do*:

F/t   T/t   (money and time are both ratios)
$(F^a - f^a)^A$ $(T^b - t^b)^B$ (both have same functional form)

This is not to suppose that the parameters need match up (e.g. that A = B, and a = b in these examples). We propose that a useful characteristic to observe in a model is whether it treats time and money symmetrically or not. If it does not, has any special justification been given why not?

1.4 A note on notation.

In defining parameters for the different models, I have tried to obey the following guidelines. I obey common usage where it is standard in general algebra (e.g., r is the common ratio of a Geometric Progression, d the common difference of an Arithmetic Progression). I have tried to use mnemonic devices (e.g., h for hyperbolic, m for the exponent on money), and use capital letters for the models themselves (e.g. E for the exponential model, H for the simple hyperbolic). Generally speaking the rate parameter is lowercase of the uppercase model: e.g., H, h.

I have also respected past literature where terminology is widely and standardly used so that it would be confusing to do otherwise (β-δ of the quasi-hyperbolic model). I have tried to use Roman letters (but not *e, l, o,* or *p*), rather than Greek. Because of the sheer number of parameters used in this survey, inevitably I have failed. I have used the convention that F is a future gain, and P is the present gain, which follows standard practice in accounting and finance, though not elsewhere. Where I wish to contrast two gains in the future, I have used the symbol f to refer to the less distant one. Thus, because of the premise that people discount delays, generally F > f. To carry the capitalization mnemonic into time, I have used T to refer to more distant time than t (T > t). So, a typical choice might be to ask which alternative is preferred: (f at time t) or (F at time T). Although that may seem quite natural, one awkward consequence of this usage is that, since most models are based around the special case of t=0, the standard symbol for time becomes T, rather than the more natural t. Also, because both t and T are now in use, to maintain mnemonic value I have resorted to using the Greek τ (tau) to subscript parameters and models that focus on a time issue, as well as for exponents used to deform objective time into subjective time. Consequently, we have the presbyopic nightmare of: T^τ and t^τ. Old eyes be warned!

Table 1 is provided to help with the many symbols. Note also the legend.

2 Simple rate parameter models

In accounting and finance, discounting can be described either as a method of valuing a future cash flow (F) as a present value (P); or in terms of the Internal Rate of Return (IRR) that would generate F from P. Present values may be used in preference to IRR to consolidate several future cash flows at different times into a single net present value (NPV). However, for simpler problems IRR is often used as a threshold or "hurdle rate" to make investment appraisals (Ryan, 2007), and it is the IRR view of the problem which turns out to be more useful for this survey. The very reasons that organizations use an IRR hurdle rate (because it is a single, unambiguous decision criterion that is easy to understand and apply organization-wide by its decision makers), are the same arguments to suggest why lay-people may spontaneously adopt a similar view of intertemporal choice (ITC). Typical ITC problems in the psychological literature do not require people to consolidate *several* future cash flows, for which NPV comes into its own, so both IRR and NPV are candidates. Note, however, that NPV is a relatively recent invention in accounting, and not one that people in business spontaneously and repeatedly re-invented as a

Table 1: Model names, notation used throughout this article, and rate parameter equations.

Model name	Model symbol	Recovered parameters	Rate parameter eqnuation
Exponential	E		$r = [\log(F/P)] / T$
Hyperbolic	H		$h = (F/P - 1) / T$
Arithmetic	A		$d = (F - P) / T$
Hyperboloid (G&M)	H_m	m	$h_m = ((F/P)^m - 1) / T$
Hyperboloid (Rachlin)	H_τ	t	$h_\tau = ((F/P) - 1) / T^\tau$
quasi-Hyperbolic (T>0)	qH	$\beta \ (= 1/\gamma)$	$q = [\ln(\beta F / P)] / T$
Fixed cost (T>0)	X	χ	$x = \log((F - \chi)/P) / T$
Read	B	θ	$b = \log(F/f) / (T - t)^\theta$
Roelofsma	V	υ	$v = \log(F/P) / [\log(\upsilon)+\log(T)]$
Constant sensitivity	C	b	$a = [\log(F/P)]^{1/b} / T$
CRDI (ψ>0)	M_+	$\beta(=1), \psi$	$\rho = \log(\beta F / P) / T^\psi$
CRDI (ψ=0)	$M_=$		$\rho = \log(\beta F / P) / \log(T)$
CRDI (ψ<0)	M_-		$\rho = \log(\beta F / P) / -T^\psi$
CADI (η>0)	N_+	$\beta(=1), \eta$	$\sigma = \log(\beta F / P) / e^{\eta T}$
CADI (η=0)	$N_=$		$\sigma = \log(\beta F / P) / T$
CADI (η<0)	N_-		$\sigma = \log(\beta F / P) / (-e^{\eta T})$
Hyperboloid on intervals	$H_{m\omega}$	m, ω	$h_{m\omega} = ((F/P)^m - 1) / [(T-t)-((F/P)^m - 1) \, \omega t]$
Size-sensitive Arithmetic	A_z	z	$d_z = P^{-z} (F - P) / T$
Killeen	K	m, τ	$k = (F^m - P^m) / T^\tau$
Killeen intervals & t_*	K_τ	m, τ	$k_\tau = (F^m - f^m) / (T_*{}^\tau - t_*{}^\tau)$
Generalized H	$H_{m\tau}$	m, τ	$h_{m\tau} = ((F/P)^m - 1) / T^\tau$
Generalized E, H, qH, X	G	β, χ	$g = \log(\beta F/ (P + \chi)) / T$
Size-sensitive K	J	m, τ, z	$j = P^{-z}(F^m - P^m) / (T_*{}^\tau - t_*{}^\tau)$
Discounting by intervals	I	m, τ, θ	$i = ((F / f)^m - 1) / (T^\tau - t^\tau)^\theta$
Tradeoff	Y	(see text) $\gamma, \tau, \varepsilon_m, \varepsilon_\tau, a_1, a_2, b_1, b_2$	
Two exponentials	$E_{\beta\delta}$	w	$P/F = w.e^{-r\beta T} + (1-w).e^{-r\delta T}$

Notes: m is the money exponent.

τ (tau) is the time exponent.

β is the quasi-hyperbolic non-present scaling parameter, and $\gamma = 1/ \beta$ is the present-premium.

z is the size exponent in J.

θ is the interval exponent in I (DBI).

F (future value) is the more distant future amount delivered at time T.

f is the less distant amount delivered at time t. If t= 0, f = P, the present value.

T_* and t_* are "started" times (Tukey).

W and w are weights placed on different processes.

kind of common sense, despite centuries of dealing with relevant problems (Rubinstein, 2003). The non-obvious nature of NPV therefore suggests that lay people might be more likely to adopt an IRR view of ITC problems than an NPV one, a hypothesis that deserves empirical investigation. We generalize the concept of IRR from r, the interest rate in E, to cover model *rate parameters* which capture the intensity of discount operating in each of the other models we consider.

2.1 Exponential discounting (E, r)

The standard financial method of summarizing growth rates by which to translate present values into future values and vice versa is by a geometric mean taken on the ratio of increase. Thus, the compounding growth model for discrete time is:

$$F = P(1 + r)^T \qquad (7)$$

with F (future value), P (present value), r (growth rate) and T (compounding periods) taking their usual meanings, so that

$$r = (F/P)^{(1/T)} - 1 \qquad (8)$$

As an example, if \$200 (P) becomes \$400 (F) over a period of 4 years (T), then r = $(400/200)^{(1/4)} - 1 = .19$, or a 19% increase per annum. We can also compute P as:

$$P = F[1/(1 + r)]^T, \qquad (9)$$

where $[1 / (1 + r)]^T$ is known as the discounting factor: it allows us to translate a future cash flow into its present value. When faced with a choice between:

   (1) P, an instantaneous payoff now, and
   (2) F at time T,

the decision maker, operating normatively as an exponential discounter, would apply the discounting factor to F using a given r, and choose the larger of P and the discounted F. Another way of describing this process is to focus on equation (8) and compute r for a given choice of P, F, and T. This is equivalent to the internal rate of return (IRR) as used by accountants, for instance to judge whether a planned project will meet a criterion level of return. If r exceeds a given criterion rate r_0, then the accountant would accept F. We posit that r_0 is not a commercially available rate of return, but a person-specific one. Each person uses their own r_0 as a criterion rate of return by which to judge whether r, computed from choice {P, F, T}, exceeds criterion (hence choose F) or not (hence choose P). We do not, however, claim that r_0 is impervious to context, manipulation, and simple fluctuation with time.

