Cognitive Psychology 111 (2019) 53–79

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Variable time preference

Lisheng He[a,*], Russell Golman[b], Sudeep Bhatia[a,1]

[a] University of Pennsylvania, United States
[b] Carnegie Mellon University, United States

ARTICLE INFO

Keywords:
Decision making
Intertemporal choice
Stochastic choice
Preference variability
Computational modeling

ABSTRACT

We re-examine behavioral patterns of intertemporal choice with recognition that time preferences may be inherently variable, focusing in particular on the explanatory power of an exponential discounting model with variable discount factors – the *variable exponential model*. We provide analytical results showing that this model can generate systematically different choice patterns from an exponential discounting model with a fixed discount factor. The variable exponential model accounts for the common behavioral pattern of decreasing impatience, which is typically attributed to hyperbolic discounting. The variable exponential model also generates violations of strong stochastic transitivity in choices involving intertemporal dominance. We present the results of two experiments designed to evaluate the variable exponential model in terms of quantitative fit to individual-level choice data. Data from these experiments reveal that allowing for a variable discount factor significantly improves the fit of the exponential model, and that a variable exponential model provides a better account of individual-level choice probabilities than hyperbolic discounting models. In a third experiment we find evidence of strong stochastic transitivity violations when intertemporal dominance is involved, in accordance with the variable exponential model. Overall, our analytical and experimental results indicate that exponential discounting can explain intertemporal choice behavior that was supposed to be beyond its descriptive scope if the discount factor is permitted to vary at random. Our results also highlight the importance of allowing for different sources of randomness in choice modeling.

1. Introduction

Intrinsic variability plays a central theoretical role in psychological research on high-level cognition. The assumption of variability, often in the form of random model parameters, not only explains inconsistencies in individuals' responses across multiple identical trials; when allowed to interact with attention, memory, and valuation, it is also capable of generating a systematic effect on behavior. In recent years, this type of variability has been shown to account for errors and response time patterns in perceptual and lexical choice (Brown & Heathcote, 2008; Ratcliff & Rouder, 1998), biases in probability judgment (Costello & Watts, 2014; Hilbert, 2012) and social judgment (Erev, Wallsten, & Budescu, 1994; Fiedler & Unkelbach, 2014), paradoxes in risky decision making (Bhatia & Loomes, 2017; Denrell, 2015), the appearance of inconsistent or intransitive preferences (Regenwetter, Dana, & Davis-Stober, 2011; Tsetsos et al., 2016) and contextual preference reversals (Howes, Warren, Farmer, El-Deredy, & Lewis, 2016). In many of these cases, variability added to a standard (baseline) model is enough, by itself, to provide a rich account of observed behavioral patterns, making additional –more complex– psychological assumptions about irrational biases in the judgment or decision process

* Corresponding author at: Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States.
E-mail address: hlisheng@sas.upenn.edu (L. He).

[1] Funding for Sudeep Bhatia was received from the National Science Foundation grant SES-1626825.

https://doi.org/10.1016/j.cogpsych.2019.03.003
Accepted 18 March 2019

unnecessary.

In this paper we provide a formal characterization, theoretical analysis, and experimental test, of the role of preference variability in intertemporal choice, that is, choice between rewards occurring at different points in time. Many common choices involve intertemporal tradeoffs. For example, we may choose to receive occupational training (at the opportunity cost of earning money sooner) for a better-paid job in the future; we may buy a car now with a monthly instalment in the future; or we may exercise now for better health later. In intertemporal choice settings, exponential discounting is often considered to be the normative model, as it satisfies a set of highly compelling axiomatic restrictions on choice behavior (e.g. dynamic consistency, Koopmans, 1960). However, psychological research has documented violations of exponential discounting, such as decreasing impatience, the finding that discount factors appear to increase over time (Green, Fristoe, & Myerson, 1994; Kirby & Herrnstein, 1995; Thaler, 1981; see Read, McDonald, & He, 2018 for a recent review of the violations). For example, people may prefer to receive $10 now rather than $12 tomorrow, but also prefer to receive $12 in 366 days rather than $10 in 365 days. This choice behavior is typically seen as evidence in favor of alternative discounting models, such as those equipped with a hyperbolic or quasi-hyperbolic discount function (Laibson, 1997; Loewenstein & Prelec, 1992; Mazur, 1987; O'Donoghue & Rabin, 1999).

Here, we revisit the explanatory power of exponential discounting by considering different sources of variability in the decision process. Like many other decision models, exponential discounting is generally modeled alongside some stochasticity in response generation, with decision makers occasionally making mistakes in translating discounted utilities into choices (Blavatskyy & Pogrebna, 2010; Luce, 1959; McFadden, 1974; Thurstone, 1959; Wilcox, 2008; see e.g. Blavatskyy & Maafi, 2018; Peters, Miedl, & Büchel, 2012; Scholten, Read, & Sanborn, 2014 for applications in exponential discounting). However, there may also be some intrinsic variability in time preferences, that is in people's desire to be patient or impatient (e.g. Lu & Saito, 2018; see Becker, DeGroot, & Marschak, 1963; Gul & Pesendorfer, 2006; Loomes & Sugden, 1995; Regenwetter & Marley, 2001 for extensive discussion of preference variability in decision making). How can we formally incorporate this type of preference variability within an exponential discounting model? Can the resulting model accommodate choice patterns such as decreasing impatience, and does it generate novel predictions that distinguish it from discounting models without preference variability? Finally, can such a model provide an adequate quantitative account of individual behavior in intertemporal choice tasks?

We address these questions by proposing an exponential discounting model that allows for both stochasticity in response generation (in the form of a logit choice rule mapping discounted utilities to choice probabilities) and trial-to-trial variability in the discount factors that generate these utilities. We examine the properties of this *variable exponential model* with analytical derivations, simulations, and three experiments. Our analysis tests the descriptive validity of the variable exponential model and evaluates when it is and is not necessary to deviate from exponential discounting to describe behavioral patterns in intertemporal choice data. By performing these tests, we hope to obtain a deeper understanding of the effects of preference variability in intertemporal choice, complementing the rich existing theoretical literature on the intrinsic variability in cognition and behavior.

2. Temporal discounting

The simplest intertemporal choice task requires a decision maker to evaluate an option $X = (x, t)$, offering a payoff x with a time delay of t. In line with most existing intertemporal choice research, our analysis focuses on monetary gains, although the same idea applies to the discounting of monetary losses and non-monetary goods. Discounting models of intertemporal choice assume that people choose the option that has the highest discounted utility, which weighs the payoff based on the magnitude of the time delay. Thus, for a discount function $d(\cdot)$, the discounted utility of $X = (x, t)$ is given by:

$$U(X) = d(t)\cdot x \tag{1}$$

In a choice set consisting of many different choice options with payoffs at different time delays, the option with the highest discounted utility, according to Eq. (1), is the one that is chosen. When each option offers multiple payoffs (each with a different time delay), discounting models assume that the payoffs are individually discounted based on their time delay, and aggregated into a single utility measure.

Note that it is sometimes assumed that payoffs are transformed non-linearly according to a value function, prior to being discounted (Samuelson, 1937; Loewenstein & Prelec, 1992; see Dai & Busemeyer, 2014; Regenwetter et al., 2018; Scholten et al., 2014 for concrete examples). The non-linear value function is particularly popular for modeling risky choice, where it generates risk aversion or risk seeking. However, a non-linear value function does not have a similar explanatory role in intertemporal choice (that is, it does not, by itself, explain any well-known intertemporal choice effect). For example, assuming $v(x) = x^{\gamma}$ does not allow us to accommodate the magnitude effect in intertemporal choice (the finding that people are more patient for larger payoffs than for smaller payoffs) (Green, Myerson, & McFadden, 1997; Loewenstein & Prelec, 1992). Likewise, permitting nonlinearity in payoff valuation does not help explain the decreasing impatience effect (such an effect holds independently of any underlying value function). We assume a linear value function for simplicity.

2.1. Exponential discounting

Exponential discounting, initially introduced by Samuelson (1937), involves a parsimonious discount function. For a per-period discount factor δ, it assumes that the discounted weight on the payoff value is simply given by:

$$d(t) = \delta^{t} \tag{2}$$

Here smaller values of δ correspond to increased discounting (that is, less patience), and generate smaller weights on later payoffs relative to sooner payoffs. It is assumed that $0 \leq \delta \leq 1$. The extreme case $\delta = 1$ corresponds to a complete absence of time preference. Without additional probabilistic specifications, the exponential discounting model predicts that people deterministically select the option with the highest discounted utility.

This one-parameter discount function is not only parsimonious; it also derives from a small set of simple axiomatic assumptions about intertemporal choice (Koopmans, 1960). For this reason, exponential discounting is sometimes considered to be the normative model of intertemporal choice, and is the most commonly used discounting model in intertemporal choice research, especially for applications in economics (see Frederick, Loewenstein, & O'Donoghue, 2002 for a discussion). Exponential discounting is also often used in cognitive models, particularly those involving reinforcement learning (e.g. Mnih et al., 2015).

2.2. Decreasing impatience

Exponential discounting, however, has shortcomings from a descriptive perspective. Notably it is unable to account for observed patterns of *decreasing impatience* for intertemporal choices. For a choice between $X = (x;\, t)$ and $Y = (y;\, t + k)$, a decision maker displays decreasing impatience if the choice switches from X to Y as t increases (but k is kept constant). Previous studies suggest that people tend to display decreasing impatience for intertemporal choice problems (e.g. Diamond & Köszegi, 2003; O'Donoghue & Rabin, 1999; Kirby & Maraković, 1995). Exponential discounting, however, predicts that changing the delay to the sooner payoff (t) without changing the difference in delays between the payoffs (k), should not influence the utility ranking of the two options (Green et al., 1994; Kirby & Herrnstein, 1995; Millar & Navarick, 1984; Solnick, Kannenberg, Eckerman, & Waller, 1980; also Ainslie & Herrnstein, 1981; Green, Fisher, Perlow, & Sherman, 1981 for evidence of similar violations in animals).

Consider, for example, two related choice problems: A proximal choice between options X_P = (\$5; 0 months) and Y_P = (\$10; 1 month) and a remote choice between options X_R = (\$5; 1 month) and Y_R = (\$10; 2 months). As both the payoffs and the difference in time delays are the same for the proximal and remote choices, exponential discounting with a fixed discount factor δ predicts that participants should either select the sooner payoff in both choices or the later payoff in both choices. However, participants may select the sooner payoff in the proximal choice, but the later payoff in the remote choice, which is an example of decreasing impatience.

2.3. Hyperbolic discounting

In response to such violations of a fundamental implication of exponential discounting, researchers have suggested that the shape of the discount function is not exponential but hyperbolic. The conventional hyperbolic discounting model has one free parameter, $\alpha \geq 0$ (Mazur, 1987), so that the discount function is:

$$d(t) = \frac{1}{1 + \alpha t} \tag{3}$$

This model retains the parsimony of the one parameter exponential discounting model, while also being able to accommodate findings of decreasing impatience. Specifically, the shape of the hyperbolic discount function is such that changing the delay to the sooner payoff without changing the difference in delays between the payoffs can alter the utility ranking within pairs of options. However, some researchers have found that this one-parameter model over-predicts decreasing impatience (i.e. is "too hyperbolic") and is thus unable to provide an adequate account of intertemporal choice behaviors observed in the laboratory (see e.g. Luhmann, 2013; Read & Roelofsma, 2003).

Two generalized hyperbolic discounting models allow for more flexibility in the hyperbolic discount function by incorporating additional parameters. One was initially proposed by Mazur (1987). It takes the hyperbolic discounting model in Eq. (3) as the base model and applies a non-linear time perception function by raising the delay t to a power $\tau \geq 0$ (Stevens, 1957; see Han & Takahashi, 2012; Takahashi, Oono, & Radford, 2008; Zauberman, Kim, Malkoc, & Bettman, 2009 for evidence of non-linear subjective perception of time), so that the discount function becomes:

$$d(t) = \frac{1}{1 + \alpha t^{\tau}} \tag{4}$$

The other generalized hyperbolic discounting model was proposed by Loewenstein and Prelec (1992). Instead of raising t to a power, this model raises the whole denominator to a power β/α, with $\beta \geq 0$. Thus, the discount function becomes:

$$d(t) = \frac{1}{(1 + \alpha t)^{\beta/\alpha}} \tag{5}$$

With the additional parameters, the two generalized hyperbolic discounting models are able to modulate the extent of hyperbolic discounting. This allows them to predict decreasing impatience more accurately than the one-parameter model. Both generalized models reduce to the one parameter model as a special case (when $\tau = 1$ or when $\beta = \alpha$ respectively). Additionally, the Loewenstein and Prelec (1992) generalized hyperbolic discounting model reduces to exponential discounting when α approaches 0. Note that Loewenstein and Prelec's hyperbolic model can also be viewed as an exponential discounting model with non-linear time perception (Takahashi, 2005). To illustrate, let $\beta = -\log(\delta) \geq 0$ and assume that subjective perception of time follows a decreasingly elastic function $w(t) = \frac{1}{\alpha}\log(1 + \alpha t)$, then the exponential model with subjective perception of time becomes

$\delta^{w(t)} = e^{-\beta w(t)} = e^{-\beta \cdot \frac{1}{\alpha}\log(1+\alpha t)} = \frac{1}{(1+\alpha t)^{\beta/\alpha}}$, which turns out to be the Loewenstein and Prelec hyperbolic model.

Another popular member in the hyperbolic discounting family is the quasi-hyperbolic discounting model (Laibson, 1997; Phelps & Pollak, 1968):

$$d(t) = \begin{cases} \delta^t = 1, & \text{when } t = 0 \\ \beta\delta^t, & \text{when } t > 0 \end{cases} \tag{6}$$

This model does not follow the conventional hyperbolic discounting model in Eq. (3). Instead, it is a direct generalization of the exponential discounting model, with a present bias parameter β, with $0 \le \beta \le 1$. When an outcome is delayed, the discount function is $d(t) = \beta\delta^t$. In contrast, the discount weight for an immediately available outcome is $d(0) = \delta^0 = 1$. The quasi-hyperbolic discounting model is able to predict decreasing impatience when one of the options involves an immediate payoff, but behaves like the exponential model when evaluating non-immediate payoffs. Additionally, the quasi-hyperbolic discounting model reduces to exponential discounting when $\beta = 1$.

3. Probabilistic specifications

One critical issue with any discounted utility model, regardless of whether the discounting is exponential or hyperbolic, is its inability to account for apparent randomness in human choice behavior. Discounted utility models assume that decision makers deterministically select the choice option with the highest discounted utility. However, decision makers sometimes make different choices even when presented with identical choice options (see Luce & Suppes, 1965, for an early discussion; Rieskamp, Busemeyer, & Mellers, 2006, provide a more recent review of key issues). In order to use discounting models to describe stochastic choice data, these models need to be recast in probabilistic terms.

3.1. Stochastic choice

The typical approach to applying discounted utility models to stochastic choice data is to assume that there is some stochasticity in translating utilities into choice – that is, some stochasticity in response generation. This is most frequently modeled with a logit specification (McFadden, 1974; also see Luce, 1959; Thurstone, 1959 for related techniques). For options X and Y, the probability of selecting X over Y is given by:

$$p[X;\, Y] = \frac{1}{1 + \exp\{-\theta(U(X) - U(Y))\}} \tag{7}$$

Here $\theta \ge 0$ is a parameter that determines the responsiveness of choice to utility differences. Smaller values of θ correspond to noisier choices. As $\theta \to \infty$ we obtain deterministic choices of the option with the highest utility. The utility function $U(\cdot)$ could be defined by either the exponential or the hyperbolic models. Intertemporal choice models defined as in Eq. (7) are a special case of the standard logit model that is widely used in economic analysis (Cheremukhin, Popova, & Tutino, 2015; Swait & Marley, 2013).

3.2. Preference variability

There is also, however, another potential source of randomness in choice: preference variability. The parameters of utility-based models often provide a formal representation of decision makers' preferences. These preferences may not be constant over the time course of an experiment; that is, they may themselves fluctuate in a random manner (Becker et al., 1963; Loomes & Sugden, 1995; also see Gul & Pesendorfer, 2006; Loomes, 2005; Loomes, Moffatt, & Sugden, 2002; Regenwetter & Marley, 2001; Regenwetter et al., 2011). Within an exponential discounting model, this type of variability would correspond to a distribution of discount factors described by a probability density function $f(\delta)$. The discount factor δ varies from trial to trial, according to $f(\cdot)$, causing the discounted utilities to vary from trial to trial. In a given trial, the option with the higher utility contingent on the sampled δ would be chosen. In a choice between options X and Y, the probability of choosing X is given by:

$$p[X;\, Y] = \int g(X,\, Y|\delta) f(\delta) d\delta$$

$$\text{with } g(X,\, Y|\delta) = \begin{cases} 1 & \text{if } U(X) > U(Y) \\ 0.5 & \text{if } U(X) = U(Y) \\ 0 & \text{if } U(X) < U(Y) \end{cases} \tag{8}$$

where $f(\delta) = 0$ when $\delta \notin [0, 1]$. Here, $g(X,\, Y|\delta)$ is a deterministic choice function (generating probabilistic random choices only when the two utilities are equal), and $U(\cdot)$ is the discounted utility function defined in Eq. (1) with the exponential discount function $d(t) = \delta^t$ as in Eq. (2).

3.3. Combined specification

Of course both preference variability and stochastic choice can influence intertemporal choice simultaneously. In this setting we would have both variability in discounting parameters for generating utilities and stochasticity in translating utilities into choice.

Choice probabilities with the variable exponential model can be obtained by integrating $p[X; Y]$ as defined in Eq. (7), over the range of feasible values of δ, weighted by their respective probabilities. Thus, in a choice between option X and option Y, the probability of choosing X would be given by:

$$p[X; Y] = \int g(X, Y|\delta) f(\delta) d\delta$$
$$\text{with } g(X, Y|\delta) = \frac{1}{1 + \exp\{-\theta(U(X) - U(Y))\}} \tag{9}$$

Note that the variable exponential model as defined in Eq. (9) is a mixed logit model for intertemporal choice (and therefore also a random utility model), as it embeds both stochastic choice in the form of a logit choice function and stochastic "tastes" of impatience in the form of variable discount factors δ (McFadden & Train, 2000).

In our model fits, we will constrain the variable exponential model by assuming that $f(\delta)$ is symmetrically uniformly distributed around $\mathrm{E}[\delta] = \bar{\delta}$ with a radius of $\eta \geq 0$, i.e. $\delta \sim \text{Uniform}[\bar{\delta} - \eta, \bar{\delta} + \eta]$, where $0 \leq \bar{\delta} - \eta \leq \bar{\delta} + \eta \leq 1$. With this specification $\eta = 0$ generates an exponential model with only stochastic choice (Eq. (7)), whereas $\theta \to \infty$ approximates an exponential model with only preference variability (Eq. (8)). Our analytical results in Section 4 do not require any particular specification of the distribution of discount factors $f(\delta)$.

Note that we do not permit variability in θ (as considered in Golman (2012)), which would entail variability in the stochastic response generation process, or consider alternative functional forms for the stochastic choice rule $g(\cdot)$. We adopt this logit choice rule specification simply because it is convenient to work with. While parameter variation in the stochastic choice rule (along with some motivation for its functional form) would be cognitively reasonable, we cannot anticipate resultant qualitative changes in choice behavior that would warrant this complication of the model. Allowing for variation in one's time preferences, on the other hand, does yield interesting qualitative behavioral predictions, as we show in Section 4.

3.4. Related approaches

Stochastic choice and preference variability have been individually applied to model choice behavior in many previous papers (e.g. Luce, 1959; McFadden, 1974; Thurstone, 1959 for applications of stochastic choice; Becker et al., 1963; Loomes & Sugden, 1995; Regenwetter & Marley, 2001 for applications of preference variability). The combination of them has, however, only been studied recently. For example, Regenwetter and Robinson (2017), Regenwetter et al. (2018) have examined risky and intertemporal choices in models that permit both stochastic choice and preference variability (in fact, the variable exponential model in Eq. (9) can be seen as a special case of Regenwetter et al. (2018) general intertemporal choice model with "probabilistic preferences compounded with probabilistic response"). Their work suggests that this type of randomness could substantially change the interpretation of experimental data in decision making. In intertemporal choice, their analysis shows that apparent violations of transitivity can be explained by transitive models if both stochastic choice and preference variability are permitted.

In the domain of risky choice, Bhatia and Loomes (2017) apply both stochastic choice and preference variability within an expected utility theory framework, and show that the resulting model can predict many risky choice phenomena, such as choice patterns commonly seen to support Prospect Theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) accounts of risk taking. Expected utility theory is typically considered to be a rational or normative model (Von Neumann & Morgenstern, 1947), and the results of Bhatia and Loomes indicate that such a model, when equipped with stochastic choice and preference variability, is perfectly capable of generating seemingly irrational patterns of choice behavior observed in risky decision making research. Our specification of an exponential discounting model with both stochastic choice and preference variability can be seen as modifying the approach proposed by Bhatia and Loomes (2017) for the intertemporal choice domain.

4. Properties

4.1. Modal choice patterns

Let us begin by examining modal choice patterns generated by the variable exponential model. In a binary choice setting, the modal choice is the option that is chosen with the higher choice probability (i.e., choice probability greater than 0.5). The exponential discounting model with stochastic choice but no preference variability (Eq. (7)) necessarily chooses the most desirable option –that is, the one with the highest discounted utility– with higher probability whenever $\theta > 0$ (when $\theta = 0$ both options are chosen with equal probability). This is an immediate consequence of the manner in which utility differences enter the logit function: Desirable options always have larger utilities, leading to choice probabilities greater than 0.5. Of course, as choice is not completely deterministic, undesirable options can be selected from time to time. Mistakes of this form happen in inverse proportion to the utility differences between the available options.

We also obtain a similar result for an exponential discounting model with only preference variability (Eq. (8)). Trial-to-trial variability in δ alters decision makers' utilities when they are exposed to the same decision problems repeatedly, thus leading to randomness in choice. However, when δ is distributed symmetrically around $\mathrm{E}[\delta]$, the option that is favored by $\mathrm{E}[\delta]$ in the simple binary choices examined in this paper, is the one that is chosen more than 50% of the time. The rationale is straightforward. If, for a given value of $\mathrm{E}[\delta]$ between 0 and 1, the decision maker prefers $X = (x; t)$ over $Y = (y; t + k)$, then the decision maker also prefers X over Y for all $0 < \delta < \mathrm{E}[\delta]$. Given that utility differences are continuous in δ, there is furthermore, some value of

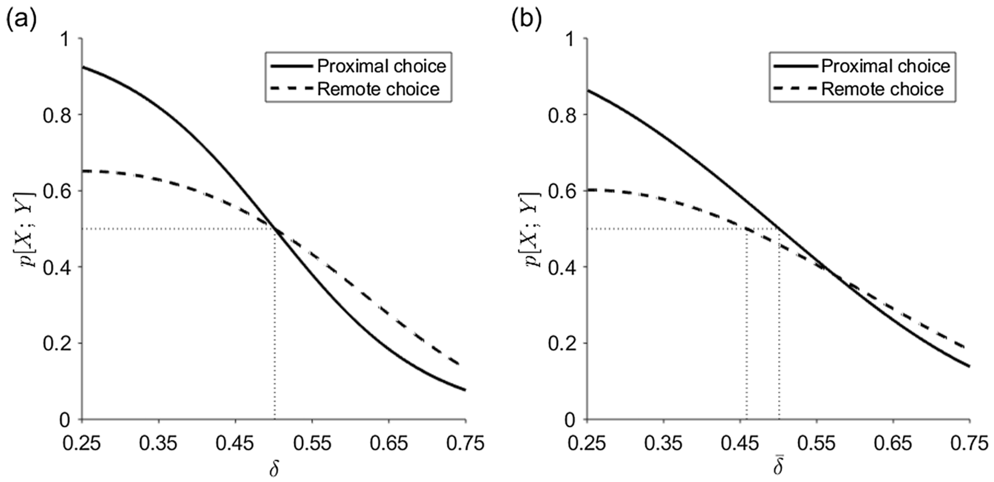

Fig. 1. The probability of X being chosen over Y as a function of (expected) discount factor in proximal and remote choices respectively when (a) only stochastic choice ($\theta = 1$) is assumed and (b) both stochastic choice ($\theta = 1$) and preference variability ($\eta = 0.25$) are assumed.

$1 > \delta > E[\delta]$ for which X is also preferred over Y. As δ is distributed symmetrically around $E[\delta]$, this implies that the eventual probability of choosing X over Y is greater than 50%.

The modal choice probabilities generated by an exponential discounting model with both stochastic choice and preference variability (Eq. (9)) can, however, deviate from the predictions of the corresponding deterministic exponential model with discount factor of $E[\delta]$ (and thus also deviate from the predictions of the associated stochastic-choice-only and preference-variability-only models). Consider the proximal and remote choices in Section 2.2. If we only allow for stochastic choice, the modal choice for both choice problems is X when $\delta < 0.5$ and Y when $\delta > 0.5$ (see the solid and dashed lines in Fig. 1a). However if we allow for both stochastic choice ($\theta = 1$) and preference variability ($\eta = 0.25$), the modal choice for the remote choice problem is Y, rather than X, when $0.46 < \bar{\delta} < 0.5$ (see the dashed line in Fig. 1b) (note that the assumption of preference variability does not change the modal choice for the proximal choice problem (see the solid lines in Fig. 1a and b)). The reason for this is that the utility difference between X and Y, $U(X) - U(Y)$ is non-linear in δ. This means that variability in δ distorts the expected differences in utility between the two options, so that the expectation of the utility difference between X and Y, $E[U(X) - U(Y)]$, is not the same as the utility difference of these options, $U(X) - U(Y)$ for $E[\delta]$. When preference variability is applied by itself (as in Eq. (8)), this distortion does not alter modal choice, as the choice rule is based only on whether $U(X) > U(Y)$ or $U(X) < U(Y)$ and not on the magnitude of $U(X) - U(Y)$. However, when these utility differences are combined with stochastic choice (as in Eq. (9)) the distorted utility differences lead to distorted choice probabilities. Note that this can happen even if the distribution of δ is symmetric around $E[\delta]$.