If we increase the number of compounding periods to be n per annum, equation (9) becomes:

$$F = P(1 + r/n)^{nT} \qquad (10)$$

and if n increases without limit so that compounding is continuous, this becomes

$$F = Pe^{rT}. \qquad (11)$$

The discount factor therefore becomes e^{-rT}, hence the name "exponential discounting", though the geometric form in (10) is often still used. The growth rate parameter r is determined from (11) as:

$$r = [log(F/P)] / T. \qquad (12)$$

In practice, the exponential and geometric view of growth and discounting do not differ much in their computed values of r. In the above example, equation (12) gives r = 17%, rather than the previously calculated 19% from equation (8).

2.2 Hyperbolic discounting (H, h)

Accountants also frequently calculate the (arithmetic) average rate of return (Ryan, 2007, p. 12). Similarly, if a quantity increases by x% over T years, then the arithmetic average of the rate of return should be (x/T)%. In the example cited above there was a 100% increase, which averages out to 25% per annum. Although frequently used to appraise investments and set a benchmark figure for project acceptance, this figure cannot be used in the compound growth formula. Instead of doubling the original as required by a 100% increase, its use would lead to a calculated increase of $(1.25^4 = 2.44)$ times the original. These arithmetic calculations are equivalent, in algebra, to computing a growth rate of h by:

$$h = (F/P - 1) / T. \qquad (13)$$

In the example given in 2.1, h = (2 − 1) / 4 = .25. Equation (13) also implies:

$$P = F[1/(1 + hT)]. \qquad (14)$$

According to equation (14), the discount factor [1 / (1 + hT)] has a hyperbolic form, hence the name "hyperbolic discounting." Also, whereas E is underpinned by a model of compound interest, H is underpinned by a model of simple interest, with no compounding (Rachlin, 2006). This is also apparent if we let n → 1/T in equation (10) for a single compounding period, which reduces to the hyperbolic model. Thus discrete-time compounding is a half-way house between the extremes of E and H under the control of the frequency of compounding, n. When n is extremely large, the compounding periods become very small, until in the limit we have continuous compounding (therefore E). When n=1, the interest is accrued at the end of the investment period, so in effect we have simple interest (therefore H).

The fact that the 2.44 exceeds 2 (and 25% exceeds 19%) in the above example is no accident, and is due to the mathematical relationship that the arithmetic mean $\geq$ geometric mean for positive numbers. It follows that, in the period up to T, hyperbolic discounting (which uses the arithmetic mean) will always make a higher estimate of future growth than exponential discounting: therefore, the immediate future will be discounted more heavily in the hyperbolic model. After T, the situation reverses.

We can also use equation (13) to determine the hyperbolic discounting rate parameter h, implicit in any choice {P, F, T}, which we then use to compare against someone's internal h_0. As with exponential discounting, if people apply a hyperbolic model, then if $h > h_0$ they will choose F; if $h < h_0$ they will choose P.

Although the hyperbolic discounting model in equation (14), given in Mazur (1987), is the one most widely used in current research, it is founded on models that predate this formulation. As reviewed in Ainslie (1975), Chung and Herrnstein's (1967) found that the rate at which pigeons would peck was proportional to the reciprocal of the time delay, as an instance of the "matching law". This led Ainslie to propose the first hyperbolic model, presented in Mazur (1987) as: P = F / kT, where k is a constant of proportionality. The problem with this formulation is that as $T \rightarrow 0$, the present value of any reward becomes infinite, but can be remedied by adding a constant to the denominator (Herrnstein, 1981), and it is then a short step to fix that constant as the value 1.

Finally, note that the word "hyperbolic" is often used loosely to cover any discounting in which the discount rate is higher at shorter delays than it is at longer delays. We use the term more precisely to refer to the particular model described in (13) and (14).

2.3 Arithmetic discounting (A, d)

In arithmetic discounting, future and present values are compared by constant increments d, exactly as in an arithmetic progression:

$$d = (F - P)/T \qquad (15)$$

$$P = F - Td \qquad (16)$$

Instead of multiplying F by a discount factor to yield P, we subtract a discount decrement (Td). In one choice people are offered P: in the second they are offered P + (F − P). If the participant decomposes the choice in this way, then the decision is whether it is worth waiting for the excess (F − P). If they feel their time is worth d_0 per day to wait, then if $(F - P) > Td_0$ they chooses F. Whereas r and h are dimensionless, being essentially expressions of interest rates, d is measured in units of money per time (e.g. dollars per day).

There is another justification for equations (15) and (16). Killeen (2009) started from the assumption that the marginal change in utility with respect to time follows a power law; also, that utility and value are themselves related by a power law. Taken together these were shown to yield an equation of the form: $P = (F^m - T^\tau d)^{1/m}$ (Killeen, 2009, equation 6, p.605). We examine the general form of this equation in a later section (4.1). But for now, notice that if c = 1, subjective time is objective time; and that if k = 1, the utility of money is just its face value, giving equation (16). Doyle and Chen (2012) found empirical support for A, relative to E and H, both in other researchers' past data, and in their own. They surmised that people were treating delay discounting by analogy with "wages for waiting" rather than by analogy with "investment growth" as implied by E and H.

3 Rate parameter + 1 models

Each of the models in this section has one additional (free) parameter that must be estimated ("recovered") from the data. The next sub-section 3.1.1, for instance, models utility of money with a power law. An additional parameter offers the possibility of more accurately modeling behavioral discounting, and of also "recovering" additional information, such as how money is subjectively perceived, as captured in the exponent used in Section 3.1.1. But additional parameters also need additional data to avoid the problems of over-fitting; and this data must be distributed to provide sufficient variety in order to estimate the additional parameter.

3.1 Hyperboloid models

3.1.1 Green and Myerson (H_m, h_m; m)

Myerson and Green (1995) proposed a model which generalizes the hyperbolic. Whereas in this and other publications these authors consistently use the symbol s for the exponent, we use 1/m for consistency with the rest of this article. Their model is:

$$P = F/(1 + h_m T)^{1/m} \qquad (17)$$

This model is a special case of Loewenstein and Prelec's (1992) generalized hyperbola formulation. Further simplifying, if m = 1, we have the hyperbolic. When 1/m (=s) < 1 the tendency of H to discount more steeply than E at short durations (but less so over longer durations) is exaggerated. Equation (17) may be re-arranged to give a rate parameter that is the equivalent of r, h, or d:

$$h_m = ((F/P)^m - 1)/T \qquad (18)$$

The subscript m on the rate parameter emphasizes that money may be treated subjectively, and requires the researcher to estimate m (=1/s) empirically. Typical values that are reported in the literature for s (=1/m) are in the range [.45, .78] (McKerchar, Green, Myerson, Pickford, Hill, & Stout, 2009; McKerchar, Green & Myerson, 2010), and these authors find that modeling the additional parameter leads to statistically significant improvements in fit over the hyperbolic. Notably, s is typically less than 1, which means that the ratio (F/R) in equation (18) is being raised to a power greater than 1, thus acting to exaggerate the relative difference between F and P in calculating h_m. At the other extreme, when s becomes large, the exponent acts to shrink the ratio (F/R) towards 1, and in so doing (18) begins to approximate the exponential model.

To see why, we use the mathematical relationship that $(x^c - 1)/c \to log(x)$ as c → 0. Here x is (F/P) and c is m. The smaller m becomes, the closer $(F/P)^m$ -1 comes to m.log(F/P). Once the value of m is fixed at some small value, all calculations in (18) are multiplied by the *same* m, which in this way just acts as a scaling constant, and can be ignored. It follows that the exponential model is a special case of (18) when s gets to be very large. The take-home point is that since m > 1 in empirical data, actual behavioral discounting is not a compromise between E and H, but lies even further from E than H had suggested.

3.1.2   Rachlin (H_τ, h_τ; τ)

An alternative way to generalize the hyperbolic was proposed by Rachlin (2006), and also presented in Mazur (1987):

$$P = F/(1 + h_\tau T^\tau) \qquad (19)$$

Here, the exponent τ applies specifically to T rather than the whole brackets. This minor modification leads to a subtly different model. Rearranging (19) to isolate the rate parameter, we get:

$$h_\tau = (\,(F/P) - 1)\,/\,T^\tau \qquad (20)$$

In this version of the hyperboloid, the numerator (the treatment of money) is the same as for the hyperbolic. However, unlike in the previous four models, time may be treated subjectively by means of a Stevens-like power law. Similar to before, if τ=1 we have the hyperbolic. But once again, $\tau < 1$ in empirical data. Thus, subjective time is increasingly compressed at longer periods. The typical range for τ is [.67, .90] (McKerchar et al. 2009, McKerchar et al. 2010) which is just slightly lower than other estimates of the subjective time exponent (Eisler, 1976). This is the first model we have considered in which the rate parameter cannot be expressed in straightforward units of measurement.

3.1.3   m, τ, or m relative to τ?

Comparing models the money and time parameters from H_m and H_τ, in 3.1.1 and 3.1.2 we cited typical (m, τ) exponents as being, very approximately (1.5, 1) when H_m was the model, and (1, .7) when H_τ was the model. The utility of money is generally conceded to be m < 1. This may lead us to believe one of two things. Either the estimates of m and τ are discrepant with each other, and in the case of m, highly discrepant with past research. Alternatively, that what may count behaviorally, over and above the absolute values of m and τ, is their sizes relative to each other. Suppose it is a psychological fact that m > τ, i.e. that subjective money is less concave (more convex) than subjective time, then in fitting m and τ relative to each other, both models are consistent. But if one model constrains $\tau = 1$ (as H_m does), then the behavioral requirement that m > τ will force m to be greater than 1. Likewise if the other model constrains m = 1 (as H_τ does), the behavioral requirement m > τ will force τ to be less than 1. This is what we find.

The requirement m > τ is consistent with Zauberman, Kim, Malkoc and Bettman's (2009) conclusion that perception of time are more labile than perception of money, but it is not consistent with exponential discounting for which m ≪ τ.

3.2   Present bias / present premium models

3.2.1   Quasi-hyperbolic (β-δ) discounting (qH, q; β)

Quasi-hyperbolic discounting (qH) is rarely used in psychological research, though it is used extensively by economists hoping to preserve as much of the exponential model as possible. In the discrete time version of the exponential model each factor discounts the previous one by δ (= 1/(1+r)) to form a geometric progression of discount factors: $\{1, \delta, \delta^2, \delta^3, \ldots \delta^n\}$ for t = 0, 1, 2, 3…n, respectively, with $\delta < 1$. As explained in Section 2.1, when the compounding periods become very small, and n becomes correspondingly large, we have continuous compounding, and therefore E itself. Contrasting with discrete-time E, the quasi-hyperbolic model posits discount factors: $\{1, \beta\delta, \beta\delta^2, \beta\delta^3, \ldots \beta\delta^n\}$ for t = 0, 1, 2, 3…n, respectively, with $0 < \beta$, $\delta < 1$ (Laibson, 1997). If $\beta = 1$ we would have a straightforward compounding model, as described in equation (9). But if instead, $\beta < 1$, this parameter acts to give a one-off boost to discounting over the first compounding period. The purpose of β is to help tick the box of unexpectedly steep discounting at short durations, which is the hallmark of actual behavioral data, and which first motivated the use of hyperbolic discounting. Thereafter, the model assumes that discounting follows the standard normative model. Paralleling the development in Section 2.1, the quasi-hyperbolic can also be

used for continuous discounting (e.g., Benhabib, Bisin, & Schotter, 2010). The discount factor D is:

$$D = 1 \ \ if \ T = 0; \quad D = \beta e^{-qT} \ \ if \ T > 0 \qquad (21)$$

The rate parameter for quasi-hyperbolic discounting is therefore:

$$q = [log(\beta F/P)] \ / \ T \ \ for \ T > 0 \qquad (22)$$

All F first get scaled by the additional parameter β before being treated by the exponential model. Obviously, if β = 1, then q = r as in equation (12). It is thus clear that qH discounting is hyperbolic only in its intent to mimic the steep initial discounting of the hyperbolic model, but in every other sense it is an exponential model.

The qH model may have been motivated pragmatically to preserve exponential discounting and all the useful mathematics that goes with it, while also modeling initial steep discounting. Nonetheless, the idea that F (but not P) gets a special "tax" simply because it is not received right now does have a more constructive justification. There is a *qualitative* difference between now and any time to come, which we all appreciate intuitively, and which language mirrors in verb tenses. From this perspective, the discontinuity in qH between t = 0 and t > 0 is not an ugly kludge to workaround bothersome behavioral evidence, but a clever modeling device that allows what happens now (P) to be treated differently from all future events (F). In this way qH, a model devised by economists for economists, has the capacity to capture a psychological distinction that the other models cannot.

To help rescue this insight from the limitation of being identified exclusively with the exponential model, let γ = 1 / β. Whereas β captures the idea that F is less than it should be, γ captures the idea that P is more than it should be, and we call γ the *present premium*. Analogous ideas are met in risky decision, where a certain reward (p = 1) are valued at a premium over rewards with p < 1; and also in the premium due to mere possession or the endowment effect (Thaler, 1980; Sen & Johnson, 1997). Rewards that are *certain*, *mine*, and *now* are all over-valued. Potentially, the models we have considered or are about to consider, could incorporate a present-premium parameter by simply replacing P with γP (or F with βF) in all calculations. Benhabib, Bisin and Schotter (2004) formulate exactly such a hybrid between H_m and qH in this way (see also Section 4.4, and Tanaka, Camerer & Nguyen, 2010).

Nonetheless, there are still issues to be confronted with qH. In particular, if β < P/F in a particular choice, the model predicts P should always be preferred to F, no matter how short the delay in F—unless someone's internal rate parameter is negative. Laibson (2003) suggested "calibrating" qH with $\beta \simeq .5$, meaning that 21

of Kirby, Petry, and Bickel's (1999) 27 choice questions would be in the negative rate parameter category. Even β = .90 would put 8 questions into that category. Summarizing several empirical studies in economics, van de Ven and Weale (2010) noted βs of .296, .308, .674, .687, .846, .942; and .825 in their own work; and in Albrecht, Volz, Sutter, Laibson and von Cramon (2011) the median β from individual experimental data was .86, with 24 of 27 participants having β < 1. If these estimates are close to what individuals use in binary choice, then β < P/F would occur more than just occasionally in typical data.

The quasi-hyperbolic is popular with behavioral economists. It lends itself to convenient testing against the normative model E, by testing whether β is significantly different from (less than) 1. Unfortunately, in the literature qH has rarely been tested against the kinds of models surveyed here which are more potent challengers in that they themselves have consistently proven better than E. Furthermore, Kable and Glimcher (2010) argued that the *present bias* is more strictly a *soon-as-possible bias*, implying that a deflationary β should be applied to all f (t > 0), and not just P (t = 0). Finally, as has already been pointed out in Section 1.3.3, nothing ever actually happens at t = 0, in that a reward offered "now" is actually received at t > 0. Therefore, the special status of t = 0 in qH can be questioned on the grounds that it may never occur in practice.

3.2.2   Benhabib et al.'s fixed cost model (X, x; χ)

As an alternative to the *percentage* decrement for non-present rewards, Benhabib, Bisin and Schottter (2010) suggested using a fixed cost to model present bias—that is, an *absolute* decrement: use (F – χ) in place of F. Thereafter use the exponential model, exactly as in qH.

$$x = log((F - \chi)/P) \ / \ T, \quad for \ T > 0 \qquad (23)$$

They found a present bias / present premium of about $4 among amounts of $10 through $100. The fixed cost model better fitted their data than the qH model. A different, though very similar model is obtained if χ is added to P rather than subtracted from F (see Section 4.4). Hardisty, Appelt, and Weber (2012) found that χ was positive for both gains *and* losses, suggesting a general "want-it-now" bias.

3.3   Read's interval model (B, b; θ)

Read (2001) presented his model in terms of a discount fraction D that occurs within the time window [t, T]. D is the fraction by which F is reduced by being at the end of the interval (T), rather than the start (t): namely, f / F. The exponent θ applies to the time interval, rather than time itself (as would be the case with the exponent τ, which

is not used here). The algebraic steps to isolate the rate parameter b are:

D = [1 / exp((T − t)$^\theta$)]^b
D = 1 / [exp((T − t)$^\theta$)]^b = f / F
[exp((T − t)$^\theta$)]^b = F/f
b(T − t)$^\theta$ = log(F/f)

$$b = log(F/f)/(T-t)^\theta \qquad (24)$$

From equation (24) we see that Read's is an exponential model performed on the subjectively (because $\theta \neq 1$) perceived time *interval* (T − t). It nests the standard exponential model E when the interval of inspection starts from t=0 (at which point f is to be known as P), and when subjective time is assumed to be linear with objective time (θ = 1).

While this model seeks to understand cognitive processes through the use of interval data (t > 0), it still must compete with other models at the particular case of t = 0, even if a given competitor model has not (yet) been generalized to the case of t > 0.