The effect of such distortion on choice probability is formalized in Proposition 1. Proposition 1 shows that for any underlying exponential discounting model, with non-degenerate stochastic choice, there exists some choice pair, and some amount of symmetric preference variability, so that the modal choice in the presence of preference variability deviates from the modal choice in the absence of preference variability (i.e. the choice made by the corresponding deterministic exponential discounting model, as well as the associated stochastic-choice-only and preference-variability-only models).

Proposition 1. *For any* $\bar{\delta} \in (0, 1)$ *and* $\theta > 0$ *in the variable exponential model* (*Eq.* (9)), *there exists choice pair* (X, Y), *with* $X = (x; t)$ *and* $Y = (y; t + k)$, *where* $0 < x < y$, $t \geq 0$ *and* $k > 1$, *such that*

1. $p[X; Y] > 50\%$ *if* $f(\delta) = \begin{cases} 1, & \text{if } \delta = \bar{\delta} \\ 0, & \text{if } \delta \neq \bar{\delta} \end{cases}$ (*i.e. in the absence of preference variability*), *but*
2. $p[X; Y] < 50\%$ *if* $f(\delta)$ *is non-degenerate and symmetric around* $\bar{\delta}$ (*i.e. in the presence of preference variability*).

Proof. See Appendix A.

The primary consequence of Proposition 1 is that regardless of the underlying exponential discounting model, as long as there is stochastic choice, there is also guaranteed to be some choice pair for which the modal choice in the absence of preference variability diverges from the modal choice in the presence of preference variability. It is impossible to find such a choice pair without having stochastic choice. Proposition 1 also shows that the combination of preference variability and stochastic choice can make a person appear to be more patient than his or her expected discount factor would suggest.

4.2. Decreasing impatience

One implication of this result is that modal choice patterns that appear to contradict exponential discounting may in fact be compatible with an exponential discounting model that has both preference variability and stochastic choice. This is the case for choice patterns corresponding to decreasing impatience. As an illustration of this, consider again the proximal and remote choices in the decreasing impatience example in Section 2.2. If we only allowed for stochastic choice, and set $\theta = 1$ in Eq. (7), we would obtain $p[X_P; Y_P] > 0.5$ in the proximal choice and $p[X_R; Y_R] > 0.5$ in the remote choice for all values of $\delta < 0.5$, and $p[X_P; Y_P] < 0.5$ in the proximal choice and $p[X_R; Y_R] < 0.5$ in the remote choice for all values of $\delta > 0.5$. This is shown in Fig. 1a.

Now consider adding preference variability to this formulation, with $\delta \sim \text{Uniform}[\bar{\delta} - 0.25, \bar{\delta} + 0.25]$. Note that $\text{E}[\delta] = \bar{\delta}$. In this setting, we find that $p[X_P; Y_P] > 0.5$ in the proximal choice and $p[X_R; Y_R] > 0.5$ in the remote choice for $\bar{\delta} < 0.46$, and $p[X_P; Y_P] < 0.5$ in the proximal choice and $p[X_R; Y_R] < 0.5$ in the remote choice for $\bar{\delta} > 0.50$. However, for $\bar{\delta}$ in the range (0.46, 0.50) we obtain both $p[X_P; Y_P] > 0.5$ and $p[X_R; Y_R] < 0.5$, consistent with the finding of decreasing impatience. This is shown in Fig. 1b.

In Fig. 1 we show an example of decreasing impatience with a single pair of choice problems and a pre-specified set of parameters. Do the patterns observed in Fig. 1 hold more generally? In order to test this, we conducted a computational analysis using parameter space partitioning (Pitt, Kim, Navarro, & Myung, 2006). This analysis implemented an exhaustive search through both the space of free parameters and the space of stimuli to test whether the variable exponential model generally displays decreasing impatience over time (see Appendix B for implementation details). The stimuli used in this test involved pairs of choice problems. One was a choice between a sooner payoff $X_P = (x; t)$ and a later payoff $Y_P = (y; s)$ and the other was a choice between a sooner payoff $X_R = (x; t + l)$ and a later payoff $Y_R = (y; s + l)$, where $y > x > 0$, $s > t \geq 0$ and $l > 0$. A pattern of $p[X_P; Y_P] \geq 0.5$ and $p[X_R; Y_R] < 0.5$ would indicate decreasing impatience, whereas a pattern of $p[X_P; Y_P] < 0.5$ and $p[X_R; Y_R] \geq 0.5$ would indicate increasing impatience. If the modal choices in the two choice problems are the same, the test does not provide strong evidence of either decreasing or increasing impatience. The computational analysis shows that the variable exponential model displayed decreasing impatience in 0.73% of the tests. Further partitioning of the parameter space indicates that the decreasing-impatience predictions were spread all over the parameter space (see Fig. A1 in Appendix B). In contrast, the model never displayed increasing impatience. This suggests that the variable exponential model is able to accommodate decreasing impatience, but it is rare, if not impossible, for it to accommodate the reverse.

Note the vast majority of space yielded inconclusive patterns: i.e. consistent modal choices for the two choice problems in each pair. This is due to the fact that our search involved a high-dimensional space of parameters and stimuli, most of which yield strong preferences for either the sooner or the later option (regardless of the time delay to the sooner option).

4.3. Dominance

An option is said to dominate another option if the former is better than the latter along at least one dimension and is at least as good as the latter along all other dimensions involved. In the context of intertemporal choice, a larger, sooner option dominates a smaller, later option. Stochastic choice, as conventionally applied to exponential discounting or other utility models (i.e., in the absence of preference variability), overlooks the structure of dominance in the choice set and thus over-predicts dominance violations. To illustrate, let us consider a choice between option X = (\$10; 0 months) and option Y = (\$15; 1 month) (non-dominance choice), as well as another choice between option X, and option X_D = (\$7.5; 0 months) (dominance choice). For a decision maker with $\delta = 0.5$ per month, we have $U(X) - U(Y) = U(X) - U(X_D) = 2.50$. As the difference in utilities is the same between X and Y and between X and X_D, an exponential choice model with only stochastic choice (as in Eq. (7)) would predict the same choice probability of X in both cases. In other words, the decision maker would be equally likely to make a mistake and select the less desirable option in the non-dominance choice as in the dominance choice. In reality, decision makers can detect dominance. Although they do occasionally choose dominated options, the likelihood of doing so is much lower than that typically predicted by models equipped with only stochastic choice. This has been established in risky choice (e.g. Busemeyer & Townsend, 1993; Diederich & Busemeyer, 1999; Loomes & Sugden, 1998), and, in our experiments below, we find that this also holds in intertemporal choice.

A combination of stochastic choice and preference variability, on the other hand, makes more reasonable predictions of intertemporal dominance violations. In order to provide an adequate account of choices involving intertemporal dominance, a model could include preference variability along with just a small amount of stochastic choice. For example, if we allow for $\theta = 1$, as well as $\delta \sim \text{Uniform}[\bar{\delta} - \eta, \bar{\delta} + \eta]$, with $\bar{\delta} = 0.5$ and $\eta = 0.25$, we obtain $p[X; Y] = 80.0\%$ in the non-dominance choice, but $p[X; X_D] = 92.4\%$ in the dominance choice. Thus even though the difference in utilities between X and Y and between X and X_D is the same under $\text{E}[\delta] = \bar{\delta}$, the probability of choosing X is higher when it dominates its competitor.

The intuition for this choice pattern is straightforward: Stochastic choice generates mistakes based on the utility differences between options, implying that dominance is violated too frequently when stochastic choice is applied by itself, although its predicted dominance violation rates will not exceed 0.5. Preference variability, in contrast, never violates dominance: The utility for the dominating option is always greater than that for the dominated option, regardless of the underlying value of δ, leading to a choice probability of zero for the dominated option in the absence of stochastic choice. The combination of stochastic choice and preference variability results in a prediction in between these two extreme predictions. Thus, in a model with both stochastic choice and preference variability, it is possible to choose a dominated option, but the probability of this is smaller than the probability of choosing an equally desirable non-dominated option.

We can see this in more detail in Fig. 2a and b, which contrast the probabilities of X being chosen in the non-dominance and dominance choices respectively, given the variable exponential model with $\bar{\delta} = 0.5$ fixed but different levels of choice stochasticity (θ) and preference variability (η). The predicted probability of X being chosen is much higher in the dominance choice than in the non-dominance choice for larger values of η, that is, whenever there is sufficient preference variability. Additionally, increasing choice stochasticity (by reducing θ) decreases the probability of X, the more desirable option, being selected in both the dominance and non-dominance choices. In contrast, increasing preference variability (by increasing η) only has this effect in the non-dominance

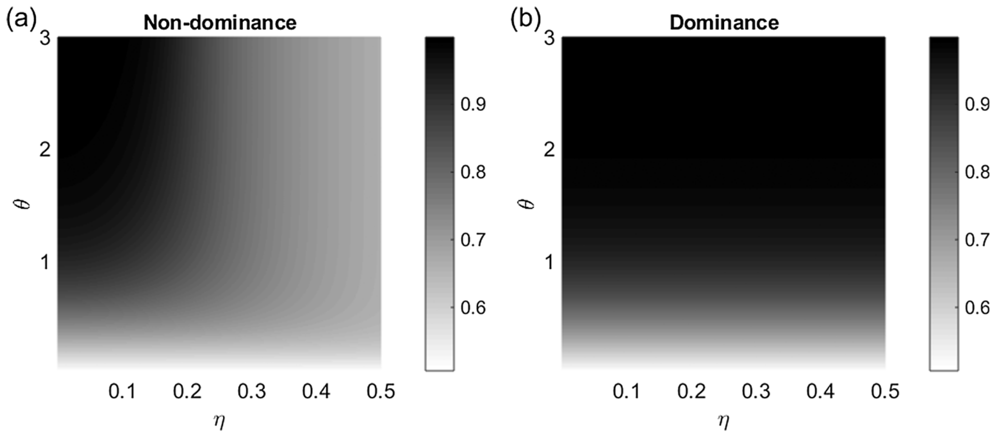

Fig. 2. The probability of X being chosen (with darker shades indicating higher probabilities) as a function of preference variability (η) and stochastic choice (θ) when (a) the alternative option Y is not dominated by X and (b) the alternative option X_D is dominated by X.

choice.

In Proposition 2 we formalize the implications of stochastic choice and preference variability for intertemporal dominance. Proposition 2 shows that for any exponential discounting model and any set of choice options involving dominance, there exist stochastic choice and preference variability parameters for which the probability of selecting a dominated option is less than the probability of selecting an equally less desirable option in the absence of dominance. Proposition 2 also establishes that it is impossible to violate dominance without at least some stochastic choice.

Proposition 2. *Consider choice options* $X = (x;\, t)$, $X_D = (x_D;\, t_D)$ *and* $Y = (y;\, s)$, *such that* $x \geq x_D$ *and* $t \leq t_D$ *(with at least one of these inequalities strict), but either* $x < y$ *or* $t > s$ *(not both), and suppose that in a choice between* X_D *and* Y, *we have* $p[Y;\, X_D] = p\,[X_D;\, Y\,] = 0.5$.

1. *If the distribution* $f\,(\delta)$ *in the variable exponential model (Eq. (9)) is degenerate (i.e., in the absence of preference variability), then in distinct binary choices,* $p[X;\, X_D] = p\,[X;\, Y]$.
2. *If the distribution* $f\,(\delta)$ *has full support on* $[0,\, 1]$, *then*

$$\lim_{\theta\to\infty} p\,[X_D;\, X] = 0,$$

but

$$\lim_{\theta\to\infty} p\,[Y;\, X] > 0,$$

i.e., for large enough θ, *we have* $p\,[Y;\, X] > p\,[X_D;\, X]$.

Proof. See Appendix A.

4.4. *Strong stochastic transitivity*

Although it does seem reasonable for decision makers to be able to detect and avoid dominance violations more frequently than they are able to reject non-dominated choices with similar desirability, such choice tendencies violate a key decision theoretic axiom: *Strong Stochastic Transitivity* (SST). SST states that for any options A, B, and C, if $p[A;\, B] \geq 0.5$ and $p[B;\, C] \geq 0.5$, then we must have $p\,[A;\, C] \geq \max\{p[A;\, B],\, p[B;\, C]\}$. SST violations are likely when dominance is involved. The intuition is as follows: Because dominance violations are relatively unlikely, we obtain $p[X;\, X_D] > p[X;\, Y]$ when X_D and Y equally attractive in pairwise choice. When it is the case that $p[X_D;\, Y] = 0.5$, we can usually obtain:

$$p\,[X;\, Y] < \max\{p\,[X;\, X_D],\, p\,[X_D;\, Y]\} = p\,[X;\, X_D] \tag{10}$$

Note that $p[X_D;\, Y] = 0.5$ is not a necessary condition for Eq. (10) to hold. For choice options Y' that are slightly worse than Y, we would have $p[X_D;\, Y'] > 0.5$, but we would still find $p[X;\, Y'] < p[X;\, X_D]$ because choice probabilities depend continuously on an option's payoff and time delay. We construct SST violations with choice options satisfying $p[X_D;\, Y] = 0.5$ simply for expositional convenience.