3.4   Roelofsma—exponential time (V, v; υ)

Arguing from the Weber-Fechner Law of logarithmic sensitivity to psychophysical stimuli, Roelofsma (1996) presented his model as:

$$P = U(F)/(1+v)^{vlog(T)} \qquad (25)$$

If we take U(F), the utility of F, as itself being logarithmic as in the exponential model E, and move from discrete to continuous time, we derive the rate parameter:

$$v = log(F/P) \,/\, [log(\upsilon) + log(T)] \qquad (26)$$

Whether log(υ) is an intrinsic part of this model is not really discussed by the author. If it is not, the obvious choice is to set υ = 1, so that this model becomes a simple rate parameter model in log money and log time, with no additional parameters to be estimated. In treating both money and time logarithmically, among the models I have surveyed, Roelofsma's is the most enthusiastic adherent of the Weber-Fechner law. Assuming υ = 1, and writing (26) as: v = [log(F) − log(P)] / log(T), we see that Roelofsma is structurally similar to the arithmetic model except that, instead of using raw values of F, P, and T in the rate parameter equation, each is replaced with its logged equivalent.

3.5   Synopsis of the models so far

E is the normatively correct model to value investments. It assumes a continuously compounding interest model of growth. But E fails to capture important aspects of how

laypeople actually evaluate intertemporal choices. qH extends E by making a distinction between present choices and future choices. In qH, all choices in the future are devalued by a constant scalar β before treating them as in E, which itself is the special case of qH when $\beta = 1$. Model X, like qH, maintains that there is a special difference between P and all future values F, but whereas qH treats the present premium multiplicatively, X does so additively.

For over two decades, H has been the psychologists' default alternative to E in that it consistently estimates people's actual discounting behaviors better than E. H is equivalent to a simple interest model of growth. The hyperboloid models are extensions of H that treat money (H$_m$) and time (H$_\tau$) as non-linear. H$_m$ is equivalent to H when its additional parameter m = 1, and E when 1/m → 0. But past evidence estimates m to be less than 1, which might imply *increasing* utility to money, contrary to classical theory and common intuition. Also, H$_\tau$ is equivalent to H when its additional parameter $\tau = 1$. The mental operations involved in A are simpler than those in all other models, which all start by forming a ratio between (possibly transformed) F and P. Model A starts by forming a difference between F and P.

The special case of t = 0 in Read's model B (all time intervals start from now; hence f = P) presents as a hybrid between E and H$_\tau$. It is an exponential model (logarithmic money) calculated on subjective (power law) time. Of course, in modeling time intervals in general, not just ones that start from the present, Read's model aims to examine phenomena that the others cannot. Roelofsma's model V treats both money and time logarithmically, which makes it equivalent to A, where all monies and times have first been logged.

Finally, because of the reciprocal T in calculations of r, h, d, h$_m$, q, and x for a given P and F, all of these rate parameters decline hyperbolically with time. In fact, one might distinguish models E, H, A, H$_m$, qH, and X from each other only by their treatment of money, and once that is determined in the numerators of their respective rate parameter equations, r, h, d, h$_m$, q, and x all decline *identically* with objective time. This neglected perspective suggests the essential difference between the five models of intertemporal choice, A, A$_z$, H, E, H$_\tau$, qH, and X has nothing to do with time *per se*, and could even be tested against each other by considering a single fixed time gap between P and F.

3.6   Time (im)patience models

From the point of view of modeling rate parameters, most of the above models have focused on their treatment of money, making few inroads into modeling the subjectivity of time. In this Section we collect together research that takes a serious look at how time is perceived—while

incidentally accepting the log(F/P) of model E as a given.

3.6.1 Ebert & Prelec's constant sensitivity (CS) function (C, a; b)

"Given that unit elasticity defines compounding discounting [E], it is natural to interpret lower-than-unit elasticities as indicating insufficient time-sensitivity relative to compound discounting. If elasticity is constant and equal to b > 0, the discount function has the constant-sensitivity form" (Ebert & Prelec, 2007, p.1425), which is:

$$P/F = exp(-(aT)^b) = e^{-(aT)^b} \qquad (27)$$

As b → 0 we have diminishing time sensitivity. If b ≫ 1, then the CS function begins to appear as a step function, meaning that time is ever more clearly dichotomized into "near future" and "far future" (at the boundary $1/\alpha$). See their Figure 1. We can isolate the rate parameter associated with the CS model from equation (27) as:

$$a = [log(F/P)]^{1/b} / T \qquad (28)$$

Clearly, when b = 1 (unit elasticity), we have model E as the special case. Ebert and Prelec note, but avoid, the similar discount function:

$$P/F = exp(-(a_x T^b)), \qquad (29)$$

leading to the rate parameter:

$$a_x = log(F/P) / T^b \qquad (30)$$

which is a special case of the generalized hyperboloid to be examined in Section 4.3, with m → 0.

3.6.2 Bleichrodt et al. Constant Relative Decreasing Impatience (CRDI) functions (M, ρ; ψ, ($\beta = 1$))

CRDI is explicitly described as an analog of constant absolute risk averse functions (CARA) that appear in risky choice models. Bleichrodt, Rhode, and Wakker (2009) observed that hyperbolic discount functions were developed to account for decreasing impatience. That is, relative to exponential discounting, people consistently show greater impatience (for P) in the short term, but greater patience in the longer term. They also noted that such models failed to accommodate "increasing impatience or strongly decreasing impatience" (p. 27), particularly given the attempt to model individuals' discounting behavior: hence their search for discounting functions to fulfill this need. Let D(T) be the discount function (i.e., D(T) = P/F). Then their CRDI function is / are:

(i) D(T) = β.exp(-ρT$^\psi$) for $\psi > 0$
(ii) D(T) = β.T$^{-\rho}$ for $\psi = 0$, and T $\neq$ 0

(iii) D(T) = β.exp(ρT$^\psi$) for $\psi < 0$, and T $\neq$ 0

Examining each in turn, we have for $\psi > 0$,
 log(P/ βF) = -ρT$^\psi$, thus:

$$\rho = log(\beta F/P) / T^\psi \qquad (31)$$

The following points should be noted. First, money is treated as in qH, with β fulfilling the same role in both models. Second, there is a power law on time; if $\beta = 1$, the model is exactly the route *not* taken by Ebert and Prelec (2007), as in equation (30). Third, if $\psi = 1$, we have objective clock time, and thus "constant impatience"; if $0 < \psi < 1$, we have the conventional view that time becomes increasingly contracted the more distant in the future it is, and thus we have "decreasing impatience"; but if $\psi > 1$, the more distant the time, the more *stretched* it becomes, which therefore models "increasing impatience".

Examining $\psi = 0$, we have:

$$\rho = log(\beta F/P)/log(T) \qquad (32)$$

If $\beta = 1$, we have Roelofsma's model (presuming $\upsilon = 1$), described in Section 3.4.

Finally, if $\psi < 0$, we have:

$$\rho = log(\beta F/P) / (-T^\psi) \qquad (33)$$

In this case the denominator asymptotes to zero, *but from below*. It is designed to model impatience that decreases more rapidly than for the first part with $0 < \psi < 1$. In order for ρ to be positive (Bleichrot et al, 2009, Definition 4.1, p. 31), we must have that $\beta <$ P/F so that the log in the numerator is negative allowing the minus signs to cancel.

As ψ moves through zero from above, for a given T > 0, the denominators in parts (31), (32) and (33) become 1, logT and –1, respectively. This means that any algorithm that hopes to recover the impatience parameter ψ from data has a discontinuity to negotiate at $\psi = 0$. It may therefore be better to think of (i), (ii) and (iii), or equations (31), (32) and (33), as three distinct sub-models, rather than a single model.

3.6.3 Bleichrodt et al. Constant Absolute Decreasing Impatience (CADI) functions (N, σ; η, ($\beta = 1$))

Their analog of a CARA Constant Relative Risk Averse function is the CADI function given by:
 D(T) = β.exp(-$\sigma e^{\eta T}$) for $\eta > 0$
 D(T) = β.exp(-σT) for $\eta = 0$
 D(T) = β.exp($\sigma e^{\eta T}$) for $\eta < 0$
"[β] is a scaling factor without empirical meaning in the sense that it does not affect preferences"; "[η] is the constant that indicates the convexity of log(D)"; and the rate parameter σ "determines the degree of discounting."

Looking first at the $\eta > 0$ part:

$$\sigma = log(\beta F/P) \, / \, e^{\eta T} \qquad (34)$$

The denominator $e^{\eta T}$ is a convex function, indicating increasing impatience. It is the analogous equation to the first part of 3.6.2, in the region $\psi > 1$.

For $\eta = 0$, we have:

$$\sigma = log(\beta F/P) \, / \, T \qquad (35)$$

which is the Exponential model, and for $\eta < 0$ we have:

$$\sigma = log(\beta F/P) \, / \, (-e^{\eta T}) \qquad (36)$$

$e^{\eta T}$ is a decaying function with time if $\eta < 0$, but because of the negative sign, the denominator becomes a concave increasing term, and is therefore analogous with 3.6.2, part(iii) and equation (33).

The same kind of remarks made for the CRDI functions, concerning algorithms and betas can be repeated here.

3.7 Size-sensitivity in the arithmetic model $(A_z, d_z; z)$

Returning to the arithmetic model, one shortcoming is its insensitivity to size. That is, if someone prefers to receive \$11 in 2 days rather than \$1 today, model A assumes that person would also prefer \$1,000,011 in 2 days to \$1,000,001 today. Both imply d = 5 so that if d exceeds the internal d_0 in the first scenario, implying that F should be preferred to P, d will also exceed d_0 in the second. However, in the second scenario, just contemplating the impact of becoming a millionaire right now is likely to affect the internal criterion d_0 in the sense of requiring greater "wages for waiting." A more flexible model is therefore to suggest that the rate parameter may vary with the size of P. For instance, d may be given by a power law scaling of a rate parameter d_z which is stable and does not vary depending on the size of the choices offered,. Hence, $d = P^z d_z$. Therefore, we have:

$$d_z = P^{-z}(F - P) \, / \, T \qquad (37)$$

If z = 0, d_z is not size-sensitive at all, and we have model A. If z = 1, d_z is highly size-sensitive and in fact we have the simple hyperbolic model H. Thus the size parameter z links A with H.

An alternative is to scale d by F^{-z} (Kirby, 1997), though see Section 5.1 for arguments why P is preferred to F.

4 Rate parameter + 2 models

4.1 Killeen's additive utility model (K, k; m, τ)

Killeen (2009) started from the assumption that the marginal change in utility with respect to time follows a power law; also, that utility and value are themselves related by a power law. Taken together these were shown to yield an equation of the form: $P = (F^m - kT^\tau)^{1/m}$ (Killeen, 2009, equation 6, p. 605), where his v_t, v, α, β, t, and λ are respectively our P, F, m, τ, T,and k. He explained as follows: "this additive utility discount function is the central contribution of this article. It is additive because the (negative) utility of a delay is added to the nominal utility of the deferred good." (p. 605). Some rearrangement isolates the rate parameter to give our canonical form:

$$k = (F^m - P^m) \, / \, T^\tau \qquad (38)$$

Clearly, when both m = 1 and τ = 1, we have the simple arithmetic model. The numerator can be written as:

$(F^m - 1) - (P^m - 1) = m\frac{F^m-1}{m} - m\frac{P^m-1}{m}$

So that as m $\rightarrow$ 0, it becomes:

m.log(F) – m.log(P) = m.log(F/P)

But since m is a common factor, it can be treated as a scaling constant, meaning that K nests E when m $\rightarrow$ 0 and τ = 1. It also nests the particular t = 0 case of Read's model, which occurs in K when m $\rightarrow$ 0 and τ is a free parameter. Killeen's used his model to survey aggregated data from past research, finding that m $\simeq$ .15, and $\tau \simeq$ .53. Both parameters are smaller than other researchers have estimated, and in particular m < τ, which is contrary to evidence collected by researchers in connection with H_m and H_τ. Note, however, that within the datasets that Killeen examined, the number of observations used to fit the model was never greater than eight, which is rather small for non-linear estimation.

4.2 Killeen, intervals, and started time (K$_\tau$, k$_\tau$; m, τ)

Killeen's additive utility model can be generalized to time intervals that start in the future (at time t $\neq$ 0), either with a power law on the interval, as modeled in Read (2001) and Section 3.3:

$$k_\theta = (F^m - f^m) \, / \, (T - t)^\theta \qquad (39)$$

or with a power on each point in time itself:

$$k_\theta = (F^m - f^m) \, / \, (T^\theta - t^\theta) \qquad (40)$$

There is an important distinction between these models for time exponents θ or τ if they are $\ll$ 1, which exaggeratedly condense distant time. In our first generalization of

Killeen, as $\theta \to 0$, the denominator $(T - t)^{\theta} \to 1$, which means that people would be completely time-insensitive, and therefore presumably always choose F over f. However, in the second generalization, as $\tau \to 0$, the denominator $(T^{\tau} - t^{\tau}) \to \tau.\log(T/t)$, meaning that at this extreme people become only logarithmically (in)sensitive to time. Started time does not affect (39) since the start is added to both T and t, which cancel out.

Because of the symmetry in the treatment of money and time in the k_{τ} generalization[3], and to allow the possibility that subjective time is logarithmic, as argued in recent research (Zaubermann et al, 2009), we now focus attention on k_{τ}. But immediately we hit a problem, which is that if t = 0 (as it usually is in intertemporal choice questionnaires), equation (39) cannot be distinguished from (40), so that we end up with the same time insensitivity as in k_{θ} when $\tau \to 0$. Persisting with the logarithmic limit idea doesn't help either because of divide-by-zero problems.

As argued earlier, because of log(0) problems a useful generalization of Killeen is to use started time[4]:

$$k_{\tau} = (F^m - f^m) / (T_*^{\tau} - t_*^{\tau}) \qquad (41)$$

where $T_* = T + \varepsilon$, and $t_* = t + \varepsilon$, and ε is small. This form is flexible enough to capture many possible states of the world, as reflected in the recovered exponents m and τ. If $m \to 0$, subjective money (F, f) would be logarithmic. If $\tau \to 0$ started time (T_*, t_*) would be logarithmic. If m and τ were close to 1, money and (started) time would be linear (objective), while intermediate values of m and τ would indicate that subjective time and money have less-than-logarithmic curvature. It is also possible that both m and / or τ could be > 1.

4.3 Dual hyperboloid with subjective money and time ($H_{m\tau}$, $h_{m\tau}$; m, τ)

As a robustness analysis, Doyle and Chen (2012) used power exponents to model time and money for A, which then becomes model K, and for the hyperbolic, which then becomes:

$$h_{m\tau} = ((F/P)^m - 1) / T^{\tau} \qquad (42)$$

This nests a number of models already considered. The rate parameter h_m for Green and Myerson's hyperboloid model is the special case when $\tau = 1$. The rate parameter h_{τ} for Rachlin's hyperboloid occurs when m = 1; and the

hyperbolic rate parameter h occurs when both m = 1 and $\tau = 1$. Furthermore, when $m \to 0$ we have:

$$h_{m\tau} = log(F/P) / T^{\tau} \qquad (43)$$

Which is model B (Read), and b = $h_{m\tau}$. Note also that the subjective-time exponential only has one additional parameter, because any attempt to introduce a new exponent for money results in $log(F^m/P^m) = log((F/P)^m)$ = mlog(F/P), so that m just becomes a scaling constant. The model Ebert and Prelec (2007) reject, equation (30), occurs when $m \to 0$.

In conclusion, most of the models considered so far can be generated from equation (42) as special cases. However, the CS model that Elbert and Prelec do propose, equation (28), has the exponent acting on all of log(F/P), not just what is in the brackets. Therefore, CS cannot be derived.

Similarly, the general hyperbola (Harvey, 1986) cannot be derived as a special case of equation (42). It has the form P = F / $(1 + \alpha T)^{\beta/\alpha}$. As Loewenstein and Prelec (1992) state: "The α-coefficient determines how much the function departs from constant discounting; the limiting case, as α goes to zero, is the exponential discount function, $e^{-\beta T}$." When $\beta=\alpha$, it is the simple hyperbolic, which we can write as P = F / $(1 + \beta T)$, as in Section 2.2. Therefore β may be treated as a rate parameter for the special case of both hyperbolic and exponential, and in the more general case, $\beta = \log(F/P) / \log(1 + \alpha T)^{1/\alpha}$. Finally, as α becomes large, the discounting function becomes increasingly step-like.

4.4 Benhabib et al.'s generalization of E, H, qH, and X (G, g; β, χ)

Benhabib, Bisin and Schotter (2004) presented their model as a discount function:

$$D = \beta(1-ngT)^{1/n}+\chi/F \ \ for \ T > 0; \ \ D = 1 \ \ for \ T = 0 \qquad (44)$$

Substituting D = P/F and re-arranging, we have:
$$(P + \chi) / (\beta F) = (1 - ngT)^{1/n}$$
When $n \to 0$, we have:
$$(P + \chi) / (\beta F) = e^{-gT}$$
Thus:

$$g = log(\beta F/(P + \chi)) / T, \ \ for \ T > 0 \qquad (45)$$

which is E if the quasi-hyperbolic multiplier $\beta = 1$, and the analogous additive term $\chi = 0$. If $\beta = 1$, and $\chi = 0$, and n = −1, we have H, the simple hyperbolic.

We begin to see here how elements of different models may be combined to good effect. In this example the authors were able to test whether the multiplier β or additive element χ was more predictive of behavior. In their

---

[3]If power laws govern the subjective perception of all simple psychophysical quantities, and by extension time and money, then the burden of proof is surely on justifying that time and money should not be treated symmetrically.