In Proposition 3 we show that such violations are guaranteed in the variable exponential model with any amount of positive symmetric preference variability; that is, for any such model, it is possible to find choice options X, X_D and Y, such that SST is violated. Note that models with only stochastic choice, in the manner formalized in Eq. (7), can never violate strong stochastic transitivity (Blavatskyy, 2014; Luce & Suppes, 1965; see Loomes & Sugden, 1998; Marley & Regenwetter, 2017, chap. 7 for an extensive discussion in the context of risky choice).

Proposition 3. *Given any non-degenerate distribution $f(\delta)$ and $\theta > 0$ in the variable exponential model (Eq. (9)), there exist options $X = (x; t)$, $X_D = (x_D; t)$ (where $x_D < x$), and $Y = (y; s)$ (where $y > x$ and $s > t$) such that:*

1. $p[Y; X] \geq 0.5$ *and* $p[X; X_D] \geq 0.5$, *but* $p[Y; X_D] < \max\{p[Y; X], p[X; X_D]\}$ *or*
2. $p[X; X_D] \geq 0.5$ *and* $p[X_D; Y] \geq 0.5$, *but* $p[X; Y] < \max\{p[X; X_D], p[X_D; Y]\}$.

Proof. See Appendix A.

4.5. Summary and discussion

In this section we have outlined the properties of an exponential discounting model with both stochastic choice and preference variability. The combination of the two forms of randomness can distort choice probabilities, and lead to different modal choices than the corresponding deterministic exponential discounting model associated stochastic-choice-only or preference-variability-only models. This allows the exponential discounting model with both stochastic choice and preference variability to explain patterns of decreasing impatience, which are typically attributed to hyperbolic discounting. The model studied in this section also makes reasonable predictions in choices involving dominated options (which correspond to violations of strong stochastic transitivity). The results of our analysis suggest that it may not be necessary to assume complex forms of intertemporal discounting if we permit variability in the discount factors that characterize preferences in an exponential discounting model.

5. Experiments

In this section, we evaluate the quantitative fit of the proposed model. For this purpose, we analyze the results of two experiments, contrasting an exponential discounting model with both stochastic choice (in the form of a logit choice function) and preference variability (in the form of uniformly distributed discount factors), with various hyperbolic models and the exponential model equipped with only stochastic choice. We also evaluate the fit of a quasi-hyperbolic model equipped with both stochastic choice and preference variability, which we call the *variable quasi-hyperbolic model*.

The results of our tests will show whether an exponential discounting model with variable discount factors is able to provide a reasonably good account of data, and how well it performs against other alternatives to the standard exponential discounting model. These experiments will also examine dominance choices. This will help us further evaluate the unique properties of variable discount factors in an exponential discounting model of intertemporal choice.

5.1. Experiments 1 and 2

5.1.1. Methods and materials

A total of 89 undergraduate students from a university in the United States participated in the first two experiments: 44 participants (31 female; aged 20.26 ± 1.25) in Experiment 1 and 45 participants (25 female; aged 19.84 ± 1.49) in Experiment 2.

Experiment 1 involved hypothetical binary choices between an option $X = (x; t)$ offering a payoff of x after a time delay t, and an option $Y = (y; t + k)$ offering a payoff of y after a time delay $t + k$. We set $x = \$100$ in all trials and chose t from the set {0 months (immediately), 3 months, 6 months, 9 months} and k from the set {3 months, 6 months, 9 months}. We kept the length of the intervals between the delayed payoffs systematically varied to control for subadditive discounting, a well-known phenomenon that people tend to discount more steeply over short intervals than over long intervals (Kinari, Ohtake, & Tsutsui, 2009; McAlvanah, 2010; Read & Roelofsma, 2003; Read, 2001). Read (2001) has argued that much of the early evidence for decreasing impatience (and thus hyperbolic discounting models) was confounded with subadditive discounting. y was determined by applying an annual interest rate from the set {−50%, 50%, 100%, 500%, 1000%} to the corresponding time delays. This generated a total of 60 unique choice pairs. The relatively large size of the stimulus set also entailed highly powered tests of decreasing impatience, relative to some earlier studies that involved just one or very few choice pairs. Note that the use of a negative interest rate implied that 12 of these choice pairs involved a dominated option (offering a smaller reward with a larger time delay than its competitor).

As shown below, Experiment 1 involved fairly high choice probabilities for the delayed option Y (in non-dominance choices). Although this should not alter our key conclusions, we wished to replicate our tests with stimuli generating roughly equivalent choice probabilities for X and Y. Thus we ran a second experiment, with stimuli generated using the methods above, but with annual interest rates in the set {-25%, 25%, 75%, 125%, 175%}. By using smaller interest rates we obtained smaller values of y for corresponding values of t and k, leading to higher choice proportions for X. Again the use of a negative interest rate implied that 12 of the 60 choice pairs involved a dominated option. The details of the choice tasks in both experiments can be found in Appendix C.

Both experiments were run in group sessions with eight or nine participants per session. Participants were individually seated in a cubicle, and performed the experiments on a computer. Prior to the start of the experiments, participants were asked to complete five practice trials to familiarize themselves with the computer interface. They then completed 240 trials, consisting of four repetitions of each of the 60 unique choice pairs in the experiment. The placement of options X and Y (left or right) was counterbalanced, and the ordering of the trials was randomized. Participants indicated their choices through keyboard presses.

We excluded data from five participants in Experiment 1 and three participants in Experiment 2 because they chose either X or Y in all the non-dominance choices. This left 39 participants in Experiment 1 and 42 participants in Experiment 2 for our analysis.

5.1.2. Models

The main goal of the two experiments was to test whether the variable exponential model with both stochastic choice and preference variability is able to provide a good quantitative account of choice data. For this purpose, we fit the core exponential discounting model embedded in a logit choice rule for stochastic choice, with a variable discount factor $\delta \sim \text{Uniform}[\bar{\delta} - \eta, \bar{\delta} + \eta]$ with $0 \le \bar{\delta} - \eta \le \bar{\delta} + \eta \le 1$ (Eq. (9)). This model involves three free parameters: $\bar{\delta}$, η and θ.

We also wished to contrast the variable exponential model with the various hyperbolic models proposed in prior work. We considered the one parameter hyperbolic discounting model (Eqs. (1) and (3)) proposed by Mazur (1987), which we refer to as ***Mazur*-1 *hyperbolic***, as well as the two-parameter hyperbolic model (Eqs. (1) and (4)) proposed by Mazur (1987), which we refer to as ***Mazur*-2 *hyperbolic***. We also used the two-parameter generalized hyperbolic discounting model (Eqs. (1) and (5)) proposed by Loewenstein and Prelec (1992), which we refer to as the ***LP hyperbolic***. Finally, we considered the quasi-hyperbolic model (Eqs. (1) and (6)) proposed by Laibson (1997). All hyperbolic discounting models were embedded within the logit choice function (Eq. (7)) to allow for stochastic choice. In this specification, the Mazur-1 hyperbolic model has two free parameters (α and θ), the Mazur-2 hyperbolic model has three free parameters (α, τ and θ), the LP hyperbolic model has three free parameters (α, β and θ), and the quasi-hyperbolic model also has three free parameters (β, δ and θ).

We added another two models to the model comparison tournaments to test whether the variable exponential model can adequately accommodate intertemporal choice data. One was the standard exponential model with only stochastic choice. This model was simply the exponential discount function (Eqs. (1) and (2)) embedded in the logit choice rule (Eq. (7)). This model has two free parameters (δ and θ). As the variable exponential model nested the standard exponential model as a special case, the comparison between the two would tell whether the additional preference variability in the variable exponential model paid off given its cost in model parsimony. The other was the variable quasi-hyperbolic model. As is obvious from its name, this model is the quasi-hyperbolic model equipped with both stochastic choice and preference variability, with a variable discount factor $\delta \sim \text{Uniform}[\bar{\delta} - \eta_\delta, \bar{\delta} + \eta_\delta]$ ($0 \le \bar{\delta} - \eta_\delta \le \bar{\delta} + \eta_\delta \le 1$) and a variable present bias $\beta \sim \text{Uniform}[\bar{\beta} - \eta_\beta, \bar{\beta} + \eta_\beta]$ ($0 \le \bar{\beta} - \eta_\beta \le \bar{\beta} + \eta_\beta \le 1$). It has five free parameters ($\bar{\beta}$, $\bar{\delta}$, η_β, η_δ and θ, where $0 \le \bar{\beta} - \eta_\beta \le \bar{\beta} + \eta_\beta \le 1$ and $0 \le \bar{\delta} - \eta_\delta \le \bar{\delta} + \eta_\delta \le 1$). This model nests the variable exponential model as a special case and has additional present-biased specifications in β (as is governed by $\bar{\beta}$ and η_β). Thus the comparison between the variable exponential and the variable quasi-hyperbolic models tests whether the variable exponential model can adequately accommodate the decreasing impatience effect in the data without additional present-biased specifications.

5.1.3. Bayesian model estimation and comparisons

Model fitting requires likelihood functions from the models. Although there are simple analytical representations for the likelihood functions of discounting models with just stochastic choice, we needed to computationally approximate the likelihood function for the variable exponential and the variable quasi-hyperbolic models. To this end, we used computational simulation. For each set of candidate values of $\bar{\delta}$ and η in the variable exponential model, we selected 1000 evenly-spaced discount factors between $\bar{\delta} - \eta$ and $\bar{\delta} + \eta$ and calculated the probability of Y being chosen for each of these discount factors. The average of these estimates yielded the probability of Y being chosen given a variable discount factor $\delta \sim \text{Uniform}[\bar{\delta} - \eta, \bar{\delta} + \eta]$. Likewise, for the variable quasi-hyperbolic, we selected 1000 evenly-spaced discount factors between $\bar{\delta} - \eta_\delta$ and $\bar{\delta} + \eta_\delta$ and 1000 evenly-spaced present biases between $\bar{\beta} - \eta_\beta$ and $\bar{\beta} + \eta_\beta$. We then randomly paired the discount factors and the present biases and estimated the probability of Y being chosen for each of the 1000 pairs. The average of these estimates yielded the probability of Y being chosen.

We used the Bayesian approach for model estimation and comparisons. The Bayesian approach allowed us to make a coherent statement about the direction and magnitude of evidence for each model from each individual participant while model complexity is automatically penalized (Gelman et al., 2014; Kruschke, 2015). We used maximum a posteriori (MAP) for parameter estimation. MAP is a Bayesian equivalent to the frequentist maximum likelihood estimate (MLE). It refers to the parameters that obtain the highest posterior probability (i.e. considering both the prior probability and the likelihood value) over the parameter space for a given model. Note that the parameter posterior probability for a given model is different from the model posterior probability in model comparisons, which we will introduce shortly below.

To compare multiple models simultaneously, we used the marginal likelihood, $p(D|M) = \int p(D|M, \Theta)p(\Theta|M)d\Theta$, where Θ is the free parameters for a model, as the model selection criterion. With marginal likelihoods, the Bayes factor (*BF*) between two models can be directly estimated as $BF_{12} = \frac{p(D|M_1)}{p(D|M_2)}$. Assuming equal prior probabilities for all models, we can also calculate model posterior probability as $p(M_i) = \frac{p(D|M_i)}{\sum_{j=1}^{N} p(D|M_j)}$, where N is the total number of models involved. We applied the simple Monte Carlo approximation to estimate the marginal likelihoods (see e.g. Scholten, Read, & Sanborn, 2016). The simple Monte Carlo approach can obtain stable and accurate approximation when the posterior distributions of the parameters are not too much narrower than the prior distributions. This is often the case when models are fitted to individual choice data, where the evidence from data would not be overwhelmingly strong and thus the posterior distributions would not be too narrow. Compared with the Markov chain Monte Carlo (MCMC) approximation of Bayesian analysis, the simple Monte Carlo approximation is much less computationally intense.