[4]Money never needs to be "started", because f = 0 is not a realistic scenario, though in principle money could be "started" too.

analyses, it was the latter. They also noted that their particular data had little power to distinguish between functional forms H (n = -1) and E (n = 0), though the main point to note here is that their generalization model does allow for testing this possibility.

4.5　Hyperboloid over intervals ($H_{m\omega}$, $h_{m\omega}$; m, ω)

Green, Myerson and Macaux (2005) extended Green and Myerson's model described in Section 3.1.1 to include time intervals that do not begin at t=0. Starting from their statement of the model in their equation A16, we derive the following equation for the rate parameter h_{mw}. Letting M = $(F/f)^m$ – 1 be the numerator, exactly as in Section 3.1.1, then:

$$h_{m\omega} = M \,/\, [(T - t) - \omega t M] \qquad (46)$$

Clearly, when t = 0, this model reduces to H_m. But when t > 0, the authors identified different models according to the values of ω and m.

Equation (46)a: ω = 0 is their *elimination by aspects* model. One aspect that is common to the two choices (f, t) and (F, T) is the wait to the first choice, and so it is eliminated, reducing the time component to (T – t). We should also note that f is common to the monetary choices, but this aspect is left untouched in 37a.

Equation (46)b: m = ω = 1 is their *present value comparison* model. In this model, people are assumed to discount both F and f to present values, using the simple hyperbolic model in Section 2.2 (i.e., m = 1).

Equation (46)c: m = 1 is their *common aspect attenuation* model. Once again, F and f are assumed to be hyperbolically discounted to a present value, but this model assumes that an extra attenuation factor, operationalized by the weight ω, may act on f.

One point to note from the general statement in (46) is that although a money-only component can be identified in the numerator, the subjective time component in the denominator is not independent of money, because the denominator in (46) has the additional term ωtM, and M has terms involving F and f. This contrasts with all other models presented thus far. Finally, equation (46) implies that if (F $\gg$ f), M could become large enough so that (T – t) < ωtM, thus turning the rate parameter negative.

5　Rate parameter + 3 models

5.1　Size-sensitive J model (J, j; z, m, τ)

If we combine the size-sensitivity component described in 3.7 with the started-time version of Killeen's model extended in Section 4.2 to cope with discounting over non-present time intervals, we get a general model that generates many of the above as special cases:

$$j = P^{-z}(F^m - P^m) \,/\, (T_*^\tau - t_*^\tau) \qquad (47)$$

First, we use the triplet (z, m, τ), each taking values of 0 or 1 to describe whether money and time are modeled as linear (m = 1, τ = 1) or logarithmic (m $\rightarrow$ 0, τ $\rightarrow$ 0); and whether there is size-sensitivity (if z = 0 there is not). J nests the following simpler models: Exponential is (0,0,1); Hyperbolic is (1,1,1); Arithmetic is (0,1,1); Roelofsma, if υ is take to be 1, is (0, 0, 0). Moreover, other models can be stated as partially constrained versions of J: Green and Myerson is (m, m, 1), i.e. z = m; Rachlin is (1, 1, τ); Read with t = 0 and t_* = 1 is (0, 0, τ); Size-sensitive Arithmetic is (z, 1, 1); the general hyperboloid $H_{m\tau}$ in 4.3 is (m, m, τ); and Killeen is (0, m, τ). Also nested in J are several of the memory trace models described in Yi, Landes, and Bickel (2009), namely "power" (0, 0, 0)—see also Roelofsma; "exponential power" (0, 0, .5), and "hyperbolic power" (1, 1, .5).

In being able to generate these alternative models, J has a number of useful properties. First, there is the simple matter of parsimony. We need hold in mind a single equation from which many others may be generated. Second, the generation process encourages us to consider and investigate other models that could be generated from the triplets, but have not. For instance, there are eight possible fully-constrained, simple rate parameter models. Each is a combination of the 0 or 1 for each of the triplet values z, m, and τ. Only four of these models have been considered: what of the other four? Similarly, though less mechanically, new models may be considered with partial constraints. Last, but not least, one may use model J with no constraints to recover parameters for z, m, and τ. The fact that so many models may be generated from this form hints that the best fitting model may be a compromise between of all of them, as should be evident if the parameters z, m, and τ were all to fall between 0 and 1.

This model scales by P^{-z} rather than F^{-z} because the latter is less productive in generating other models as special cases, and also because it is more likely that P, will be used as the *given situation*, against which F will be compared, rather than vice versa. After all, P is mentioned first in the typical question frame (e.g. "Would you prefer to receive \$70 now, or \$100 in 20 days?"), and is on offer right now. Thus F will be evaluated in terms of P. Nonetheless, it may be possible to prime F to fulfill the role of given information, as in the question frame: "Assuming you can receive \$100 in 20 days' time, would you accept \$70 now instead?" Thus biasing P to be evaluated in terms of F. Manipulating the salience of P relative F, or vice versa, can be taken a stage further in matching tasks where people estimate their indifference points be-

tween P and F. Thus if people are required to adjust P for a fixed F they will be biased towards scaling by F^{-z}; whereas if they adjust F for a given P, they will be biased towards scaling by P^{-z}. Using F^{-z} gives rise to a parallel set of models to those presented here. But the degree to which P^{-z} outperforms F^{-z} as a scaling component, or vice versa, and their susceptibility to manipulation are all matters for future empirical investigation.

5.2 Discounting by intervals, DBI (I, I; m, τ, θ)

Like Read's model, Scholten and Read's (2006) discounting by intervals (DBI) model is designed to highlight phenomena which occur over several intervals that do not necessarily include the special one bounded at t=0. Nonetheless, as a viable model it must be able to stand comparison with other models when t=0. The rate parameter for DBI is:

$$i = ((F/f)^m - 1) \,/\, (T^\tau - t^\tau)^\theta \qquad (48)$$

This model extends Read's in two ways. First, money is modeled as in Green and Myerson with a to-be-estimated exponent m, rather than with log(F/f), as in Read. This is more flexible because the power form nests the logarithmic money form. Second, although there is still an exponent on the time interval, the times in the interval are themselves subjective. Therefore, Read's model is nested in DBI as the special case where m → 0, and τ = 1. Note that if t = 0, the denominator becomes $T^{\tau\theta}$. Given that $\tau\theta$ only occur together, they can be treated as a single parameter, and DBI becomes the generalized hyperboloid $H_{m\tau}$. Also, if θ = 1, we approximate J (m, m, τ) if τ is sufficiently far from 0 so that started and objective time are effectively equivalent.

Finally, the nesting of time and time-interval exponents in the denominator $(T^\tau - t^\tau)^\theta$ means that care must be taken in experimentally estimating the separate parameters.

6 Discounting models with no closed-form rate parameter

6.1 Two rate parameters—two exponentials discounting ($E_{\beta\delta}$, r_β, r_δ; w)

This model is related to the quasi-hyperbolic model in its intent to separate short-term from long-term processes. But, whereas qH separates events at t=0 into a qualitatively distinct category from those at t>0, this model deals with the same short/long-term issue in a more graded manner. Carrying over the same terminology as

used in qH, McClure, Laibson, Loewenstein, and Cohen (2007) suggested a β-system associated with limbic areas of the brain, that is impulsive, myopic, and discounts at a high rate; and a δ-system, associated with prefrontal and parietal cortical regions ("higher man"), which discounts at lower rates. Each of these sub-systems is assumed to discount according to the exponential model, with the overall discount rate being a weighted sum of each sub-system:

$$P/F = w.exp(-r_\beta T) + (1 - w).exp(-r_\delta T), \quad (49)$$

where if the first term represents the β-system, then $r_\beta >$ r_δ. Quite explicitly, there are two rate parameters.