The simple Monte Carlo approximation of marginal likelihoods consisted of three key steps. First, one million samples of parameters were drawn from the prior distribution of parameters. Specifically, for parameters that are bounded between 0 and 1, we used a uniform distribution between 0 and 1 as the prior distribution. For parameters that are nonnegative but have no upper bound, we used an exponential distribution with the probability density function $p(x; \lambda) = \lambda e^{-\lambda x}$, where $\lambda = 1$. Second, we calculated the likelihood of data for each of the one million samples of parameters, obtaining one million likelihood values. Third, the one million likelihood values were averaged as the marginal likelihood of the data for the given model. To further ensure that the approximation was stable and accurate, we ran the approximation twice for each model and each participant. The two runs identified the same best-performing model 99.4% of the time. In the case where the two runs identified different best-performing models, the marginal likelihoods for the two models were extremely close, with Bayes factors of 1.02 and 0.98 in the two runs respectively. Additionally we calculated the deviation of model posterior probabilities across the two runs as an absolute measure of stability. The average deviation in model posterior probabilities between the two runs was 0.0047 across all participants, models and experiments. Overall, the simple Monte Carlo approximation achieved highly stable results.

5.1.4. Results

We found that option Y (offering the payoff with the larger delay) was chosen 54.3% of the time in Experiment 1, and 39.8% of the time in Experiment 2. Restricting our analysis to only the non-dominance choices, we found that option Y was chosen 66.9% of the time in Experiment 1, and 48.6% of the time in Experiment 2. The frequency of choosing the dominated options was comparable across the two experiments: Participants in Experiment 1 and Experiment 2 chose the dominated option (option Y offering both the smaller and the more delayed reward) 3.8% of the time and 4.7% of the time respectively.

5.1.4.1. Test of decreasing impatience. We tested for decreasing impatience in the two experiments with the non-dominance trials. In Experiment 1, participants chose option Y (offering the payoff with the larger delay) 65.9% when $t = 0$ months, 66.0% when the front-end delay $t = 3$ months, 67.7% when $t = 6$ months, and 67.8% when $t = 9$ months. Correspondingly, they chose option Y 57.5% of the time when $k = 3$ months, 70.0% when $k = 6$ months and 73.0% when $k = 9$ months, a pattern consistent with subadditive discounting. Note that t means the earlier delay in a choice problem. Alternatively it is called the front-end delay of a choice problem and correspondingly k (the difference between the earlier delay and later delay) is called the interval. We formally tested for decreasing impatience with Bayesian hierarchical models using the MCMC algorithm implemented in the "rstanarm" package in R (Goodrich, Gabry, Ali, & Brilleman, 2018; R Core Team, 2018). The convergence of the MCMC simulations was checked with $\hat{R} < 1.1$ (Brooks & Gelman, 1998). When assuming that participants held the same tendency to display decreasing impatience, we found substantial evidence for decreasing impatience in the choice data, with the probability of Y being chosen increasing in t (fixed-effect coefficient: $\beta_{\text{Mean}} = 0.01$, $\beta_{\text{Median}} = 0.01$, $\beta_{95\%\ \text{HDI}} = [0.00, 0.05]$; $\beta_{95\%\ \text{HDI}}$ stands for the 95% high density interval of the estimated posterior distribution). We then ran another Bayesian hierarchical model allowing for differential decreasing-impatience degrees across participants. In this test, five participants displayed decreasing impatience with the 95% high density intervals (95% HDIs) of their respective effects above 0, while no one displayed the reverse, increasing impatience, although the fixed effect of t on the probability of Y being chosen was diluted (fixed-effect coefficient: $\beta_{\text{Mean}} = 0.01$, $\beta_{\text{Median}} = 0.01$, $\beta_{95\%\ \text{HDI}} = [-0.03, 0.05]$). These results provide some evidence of decreasing impatience overall, but also suggest substantial heterogeneity in this choice pattern across participants.

We obtained more ambiguous results regarding decreasing impatience in Experiment 2. In this experiment we observed a choice frequency for option Y of 47.6% when $t = 0$ months, 49.0% when $t = 3$ months, and 48.9% when $t = 6$ months, and 49.1% when $t = 9$ months. Although this choice frequency was increasing in t, we did not find substantial evidence for decreasing impatience in the choice data when assuming that participants held the same tendency to display decreasing impatience (fixed-effect coefficient: $\beta_{\text{Mean}} = 0.02$, $\beta_{\text{Median}} = 0.02$, $\beta_{95\%\ \text{HDI}} = [-0.01, 0.04]$). The Bayesian hierarchical model allowing for differential decreasing-impatience degrees across participants suggested that only two participants displayed decreasing impatience with the 95% HDIs above 0, while no one displayed substantial increasing impatience. Neither did we find evidence for the fixed-effect of the front-end delay t on the probability of Y being chosen (fixed-effect coefficient: $\beta_{\text{Mean}} = 0.02$, $\beta_{\text{Median}} = 0.02$, $\beta_{95\%\ \text{HDI}} = [-0.01, 0.05]$). Conversely we observed substantial evidence consistent with subadditive discounting, with 37.8% of option Y chosen when $k = 3$ months, 50.2% when $k = 6$ months and 58.0% when $k = 9$ months.

Overall, the findings of Experiments 1 and 2 suggest that decreasing impatience is not as robust as is widely held. However, this is consistent with some recent experiments with similar inconclusive effects for decreasing impatience when subadditive discounting is controlled (e.g. Attema, Bleichrodt, Rohde, & Wakker, 2010; Cavagnaro, Aranovich, McClure, Pitt, & Myung, 2016; Kable & Glimcher, 2010; Kinari et al., 2009; Read, 2001; Sayman & Öncüler, 2009; Sopher & Sheth, 2006).

5.1.4.2. Estimation of variable exponential model. One of our key interests is the estimation of preference variability in the variable exponential model. In Fig. 3a and b we show the scatter plots of the maximum a posteriori (MAP) estimation of $\bar{\delta}$ and η for the variable exponential model, across the participants in the two experiments respectively. We observed $\eta > 0$ implying that participants did show preference variability when it was allowed as in the variable exponential model.

As can be seen in the scatter plots, the estimated value of η was negatively correlated with that of $\bar{\delta}$ in both experiments. That is because we placed a constraint over the parameters, $0 \le \bar{\delta} - \eta \le \bar{\delta} + \eta \le 1$, in model fitting. As the estimated $\bar{\delta}$ were all above 0.5, the value of η was more constrained (and thus lower) with $\bar{\delta}$ closer to 1.

5.1.4.3. Model comparisons. We made Bayesian model comparisons between the variable exponential model and other intertemporal choice models. To evaluate and compare the overall model performance across participants, we summed up the marginal log-likelihood (i.e., the log of the marginal likelihood) across participants for each model as the model's aggregate marginal log-likelihood for both Experiments 1 and 2. As can be seen in Table 1, the best performing model according to the aggregate marginal log-likelihoods across participants, for both experiments, was the variable exponential model. Conversely, the worst performing model for both measures for both experiments was the Mazur-1 hyperbolic model. We also examined the proportion of participants best explained by each of the models when all models were compared simultaneously. Here we found that the best performing model

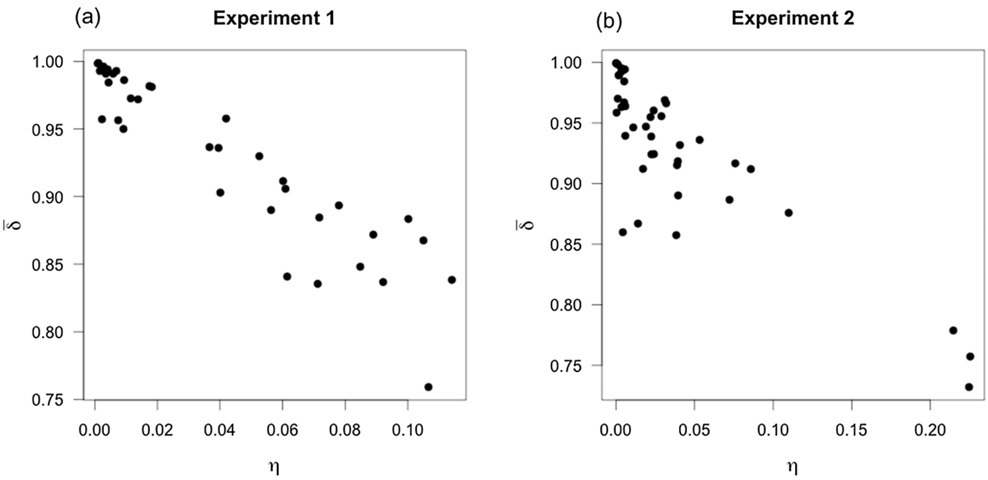

Fig. 3. The maximum a posteriori (MAP) estimates of η and $\bar{\delta}$ from the variable exponential model. (a) Estimates of Experiment 1. (b) Estimates of Experiment 2.

Table 1
Summary of Bayesian model fit statistics in Experiments 1 and 2 (all trials). The two columns correspond to the sum of marginal log-likelihoods across participants, and the proportion of participants best explained by each model.

Model name	Aggr. marginal log-likelihood		% Best marginal log-likelihood	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2
Mazur-1 hyperbolic	−3274	−3746	10%	14%
Mazur-2 hyperbolic	−3085	−3444	0%	10%
LP hyperbolic	−3052	−3428	38%	19%
Quasi-hyperbolic	−3112	−3460	0%	2%
Exponential	−3070	−3401	10%	26%
Variable exponential	−3011	−3391	21%	24%
Variable quasi-hyperbolic	−3042	−3451	21%	5%

was the LP hyperbolic model, which had the highest marginal likelihoods for 28% of the participants in the two experiments (38% in Experiments 1 and 19% in Experiment 2). This was closely followed by the variable exponential model, which provided the highest marginal likelihoods for 22% of the participants in the two experiments (20% in Experiment 1 and 24% in Experiment 2).

To further unpack the direction and the magnitude of the evidence from the model comparisons, we made pairwise contrasts between the variable exponential model and other models at the individual level. Fig. 4a and b shows the pairwise contrasts for Experiments 1 and 2 respectively. The dashed horizontal lines in these figures indicate Bayes factors above 3 or below 1/3, which are considered positive evidence for the better-performing model (values between 1/3 and 3 are considered anecdotal evidence for either model or evidence that is insufficient to tell the difference of model performance) (Kass & Raftery, 1995; Raftery, 1996).

Let us first look at the pairwise contrast between the variable exponential model and the Mazur-1 hyperbolic model at the top left corner of Fig. 4a. Almost all participants offer non-anecdotal evidence for the better-performing model, with 18 participants (46%) favoring the variable exponential model (i.e. $BF > 3$) and 20 participants (52%) favoring the Mazur-1 hyperbolic model (i.e. $BF < 1/3$). Although there were slightly more participants favoring the Mazur-1 hyperbolic model over the variable exponential model than the reverse, however, the magnitude of evidence favoring the variable exponential model was on average much stronger than the magnitude of evidence favoring the Mazur-1 hyperbolic model. We observed similar patterns in the pairwise contrast between the variable exponential model and Mazur-2 hyperbolic, LP hyperbolic and the standard exponential model respectively. In the pairwise comparisons between the variable exponential model and the quasi-hyperbolic model, however, there were more participants better explained by the variable exponential model than by the quasi-hyperbolic model (15 vs. 11 participants or 38% vs. 28%) and the average magnitude of evidence that favored the former was larger than the latter. These results suggest that the variable exponential model outperformed the exponential and hyperbolic models with only stochastic choice mostly because the average magnitude of evidence that favored the former was larger than that of the latter.