6.2 Scholten and Read's intertemporal tradeoff model—7 parameters (Y, γ, τ, ε_m, ε_τ, and three of {a_1, a_2, b_1, b_2})

Scholten and Read (2010) provide a densely argued justification for a complex model which they present in terms of an indifference point between the perk of additional compensation, and the irk of waiting for it. They define the money value function on a future payment F to be:

$$v(F) = (1/\gamma)log(1 + \gamma F) \qquad (50)$$

and a time-weighting function to be:

$$w(T) = (1/\tau)log(1 + \tau T). \qquad (51)$$

They define an "effective compensation" to be:

$$V_m = (v(F) - v(f)), \;\; where \; F > f \qquad (52)$$

and an "effective interval" between t and T to be:

$$W_\tau = (w(T) - w(t)), \;\; where \; T > t. \qquad (53)$$

The point of indifference is when:

$$Q_\tau(W_\tau) = Q_m(V_m) \qquad (54)$$

such that if $Q_\tau(W_\tau) > Q_m(V_m)$ then the irk of waiting is greater than the perk of compensation, so DM chooses f at t, rather than F at T, and vice versa if $Q_\tau(W_\tau) <$ $Q_m(V_m)$. However, Q_τ and Q_m are not simple scaling factors, but two-part linear functions:

$$Q_\tau(x) = a_1 x \; if \; x < \varepsilon_\tau$$
$$Q_\tau(x) = a_2 x \; if \; x \geq \varepsilon_\tau, with \; a_1 > a_2. \qquad (55)$$

ε_τ is a threshold below which effective time intervals (here just x) are weighted more steeply than above-threshold. Note the similarity of intent with the two exponentials model. Similarly,

$$Q_m(x) = b_1 x \ \ if \ x < \varepsilon_m$$
$$Q_m(x) = b_2 x \ \ if \ x \geq \varepsilon_m, \ \ with \ b_1 > b_2 > 0. \quad (56)$$

If the indifference equation is written out in full, it is possible to see that a discount function of the form f/F cannot be extracted: nor does limiting analysis to t=0 overcome the problem. Similarly, no simple rate parameter emerges from this model. It follows that the indifference equation is the only simple way to present the model. The model requires three parameters to specify the two-part linear function for time (a_1, a_2, ε_τ), three more to specify the equivalent parameters for money (b_1, b_2, ε_m), as well as τ and γ to specify the time-weighting and value functions, respectively. However, the authors note that "if only the relative magnitude of scaled effective differences is of interest, one of these scaling constants $\{a_1, a_2, b_1, b_2\}$ can be set to unity." (p. 934). This model therefore requires seven parameters to be estimated. So clearly, one of the biggest problems in testing this model is the number of parameters to be recovered, which means designing choice problems over which the parameters are sufficiently independent of each other, and the number of observations sufficiently large in order to estimate the parameters adequately.

The tradeoff model treats time and money symmetrically, which we argued in 1.3.4 should be the default position for a quasi-psychophysical theory of time and money. Indeed, the authors explicitly deny that people are discounting in any way that an accountant would recognize, espousing instead a thoroughly psychological perspective on intertemporal choice.

7 Model families and newborns

7.1 Models have similar forms

The world is simpler than a list of twenty or so models might suggest. Our presentation of models via their rate parameters emphasizes the following common structure:

$$Rate = s(money) \ \otimes \ S(time), \quad (57)$$

where s(.) and S(.) are functions that return subjective perceptions of the money and time aspects, respectively, and $\otimes$ is an operation by which these become combined. According to this view, if someone operates according to the exponential model it is because his/her subjective perception of money is logarithmic, while his/her subjective perception of time is linear. In (57), subjective perceptions then get combined by the operation $\otimes$. In most cases, $\otimes$ is arithmetic division, but can be subtraction, or conceivably something else. The separability of s and S in the mathematics implies that subjective perceptions of time and money should take place *independently of each other*. In most models, objective time and money are

transformed into subjective equivalents by power laws, or logs, but other transformations are possible.

To further simplify the map of the models surveyed here, note that half are special cases of model J, and the present-bias models of qH and X could also easily be incorporated into J. Two further themes that lie outside J are: the time impatience models of Section 3.6 that go to town on the modeling of time itself; and models that place power laws on *time intervals* rather than on *points of time*—Sections 3.3 and 5.2. Finally, we are left with just four models that are not captured by these themes. First there is the two-exponentials model—which is just a doubling up of the reference model E, as if we had two normative discounters in the same head: one patient (δ), the other impatient (β). Then there is Schotten and Read's (2010) trade-off model I, which for all its complexity is still recognisable as an assemblage of parameter-based treatments of subjective money, time, and intervals. $H_{m\omega}$, although an innocuous extension of the hyperboloid to time intervals turns out to be the one model for which time and money cannot be cleanly separated.

7.2 Speculative models

We end with a sample of speculative models (newborns) that have clearly been assembled from the components of existing models. Being able to mix and match almost indefinitely emphasizes the family resemblances that exist between models, and helps to compare and contrast treatments of time and money. Some of these examples are semi-serious; others are intended to provoke thought.

7.2.1 Arithmetic interval model

One of the innovations in B and I is that intervals and subjective intervals are treated as if psychophysical elements in their own right, which may then be investigated using power laws, and so on. We cross this with Killeen (extended) to give the hybrid:

$$r = (F^m - f^m)^\mu \ / \ (T^\tau - t^\tau)^\theta \quad (58)$$

For mnemonic value note that μ is "mu" Greek m; and θ "theta" sounds vaguely like T.

7.2.2 Fully additive model

Instead of treating $\otimes$ as a division in 7.2.1, we could treat it as a subtraction, echoing the difference models examined in Stevenson (1986):

$$r = (F^m - f^m)^\mu \ - \ w(T^\tau - t^\tau)^\theta \quad (59)$$

The weighting coefficient w cannot simply be interpreted as the perceived relative importance of time, because it

also rescales for m, μ, τ, θ, and units of time and money. However, Stevenson found little support for this way of combining time and money, compared with multiplicative (ratio) forms.

7.2.3 Fully multiplicative hyperbolic model

Analogously, if H treats money as a percentage increase (that gets averaged over time), then a model that also treated time as a percentage increase would look like:

$$r = [(F/f)^m - 1] \, / \, [(T_*/t_*)^\tau - 1] \qquad (60)$$

This example highlights the point that the default setting in models is to treat time as additive, whereas money is treated as multiplicative. From the point of view of Stevens' (1946) levels of measurement, both are ratio scale measurement, with properly defined zeros. So if in theory we can treat time and money symmetrically, what is to stop the naïve behavioral theorist thinking likewise?

7.2.4 Time-size and money-size sensitive model

Model J only treated money as size-sensitive. A symmetrical treatment of time and money is instead:

$$r = t_*^g (F^m - f^m)^\mu \, / \, [f^z (T_*^\tau - t_*^\tau)^\theta], \qquad (61)$$

where g is the time analog of z for money, implying that someone's criterion rate parameter may vary, not just with the size of f, but also how far in the future f is offered.

7.2.5 Arithmetic present premium model

A present premium can be incorporated into other models. Using both the multiplicative and additive components of Benhabib et al's (2004) generalizing model in Section 4.4, the arithmetic model would become, for instance:

$$r = (\beta F - (P + \chi)) \, / \, T. \qquad (62)$$

8   Conclusions

This paper has surveyed delay discounting models that have appeared in several literatures. Our presentation emphasizes that models have a clearly separable time component, a separable money component, and a method of combining these components into a rate parameter / decision parameter. Within the components that transform objective monies and times and intervals into subjective equivalents, we have seen that the power law occupies a special place in delay discounting models, particularly given that log laws may be derived as special cases. However, despite the many models presented in this survey,

it may turn out that none of them portrays ITC behavior particularly accurately—the door must always be left ajar to improvements. Although researchers have been inventive in their modeling of the time and money components, there has been less interest shown in exploring alternatives to power laws (which have been taken as the basic nuts and bolts of discounting models), or how the time and money components should be combined. Whereas past research has favored power laws over log laws or the negative exponential to represent the utility function, nonetheless even *it* may be flawed. For instance, Kirby (2011) has shown that power laws depart significantly from utility, as incidentally do the other standard forms considered. Kirby's findings have yet to be extended from utility in risky choice to utility in intertemporal choice. Even more radically, the Decision-by-Sampling model (Stewart, 2009) foregoes the parameterisation of utility and time in favor of a memory-based approach. Another possible direction for future research in individual decision making is to more explicitly model the "fuzzy math" (Stango & Zinman, 2009) of the typical insufficiently competent mathematician who is required to think about financial and other numerically presented choices. But given the large numbers of models and variants that have appeared in the literature, the re-usability of components, and the ease with which speculative models can be assembled from them, a chief problem for future research is to heed Occam's razor by limiting the number of parameters a model employs to only those that are strictly necessary. Finally, by gathering these models into one place and thereby shining a light on them, it is hoped that future research will be more proactive in testing models against each other, and in manipulating choice behavior through selectively manipulating the parameters that are supposed to drive those behaviors.

References

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*, 463-496.

Ainslie, G., & Herrnstein, R. J. (1981). Preference reversals and delayed reinforcement. *Animal Learning & Behavior, 9(4)*, 476-487.

Albrecht, K., Volz, K. G., Sutter, M., Laibson, D. I., & von Cramon, D. Y. (2011). What is for me is not for you: Brain correlates of intertemporal choice for self and other. *Social Cognitive and Affective Neuroscience*, *6*, 218-225.