The pairwise comparisons between the variable exponential and the variable quasi-hyperbolic models show a different pattern. Although the variable exponential model is nested in the variable quasi-hyperbolic model, a majority of participants (22 out of 39; 56%) in Experiment 1 favored the variable exponential model over the variable quasi-hyperbolic model ($BF > 3$) while only 8 participants (21%) displayed the reverse ($BF < 1/3$). This suggests that the additional present-biased specifications in the variable quasi-hyperbolic model might be unnecessary for the majority of our participants given the fact that variable exponential model with both stochastic choice and preference variability can account for the phenomenon of decreasing impatience. The patterns of

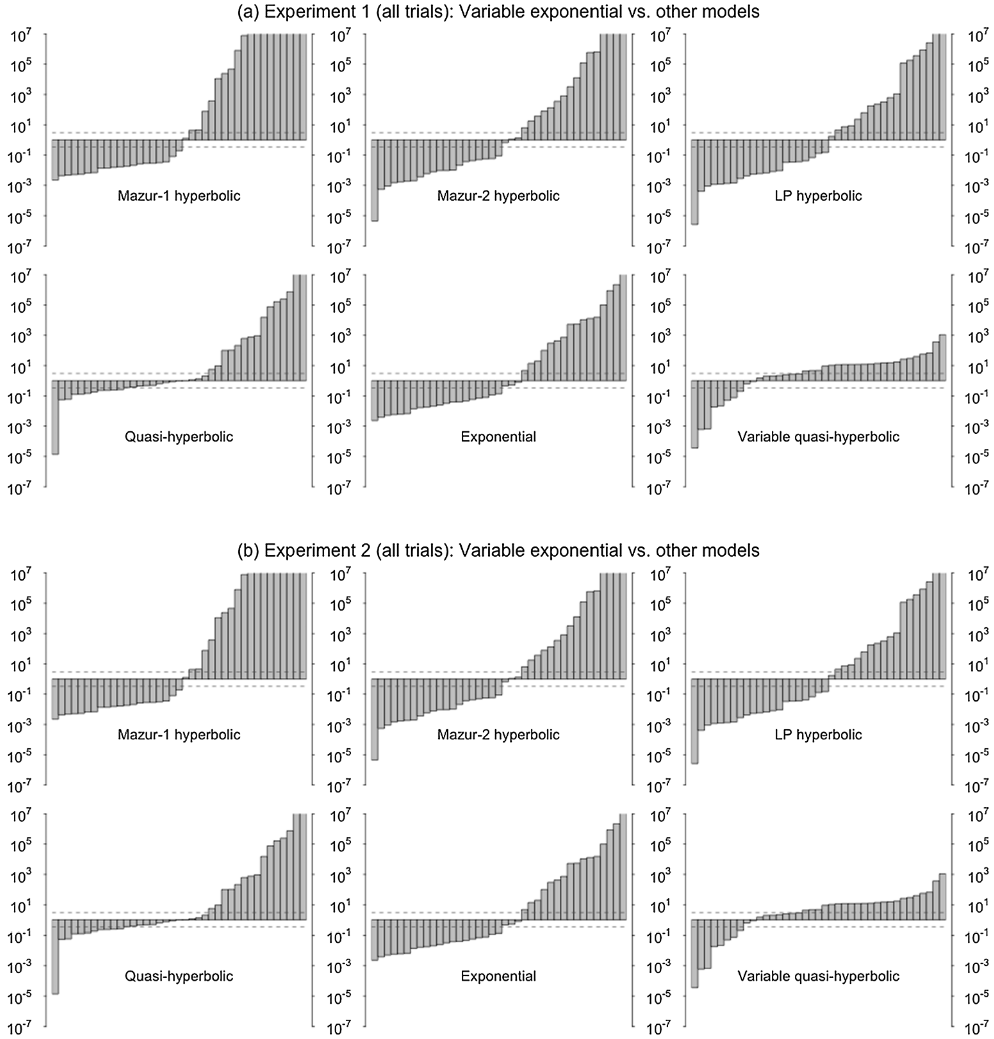

Fig. 4. Individual-level pairwise Bayes factors between the variable exponential model and other models when fitted to all trails in (a) Experiment 1 and (b) Experiment 2. Values are increasingly sorted from small (disfavoring the variable exponential model) to large (favoring the variable exponential model).

Experiment 2 in Fig. 4b show almost the same trends as in Experiment 1.

One potential issue with the model comparisons discussed above is that the advantage of the variable exponential model (or sometimes the variable quasi-hyperbolic model) could be solely due to the fairly large amount of dominance trials involved in the model fitting. This is unlike previous model fitting exercises. To eliminate this possibility, we repeated our fits with only non-dominance trials for both experiments. The results of these fits are presented in Table 2 and Fig. 5. Again, the aggregate marginal log-

Table 2
Summary of Bayesian model fit statistics in Experiments 1 and 2 (non-dominance trials only). The two columns correspond to the sum of marginal log-likelihoods across participants, and the proportion of participants best explained by each model.

Model name	Aggr. marginal log-likelihood		% Best marginal log-likelihood	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2
Mazur-1 hyperbolic	−2877	−3342	5%	14%
Mazur-2 hyperbolic	−2728	−3087	5%	12%
LP hyperbolic	−2699	−3072	41%	19%
Quasi-hyperbolic	−2769	−3104	0%	0%
Exponential	−2724	−3047	10%	14%
Variable exponential	−2684	−3027	21%	31%
Variable quasi-hyperbolic	−2725	−3078	18%	10%

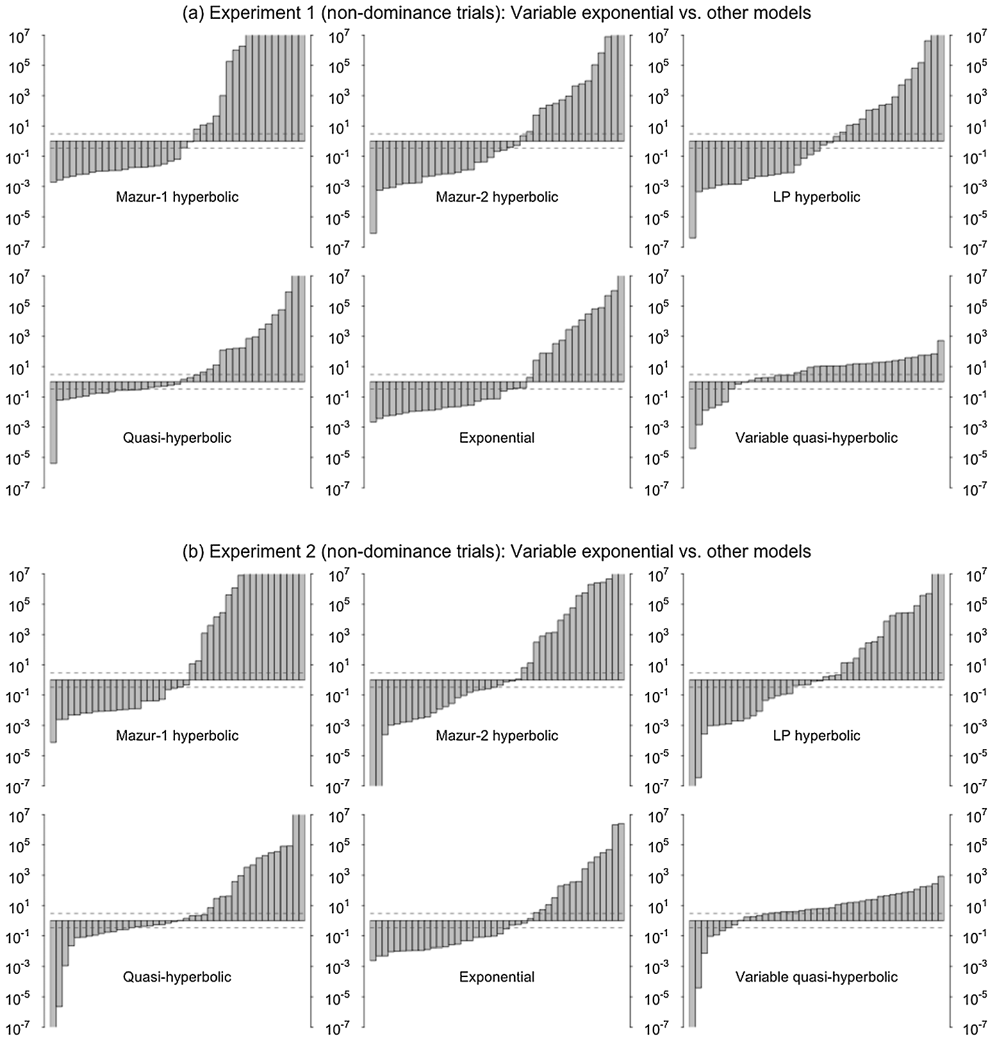

Fig. 5. Individual-level pairwise Bayes factors between the variable exponential model and other models when fitted to non-dominance trails in (a) Experiment 1 and (b) Experiment 2. Values are increasingly sorted from small (disfavoring the variable exponential model) to large (favoring the variable exponential model).

likelihoods suggest that the variable exponential model outperformed all other models in both experiments. The advantage of the variable exponential model was mostly driven by the magnitude of evidence favoring the variable exponential model being larger than the magnitude of evidence favoring the models with only stochastic choice. With the variable exponential model being able to accommodate decreasing impatience, additional present-biased specifications, as in the variable quasi-hyperbolic model, appear unnecessary for the majority of our participants. In summary, these results suggest that the key quantitative advantages of the variable exponential were robust whether or not dominance trials were included in the model fitting.

5.1.4.4. Predictions of dominance violation rates. Our theoretical analysis also shows that the prediction regarding dominance violations is an important property that distinguishes the variable exponential and the variable quasi-hyperbolic models from the models with stochastic choice only. Relative to models with stochastic choice alone, the variable exponential model with both stochastic choice and preference variability leads to fewer mistakes in choices involving dominated options. With the posterior estimation of model parameters from the Bayesian analysis, we were able to recover the posterior predictive choice probability of the dominated option for each model. As shown in Fig. 6a and b, we found that the variable exponential model predicted an average dominance violation rate of 6.1% and 5.5% for Experiments 1 and 2 respectively. Likewise, the variable quasi-hyperbolic model predicted an average dominance violation rate of 5.8% and 5.1% for Experiments 1 and 2 respectively. Although they are higher than the observed dominance violation rate of 3.8% and 4.7%, they are much lower than the dominance violation rates predicted by the models with stochastic choice alone (see the middle column of Table 3 and Fig. 6).

Not only did the variable exponential and variable quasi-hyperbolic models predict the overall dominance violation rates better than the models with stochastic choice alone, they also captured individual differences in dominance violation rates better than

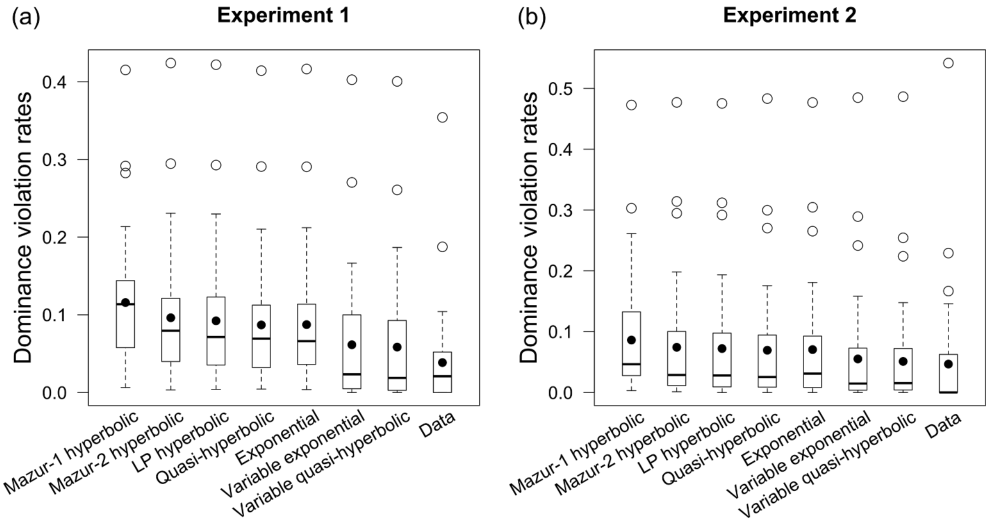

Fig. 6. Boxplots of the dominance violation rates as predicted by models and as observed in data from (a) Experiment 1 and (b) Experiment 2. The filled circles represent (predicted or actual) average dominance violation rates.

Table 3
Models' posterior predictive accuracy of dominance violation rates in Experiments 1 and 2. The middle column contains the average domination violation rates and the right column contains the Pearson's correlation coefficients between model predictions and actual data at the individual level. Values outside the parentheses are predictions when models were fitted to all trials. Values in the parentheses are predictions when models were fitted to non-dominance trials only.

Model name	Average dominance violation rate (%)		Pearson's correlation coefficient	
	Exp. 1	Exp. 2	Exp. 1	Exp. 2
Mazur-1 hyperbolic	11.6 (14.2)	8.6 (10.4)	0.70 (0.54)	0.87 (0.65)
Mazur-2 hyperbolic	9.6 (11.8)	7.4 (8.7)	0.86 (0.71)	0.86 (0.69)
LP hyperbolic	9.2 (11.3)	7.2 (8.4)	0.86 (0.71)	0.87 (0.70)
Quasi-hyperbolic	8.7 (10.5)	6.9 (7.9)	0.87 (0.71)	0.88 (0.72)
Exponential	8.7 (10.7)	7.0 (8.6)	0.87 (0.71)	0.87 (0.65)
Variable exponential	6.1 (7.2)	5.5 (6.5)	0.90 (0.74)	0.92 (0.69)
Variable quasi-hyperbolic	5.8 (6.5)	5.1 (5.3)	0.89 (0.78)	0.94 (0.78)
Data	3.8	4.7	1	1

models with stochastic choice alone. We correlated each model's predicted average dominance violation rate with the actual average dominance violation rate at the individual level. Although all models' predictions correlated with actual data fairly well (with all Pearson's correlation coefficients above 0.7), the two models equipped with both stochastic choice and preference variability constantly obtained the highest correlation coefficients in both experiments, with the coefficients above or close to 0.9 (see the right column of Table 3). Overall we found strong evidence that models equipped with both stochastic choice and preference variability outperformed models with stochastic choice alone in terms of their predictions of dominance violations.