Andersen, S., Harrison, G. W., Lau, M. & Rutström, E. E. (2011). Discounting behavior: A reconsideration. Unpublished MS, available from http://cear.gsu.edu/files/.

Benhabib, J., Bisin, A., & Schotter, A. (2004). Hyperbolic Discounting: An Experimental Analysis. Available at: http://elsa.berkeley.edu/users/webfac/dellavigna/e218_sp05/bisin.pdf.

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior, 69*, 205-223.

Bleichrodt, H., Rohde, K. I. M., & Wakker, P. P. (2009). Non-hyperbolic time inconsistency. *Games and Economic Behavior, 66*, 27-38.

Box, G. P. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B, 26*, 211-252.

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P. & Taubinsky, D. (2008). Measuring intertemporal preferences using response times. *NBER Working Paper Series, 14353.*

Chapman, G. B., & Elstein, A. S. (1995). Valuing the future: Temporal discounting of health and money. *Medical Decision Making,15*, 373-386.

Chung, S.-H., & Herrnstein, R. J. (1967). Choice and delay of reinforcement. *Journal of the Experimental Analysis of Behavior, 10(1), 67-74.*

Doyle, J. R., & Chen, C. H. (2012). The wages of waiting and simple models of delay discounting. http://ssrn.com/abstract=2008283.

Doyle, J. R., Chen, C.H., & Savani, K. (2011). New designs for research in delay discounting. *Judgment and Decision Making, 6*, 759-770.

Ebert, J. E., & Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of the near and far future. *Management Science, 53*, 1423-1438.

Eisler, H. (1976). Experiments on subjective duration 1968-1975: A collection of power function exponents. *Psychological Bulletin, 83*, 1154-1171.

Estle, S. J., Green, L., Myerson, J., & Holt, D. D. (2007). Discounting of money and directly consumable rewards. *Psychological Science, 18*, 58-63.

Green L., & Myerson, J. (2004). A discounting framework for choice with delayed and probabilistic rewards. *Pyschological Bulletin, 130*, 769-792.

Green, L., Myerson, J., & Macaux, E. W. (2005). Temporal discounting when the choice is between two delayed rewards. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1121-1133.

Hardisty, D. J. & Weber, E. U. (2009). Discounting future green: Money versus the environment. *Journal of Experimental Psychology: General, 138*, 329-340.

Hardisty, D. J., Appelt, K. C., & Weber, E. U. (2012). Good or bad, we want it now: Fixed-cost present bias for gains *and* losses explains magnitude asymmetries in intertemporal choice. http://ssrn.com/abstract=1979828.

Harvey, C. M. (1986). Value functions for infinite-period planning. *Management Science, 32*, 1123-1139.

Herrnstein, R. J. (1981). Self-control as response strength. In C. M. Bradshaw, E. Szabadi & C. F. Lowe (Eds.), *Quantification of steady-state operant behavior* (pp. 3-20). Amsterdam: Elsevier/North Holland Biomedical Press.

Kable, J. W., & Glimcher, P. W. (2010). An "as soon as possible" effect in human intertemporal decision making: Behavioral evidence and neural mechanisms. *Journal of Neurophysiology, 103*, 2513-2531.

Killeen, P. R. (2009). An additive-utility model of delay discounting. *Psychological Review, 116*, 602-619.

Kirby, K. (2011). An empirical assessment of the form of utility functions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 461–476.

Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General, 126*, 54-60.

Kirby, K. N., & Marakovic, N. N. (1996). Delay discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review, 3*, 100-104.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards then non-drug-using controls. *Journal of Experimental Psychology*: *General, 128*, 78-87.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics, 112*, 443-477.

Laibson, D. (2003). Intertemporal decision making. *Encyclopedia of Cognitive Science.* London: Nature Publishing Group.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and interpretation, *Quarterly Journal of Economics, 107*, 573–597.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analysis of behavior, Vol. 5.* Mahwah NJ: Erlbaum.

Madden, G. J., & Bickel, W. K. (2009). *Impulsivity: The Behavioral and Neurological Science of Discounting.* Washington, DC: American Psychological Association.

McClure, S. M., Laibson, D. I., Loewensein, G., & Cohen, J. D. (2007). Time Discounting for Primary Rewards. *Journal of Neuroscience, 27*, 5796-5804.

McKerchar, T. L., Green, L., & Myerson, J. (2010). On the scaling interpretation of exponents in hyperboloid models of delay and probability discounting, *behavioral Processes, 84*, 440-444.

McKerchar, T. L., Green, L., Myerson, J., Pickford, T. S., Hill, J.C., & Stout, S. C. (2009). A comparison of

four models of delay discounting in humans, *behavioral Processes, 81*, 256–259.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison Wesley.

Myerson J., & Green L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of Experimental Analysis of Behavior, 64*, 263–276.

Myerson J., Green L., & Warusawitharana, M. (2001). Area under the curve as a measure of discounting. *Journal of the Experimental Analysis of Behavior, 76(2)*, 235-243.

Peters, E., Slovic, P., Västfjäll, D., & Mertz, C. K. (2008). Intuitive numbers guide decisions. *Judgment and Decision Making, 3*, 619-635.

Prelec, D., & Loewenstein, G. (1991). Decision making over time and under uncertainty: A common approach. *Management Science, 37*, 770-786.

Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behavior, 85,* 425-435.

Rachlin, H., Raineri, A., Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior, 55*, 233-244.

Read, D. (2001). Is time discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty, 23*, 5-32.

Roelofsma, P. H. M. P. (1996). modeling intertemporal choices: An anomaly approach. *Acta Psychologica, 93*, 5-22.

Rubinstein, M. (2003). Great moments in financial economics: I, present value. *Journal of Investment Management, 1*.

Ryan, R. (2007). *Corporate Finance and Valuation*, London: Thomson.

Scholten, M., & Read, D. (2006). Discounting by intervals: A generalized model of intertemporal choice. *Management Science, 52*, 1424-1436.

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review, 117*, 925-944.

Sen, S., & Johnson, E. J. (1997). Mere possession effects without possession in consumer choice. *Journal of Consumer Research, 24(1)*, 105-117.

Smith, C. L., & Hantula, D. A. (2008). Methodological considerations in the study of delay discounting in intertemporal choice: A comparison of tasks and modes. *Behavior Research Methods, 40*, 940-953

Stango, V., & Zinman, J. (2009). Exponential growth bias and household finance. *Journal of Finance, 64*, 2807-2849.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, 103*, 677–680.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*, 153–181.

Stevens, S. S. (1961). To honor Fechner and repeal his law. *Science, 133,* 80-86.

Stevenson, M. K. (1986). A discounting model for decisions with delayed positive or negative outcomes. *Journal of Experimental Psychology: General. 115*, 131-154.

Stewart, N. (2009). Decision by sampling: The role of the decision environment in risky choice. *Quarterly Journal of Experimental Psychology, 62*, 1041-1062.

Tanaka, T., Camerer, C. F., & Nguyen, Q. (2010). Risk and time preferences: linking experimental and household survey data from Vietnam. *The American Economic Review, 100*, 557-571.

Thaler, R. (1980). Towards a positive theory of consumer choice. *Journal of Economic Behavior and Organization, 1(1)*, 39-60.

Tsukayama, E. & Duckworth, A. L. (2010). Domain-specific temporal discounting and temptation. *Judgment and Decision Making, 5*, 72-82.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison Wesley.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232.

Van de Ven, J., & Weale, M. (2010). An empirical investigation of quasi-hyperbolic discounting. *NIESR, National Institute of Economic and Social Research, Discussion Paper 355.*

Van den Bergh, B., Dewitte, S., & Warlop, L. (2008). Bikinis instigate gemeralized impatience in intertemporal choice. *Journal of Consumer Research*, 35, 85-97.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, 17, 1329-1344.

Weatherly, J. N., Terrell, H. K., and Derenne, A. (2010). Delay discounting in different commodities. *Journal of General Psychology, 137*, 273-286.

Wileyto, P. E., Audrain-McGovern, J., Epstein, L. H., & Lerman, C. (2004). Using logistic regression to estimate delay-discounting functions. *Behavior Research Methods, Instruments, & Computers, 36*, 41-51.

Wilson, M., & Daly, M. (2004). Do pretty women inspire men to discount the future? *Proceedings of the Royal Society of London B (Supplement), 271,* S177-S179.

Yi, R., Landes, R. D., & Bickel, W. K. (2009). Novel models of intertemporal valuation: Past and future outcomes. *Journal of Neuroscience, Psychology, and Economics*, 2, 102-111.

Zauberman, G., Kim, B. K., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: subjective time perception and intertemporal preferences. *Journal of Marketing Research, 46*, 543-556.