What about the model predictions of dominance violations when the models were fitted to non-dominance trials only? This corresponds to the generalization criterion in model evaluation, namely how well a model generalizes to new datasets or conditions after being estimated from old datasets or conditions (Busemeyer & Wang, 2000). Put differently, we tested how well the parameters estimated from non-dominance trials generalized to dominance trials for each of the models considered. The generalization results are also summarized in Table 3 (i.e. values in the parentheses). Overall, the generalization criterion favored the variable exponential and variable quasi-hyperbolic models, which are equipped with both stochastic choice and preference variability, over the models equipped with stochastic choice alone. The former always provided the average dominance violation rates closer to the observed data and better captured individual differences in dominance violations (via Pearson's correlation) than the latter. The only exception was that, in Experiment 2, the correlation coefficient for the variable exponential model was slightly lower than those for the LP hyperbolic and the quasi-hyperbolic models. However, the variable quasi-hyperbolic model, the other model equipped with both stochastic choice and preference variability, still achieved the highest correlation coefficient. In summary, we found that the advantage of specifying both stochastic choice and preference variability in predicting dominance violations holds even if we do not feed the model fitting with dominance trials.

5.1.5. Summary and discussion

The model fitting exercise in this section has revealed a number of novel insights regarding the effect of preference variability in intertemporal choice. Firstly, preference variability did improve the explanatory power of the exponential model when permitted and explicitly characterized as in the variable exponential model. We also found that the variable exponential model outperformed the various hyperbolic models considered in this paper, in terms of quantitative fit. Our results also suggest that the variable exponential model better accounts for dominance violations than models equipped with stochastic choice alone. Moreover, the advantage of the variable exponential model is not limited to dominance violations, since it outperformed other models in the Bayesian model comparison tournaments even when dominance trials were removed from model fitting. This may be because the variable exponential model quantitatively accommodates the decreasing impatience effect better than the hyperbolic models, even though they all qualitatively predict decreasing impatience. Additional support for this claim comes from the fact that the variable exponential model outperformed the variable quasi-hyperbolic model, which indicates that both preference variability and hyperbolic discounting may be unnecessary (and may lead to over-fitting).

5.2. Experiment 3

Experiments 1 and 2 indicate that the variable exponential model provides strong explanatory power to individual intertemporal choice data. To further test the validity of this model we evaluated its unique predictions regarding violations of strong stochastic transitivity (SST) in the presence of intertemporal dominance. These are predictions that distinguish the exponential discounting model with both stochastic choice and preference variability, from discounting models with only stochastic choice. Although such dominance-based violations have been documented in the domain of risky and multiattribute choice (e.g. Busemeyer & Townsend, 1993; Diederich & Busemeyer, 1999; Loomes & Sugden, 1998), to our knowledge, there are no experimental tests of dominance-based SST violations in the intertemporal choice setting. Experiment 3 attempts such a test.

5.2.1. Methods and materials

Forty-eight undergraduate students from a university in the United States participated in Experiment 3 for course credits (41 female; aged 19.75 ± 1.23).

To test SST, we used triplets of options X = (\$100; t months), X_D = (\$95; t months), and Y = (\$120; $t + 3$ months), of hypothetical monetary outcomes. X and X_D were available at the same time, but X_D offered a smaller amount than X, and thus was dominated by X. We created four such triplets by choosing t from the set {0 months (immediately), 3 months, 6 months, 9 months}. The details of the four triplets can be found in Appendix C. Each of the four triplets generated three different binary choice pairs from a full combination of the three options, creating a block of 12 distinct choice pairs. For example, when $t = 0$ months, we obtained a choice between X = (\$100; 0 months) and X_D = (\$95; 0 months), between X = (\$100; 0 months) and Y = (\$120; 3 months), and between X_D = (\$95; 0 months) and Y = (\$120; 3 months).

The block of 12 distinct choice pairs was repeated ten times to obtain participants' choice frequencies for each choice pair. For the analysis below, these choice frequencies were assumed to correspond to choice probabilities. Overall there were a total of 120 trials per participant. The placement of options (left or right) was counterbalanced, and the ordering of the trials within each block was randomized.

5.2.2. Results

5.2.2.1. Qualitative examination of SST violations. We examined the two types of SST violations illustrated in Proposition 3 respectively. Our experiment yielded choice probabilities for 192 triplets (48 participants × 4 unique triplets each). The following analysis treats the 192 triplets independently. Choice probabilities for 84 triplets (43.8%) satisfied the prerequisite of Type-1 SST violation (i.e. $p[Y; X] \geq 0.5$ and $p[X; X_D] \geq 0.5$). For triplets satisfying this prerequisite, we first ruled out the cases in which $p[Y; X_D] = \max\{p[Y; X], p[X; X_D]\} = 1$. In such cases, the decision maker was too patient to violate SST with the choice pairs in the experiment. This exclusion left us with 23 triplets satisfying Type-1 prerequisite that could potentially generate SST violations. Twenty of the 23 triplets (87.0%) did in fact do so, with $p[Y; X_D] < max\{p[Y; X], p[X; X_D]\}$.

We repeated a similar analysis for triplets satisfying the prerequisite of Type-2 SST violations in Proposition 3 (i.e. $p[X; X_D] \geq 0.5$ and $p[X_D; Y] \geq 0.5$). Choice probabilities for 103 triplets (53.6%) satisfied the prerequisite of Type 2. After excluding the triplets for which $p[X; Y] = \max\{p[X; X_D], p[X_D; Y]\} = 1$ (where the decision maker was too impatient for us to detect SST violations with our stimuli), we were left with 23 triplets, out of which 19 (82.6%) violated SST, with $p[X; Y] < max\{p[X; X_D], p[X_D; Y]\}$. The total number of 46 triplets suitable for SST testing came from 22 different participants, among whom 21 generated at least one Type-1 or one Type-2 SST violation. In summary, we observed numerous SST violations in the manner predicted by the variable exponential model as long as the choice problems made the detection possible.

5.2.2.2. Quantitative QTest. Although the qualitative approach used to detect SST violations in the prior section is commonly used for such tests (e.g. Becker, DeGroot, & Marschak, 1964; Mellers & Biagini, 1994; Tversky, 1969), it is silent on the magnitude of the disagreement between the data and the predictions of SST (Regenwetter et al., 2011). To provide a more rigorous test of SST, we thus pursued a quantitative test with the state-of-the-art decision-axiom testing tool, QTest (Regenwetter et al., 2014; see Regenwetter et al., 2018 for a recent application of QTest in intertemporal choice modeling). Intuitively, SST was considered as a constrained probabilistic choice model and we contrasted its performance on the data to that of a probabilistic choice model constrained only by the prerequisite of SST, but not the corresponding conclusion. For example, if $p[Y; X] \geq 0.5$ and $p[X; X_D] \geq 0.5$, the SST model predicts $p[Y; X_D] \geq \max\{p[X; X_D], p[X_D; Y]\}$ as the conclusion. Consequently the SST model is constrained by $p[Y; X] \geq 0.5$, $p[X; X_D] \geq 0.5$, and $p[Y; X_D] \geq \max\{p[X; X_D], p[X_D; Y]\}$. The prerequisite of SST, which we refer to as the "SSTpre" model, is only constrained by $p[Y; X] \geq 0.5$ and $p[X; X_D] \geq 0.5$ and allows for $0 \leq p[Y; X_D] \leq 1$. In line with Experiments 1 and 2, we used the Bayes factors offered in QTest for the comparisons of the SST and the SSTpre models. Note that this Bayes factor was not directly available in QTest because QTest only allowed us to compare the SST model with a saturated model. However, the Bayes factor could still be estimated by formula: $BF_{SST,SSTpre} = BF_{SST,saturated} \cdot BF_{saturated,SSTpre}$.

We restricted our Bayesian analysis to the 46 triplets that allowed us to test SST violations (i.e. 23 Type-1 triplets and 23 Type-2 triplets). The convergence of the Bayesian estimation was checked by running two independent simulations. Across the 23 Type-1 triplets, the aggregate Bayes factor between the SST and the SSTpre models suggested substantial SST violations, with $BF_{SST,SSTpre} = 3.9 \times 10^{-6}$. Likewise, the same aggregate Bayes factor across the 23 Type-2 triplets was 5.2×10^{-9}. Unpacking the aggregate Bayes factors, there were seven Type-1 triplets and eleven Type-2 triplets suggesting SST violations (with $BF_{SST,SSTpre} < 1/3$). Conversely, only two Type-1 triplets and two Type-2 triplets conformed to strong stochastic transitivity (with $BF_{SST,SSTpre} > 3$). Other triplets offered anecdotal evidence in discriminating the two models ($1/3 \leq BF_{SST,SSTpre} \leq 3$). Thus, even with the stringent QTest standard, we still observed evidence for SST violations in the fashion as predicted by the variable exponential model, when the choice problems made the detection possible.

Of all the 141 triplets that were not suitable for the test of SST violations (i.e. sixty-one Type-1 triplets with $p[Y; X_D] = \max\{p[Y; X], p[X; X_D]\} = 1$ and eighty Type-2 triplets with $p[X; Y] = \max\{p[X; X_D], p[X_D; Y]\} = 1$), one hundred and thirty-nine triplets (98.6%) obtained anecdotal evidence for either model (i.e. $1/3 \leq BF_{SST,SSTpre} \leq 3$) and only two Type-2 triplets achieved Bayes factors of 3.01 and 3.05 respectively, slightly favoring the SST model. These results re-affirmed that these triplets were non-informative for the test of SST violations.

Note that we relied on the observed choice frequencies as a proxy of the underlying choice probabilities to determine whether a prerequisite was satisfied. The procedure was not flawless because the observed choice frequencies may not perfectly reflect the unobserved choice probabilities. For example, choosing Y over X six times in 10 repetitions could be generated by $p[Y; X] = 0.6$, but could also be generated by $p[Y; X] = 0.3$ albeit with a smaller chance. One way to sidestep this issue is to test the full SST constraints (i.e. by considering all the six transitive patterns simultaneously: $Y \succcurlyeq X \succcurlyeq X_D$; $Y \succcurlyeq X_D \succcurlyeq X$; $X \succcurlyeq Y \succcurlyeq X_D$; $X \succcurlyeq X_D \succcurlyeq Y$; $X_D \succcurlyeq Y \succcurlyeq X$ and $X_D \succcurlyeq X \succcurlyeq Y$) against the saturated model. In this analysis, the aggregate Bayes factor across all data is $BF_{full-SST,saturated} = 3.8 \times 10^{-31}$, suggesting substantial evidence for SST violations. When looking at individual Bayes factors, we however found that 89% of the Bayes factors were between 0.65 and 0.72 (i.e. slightly tilted towards SST violations). The dilution of the detected effect sizes (compared with $BF_{SST,SSTpre} < 1/3$ in earlier analyses) was probably due to the fact that both the full-SST model and the saturated model involved large volumes of space that were far away from the high-density regions. Nevertheless, the overall direction of the evidence for SST violations is consistent with other analyses above.

5.2.2.3. Summary and discussion. We tested SST violations in the presence of intertemporal dominance because they distinguish the proposed variable exponential model from mainstream discounting models with stochastic choice alone. Only the variable exponential model, with both preference variability and stochastic choice, can accommodate such violations (though other models with both preference variability and stochastic choice, such as the variable quasi-hyperbolic model, may be able to do so as well). The exponential discounting model with only stochastic choice, as well as the hyperbolic models with only stochastic choice, always predict that SST is satisfied. Unlike Dai (2017), who found that people frequently conform to transitivity in intertemporal choice in the absence of dominance, we found that SST is violated in the presence of intertemporal dominance in Experiment 3. The violations emerged both with the conventional criterion as well as with the Bayesian implementation of the state-of-the-art QTest criterion.

How accurately can the variable exponential model quantitatively predict the data? As SST is not a parametric model and we could not fit models in Experiment 3, we did not directly evaluate the variable exponential model's quantitative predictions. However, we can still speculate about these quantitative predictions, by examining the mechanism by which the variable exponential model generates SST violations in the presence of intertemporal dominance. As shown in Proposition 3, the variable exponential model predicts SST violations by disproportionately avoiding dominated options (relative to non-dominated options). In Experiments 1 and 2, we found that the variable exponential model predicted more dominance violations than observed in our choice data. Thus, it is likely that the variable exponential model under-predicts the number of SST violations.

There are of course other ways to explain the SST violations observed in the experiment. For example, sequential sampling models, which probabilistically attend to different time periods, are capable of generating reduced choice shares for dominated options, and thus can also lead to SST violations (e.g. Busemeyer & Townsend, 1993; Diederich & Busemeyer, 1999; see also Dai & Busemeyer, 2014). Future work should attempt to quantitatively compare different explanations of dominance-based SST violations, so as to better understand the psychological mechanisms underlying probabilistic intertemporal choice.

6. General discussion

Many everyday decisions involve a tradeoff between smaller payoffs obtained earlier, and larger payoffs obtained later, and understanding how these tradeoffs are made is a key area of interest across numerous disciplines, including psychology, economics, and neuroscience. The standard model of intertemporal decision making proposes that people discount future payoffs exponentially, with the magnitude of discounting determined by their time preference –that is, their discount factor (Samuelson, 1937). This is not only a parsimonious model; it also satisfies a number of normative decision axioms, and is thus widely seen the standard approach to studying intertemporal choice (Frederick et al., 2002; Koopmans, 1960).

This paper has re-examined the explanatory power of the exponential discounting model by incorporating both stochastic choice and preference variability. To model preference variability, we have proposed a modification to the exponential model that allows the discount factor to fluctuate from one decision to the next. Our interpretation is that people do not have consistent time preferences determined by a single fixed discount factor, but instead determine their discount factor when confronted with a particular decision in a manner that depends on unobserved aspects of cognition (e.g. one's emotional state, attention to the decision stimuli, or memory for past experiences). We have derived key properties of the variable exponential model and tested the predictions of the proposed model with three experiments. Our results indicate that an exponential model of intertemporal decision making that permits variability in the degree of time preference has explanatory power. By doing so, it complements a rich existing literature in psychology on the descriptive role of choice stochasticity and parameter variability in cognition and behavior.

6.1. Inferring time preferences

Although the focus of this paper has been on understanding the theoretical properties of an exponential discounting model with variability in discount factors, our results also have implications for empirical work on time preference. Notably, the results outlined in Section 4 (e.g. Proposition 1) show that modal choices can be incomplete guides to time preferences: In the presence of both stochastic choice and preference variability, it is possible to choose one option more frequently than another, but nonetheless assign a higher utility to the other option when using the underlying (expected) discount factor. Likewise, with both types of randomness, it is possible to choose one option more frequently than another, but nonetheless assign a higher utility to the other option more than half the time. Prior work has documented the effect of different stochastic specifications on quantitative inferences based on model fits to intertemporal choice data (Dai & Busemeyer, 2014; Wulff & van den Bos, 2018; see Blavatskyy & Pogrebna, 2010 for similar findings in risky choice modeling). Our results extend these restrictions to qualitative inferences using modal choice proportions. If a person relies on a variable and stochastic choice process, modal choices do not reveal preferences; the underlying preferences cannot be identified separately from the stochastic structure of the choice process.

It may also be problematic to use modal choice proportions to infer the contextual, demographic, and neural correlates of time preference. Altering the degree of choice stochasticity and preference variability can alter modal choice proportions, implying that observed differences in these choice proportions across experimental conditions, demographic groups, or brain regions, could be due to different degrees of randomness, rather than systematic differences in time preference.

6.2. Generality of our approach

For practical reasons, we constrained the scope of our analysis in several ways. For example, we specified stochastic choice using the logit function and assumed a uniform distribution for the discount factor δ, and our claims regarding quantitative fits for the variable exponential model are contingent on stochastic choice and parameter variability meeting these assumptions. That said, we do not believe that changing these assumptions will have a large effect on our conclusions. For example, replacing the logit function with a probit function will likely yield the same relative model fit statistics. That is because the two choice functions are very similar to each other (probit only has a slightly fatter tail than logit), so much so that some computational tools even use logit as an approximation of probit (e.g. Carpenter et al., 2017). Replacing the uniform distribution over δ with a different symmetric distribution will also likely yield the same relative fit statistics. Indeed, the analytical derivations provided in Section 4 are not restricted to uniform distributions; they hold more generally. That said, there is room to extend our analysis to permit other probabilistic specifications. Doing so will shed light on how different assumptions about variability in the choice process can be combined to better describe intertemporal decision making (Blavatskyy & Pogrebna, 2010; Dai & Busemeyer, 2014; Dai, Pleskac, & Pachur, 2018; Loomes & Sugden, 1995; Regenwetter & Robinson, 2017; Regenwetter et al., 2018).

This paper has also mostly tested the assumption of preference variability through the lens of the exponential discounting model. Of course, one could also allow the parameters of other intertemporal choice models to fluctuate (such as the variable quasi-hyperbolic model in the model selection tournament). We did not extensively study these types of variable hyperbolic models, as our goal was merely to understand the effects of preference variability in a simple model of time preference, and the variable exponential model already performed well. As the variable exponential model can account for key patterns in choice behavior, more complex assumptions regarding intertemporal discounting may not be necessary. That said, future work could attempt to modify hyperbolic discounting models to admit preference variability. If these models are in fact accurate descriptors of the underlying choice process, it is likely that their parameters do not stay fixed over trials. Allowing for variability in these parameters can not only improve model fit, but can also potentially expand the set of choice patterns that are accommodated by hyperbolic models. Variability can also be applied to subjective valuation of payoffs and subjective perception of time in models that assume nonlinear transformations to them. More generally, the approach proposed in this paper can be applied to any set of parametrized intertemporal choice models, including those outside of the discounting framework (e.g. Dai & Busemeyer, 2014; Scholten & Read, 2010; Scholten et al., 2014). Many of these models do propose parametric measures of time preference, which can be reasonably assumed to vary across trials.

Rather than modifying existing utility maximization models to permit preference variability, it may be better instead to build models that make explicit the memory and attention processes responsible for forming preferences and determining choice. These processes are often studied within the accumulation-to-threshold framework, and there has been some recent work extending models within this framework to intertemporal choice tasks (e.g. Dai & Busemeyer, 2014; Dai et al., 2018; Rodriguez, Turner, & McClure, 2014; Rodriguez, Turner, Van Zandt, & McClure, 2015; Zhao, Diederich, Trueblood, & Bhatia, in press). Some of the accumulation-to-threshold also assume variability in the start point of the evidence accumulation process (e.g. the linear ballistic accumulator, Brown & Heathcote, 2008). Unlike utility maximization models, accumulation-to-threshold models come equipped with well-specified, psychologically grounded assumptions regarding the sources of variability, typically in the form of stochasticity in attention or memory retrieval. Not only do these models provide a more detailed theoretical account of the cognitive processes responsible for choice (as well as the role of stochasticity in the choice process), they also have a number of desirable descriptive properties involving intertemporal dominance, reaction time, confidence, and other relevant psychological variables. The consideration of preference variability would undoubtedly complement the rich insights from those well-grounded cognitive models. It is even possible to make some psychological variables observable by the use of attitude scales or by incorporating them as latent variables in the models (see e.g. Boehm, Steingroever, & Wagenmakers, 2018). We hope that future work will more closely incorporate the insights of these cognitive models into the study of intertemporal choice behavior.

Finally, we focused our analysis on the discounting of monetary payoffs in this paper. However, the idea of variability in time preference also applies to other domains, such as food, leisure activities and health-related states. Due to the fact that empirical evidence regarding behavioral regularities in these domains is somewhat inconclusive (see Read et al., 2018 for a recent review), we do not want to extend our claims about the power of the variable exponential model beyond monetary payoffs. Further research is needed to test the descriptive accuracy of preference variability in intertemporal choice models of non-monetary choice.

6.3. Inter-individual heterogeneity

In this paper, we present a model that combines two sources of randomness in the individual decision process: stochastic choice and preference variability. Our primary analysis examines how well this model describes individual-level intertemporal choice (as in e.g. Bhatia & Loomes, 2017; Regenwetter & Robinson, 2017; Regenwetter et al., 2018). Another body of literature has applied similar ideas to the analysis of group-level behavior, in which the choices of a number of different decision makers are aggregated into a group choice (see Adams, Cherchye, De Rock, & Verriest, 2014; Apesteguia & Ballester, 2016; Gollier & Zeckhauser, 2005; Golman, 2011, 2012; Jackson & Yariv, 2014, 2015, 2018; Weitzman, 2001). Here, different members of the group are assumed to have different tastes (a form of between-individual preference heterogeneity, analogous to our assumption of preference variability on the individual level), and may also be subject to error in making choices (matching our assumption of individual-level stochastic choice). There are well-established probabilistic models that permit both between-individual preference heterogeneity and their choice stochasticity simultaneously (e.g. McFadden & Train, 2000), and our proposed variable exponential model can be seen as a modification of this type of probabilistic model, used to describe individual-level decision making.

Research on group-level behavior has identified aggregation problems in intertemporal choice, i.e., problems inherent in combining individual preferences into group preferences. This work has found that an increasing discount factor at the group level can emerge from heterogeneous constant (exponential) discount factors on the individual level (Gollier & Zeckhauser, 2005; Jackson & Yariv, 2014, 2015; Killeen and Taylor, Unpublished manuscript; Lewandowsky, Freeman, & Mann, 2017; Weitzman, 2001). Echoing the warning of the aggregation fallacy (Regenwetter & Robinson, 2017), this work suggests that observed decreasing impatience at the group-level does not guarantee a single individual actually showing decreasing impatience. Again, there is a close link between these results and the findings presented in this paper (particularly our finding that decreasing impatience can be generated by an exponential discounting model with variable discount factors). Although much of our other analysis (such as our propositions regarding dominance and strong stochastic transitivity, as well as our experimental tests and model fits) is specific to the individual level; this nonetheless suggests that a core set of theoretical insights can be used to describe both the aggregation of heterogeneous preferences into group choices, as well as the aggregation of variable preferences into individual choices. These theoretical insights pertain primarily to the complex effects of variability, and we look forward to future research that uses these insights to build better models of cognition and behavior.

Appendix A. Proofs of propositions

Proposition 1. *For any $\bar{\delta} \in (0, 1)$ and $\theta > 0$ in the variable exponential model (Eq. (9)), there exists choice pair (X, Y), with $X = (x; t)$ and $Y = (y; t + k)$, where $0 < x < y$, $t \geq 0$ and $k > 1$, such that*

1. $p[X; Y] > 50\%$ *if* $f(\delta) = \begin{cases} 1, & \text{if } \delta = \bar{\delta} \\ 0, & \text{if } \delta \neq \bar{\delta} \end{cases}$ *(i.e. in the absence of preference variability), but*
2. $p[X; Y] < 50\%$ *if $f(\delta)$ is non-degenerate and symmetric around $\bar{\delta}$ (i.e. in the presence of preference variability).*

Proof. We will first construct a choice pair $X' = (x'; t)$ and $Y = (y; t + k)$ (with $k > 1$) such that $p[X'; Y] = 50\%$ with no preference variability, but $p[X'; Y] < 50\%$ with preference variability. Then, because the probability p is continuous in the payoffs, we can take X to offer a slightly higher payoff than X' to construct the choice pair (X, Y).

Consider X' and Y such that $x' = \bar{\delta}^k y$. Then $\bar{\delta}^t x' - \bar{\delta}^{t+k} y = 0$, so with no preference variation, $p[X'; Y] = 0.5$. We will now show that with preference variation, $p[X'; Y] < 0.5$.

Observe that for $\delta < \bar{\delta}$, we have $\delta^t < \bar{\delta}^t$ and $x' - \delta^k y > 0$, whereas for $\delta > \bar{\delta}$, these inequalities are reversed. This implies that $\delta^t (x' - \delta^k y) < \bar{\delta}^t (x' - \delta^k y)$ for any $\delta \neq \bar{\delta}$. We can plug this into the expression for $g(X', Y|\delta)$ (see Eq. (9) in the main article) to get:

$$\int \left(g(X', Y|\delta) - \frac{1}{2}\right) f(\delta) d\delta < \int \left(\frac{1}{1 + \exp\{-\theta \bar{\delta}^t (x' - \delta^k y)\}} - \frac{1}{2}\right) f(\delta) d\delta.$$

Let $h(\delta) = x^{'} - \delta^k y$ and observe that h is strictly concave. Let $L(\delta) = a + b\delta$be the line tangent to $h(\delta)$ at $\bar{\delta}$. Because h is strictly concave, we know $L(\delta)\geq h(\delta)$ with equality only at $\delta = \bar{\delta}$. This implies that

$$\int \left(\frac{1}{1 + \exp\{-\theta\bar{\delta}^t(x' - \delta^k y)\}} - \frac{1}{2}\right) f(\delta)d\delta < \int \left(\frac{1}{1 + \exp\{-\theta\bar{\delta}^t(a + b\delta)\}} - \frac{1}{2}\right) f(\delta)d\delta$$

(as long as the distribution $f(\cdot)$ is not degenerate). We can evaluate the right-hand integral by substituting in $\hat{\delta} = \delta - \bar{\delta}$, noting that $a + b\bar{\delta} = 0$ (because $L(\bar{\delta}) = h(\bar{\delta}) = 0$), and observing that $\left(\frac{1}{1 + \exp\{-\theta\bar{\delta}^t(b\hat{\delta})\}} - \frac{1}{2}\right)$ is an odd function of $\hat{\delta}$. Thus, given our assumption that $f(\cdot)$ is symmetric around $\bar{\delta}$, we find that:

$$\int \left(\frac{1}{1 + \exp\{-\theta\bar{\delta}^t(a + b\delta)\}} - \frac{1}{2}\right) f(\delta)d\delta = 0.$$

Putting all of our inequalities together, we obtain

$$\int \left(g(X', Y|\delta) - \frac{1}{2}\right) f(\delta)d\delta < 0,$$

which implies that with preference variation, $p[X'; Y] < \frac{1}{2}$.

We can now consider a choice between $X = (x; t)$ and Y where $x = x^{'} + \varepsilon$ for some sufficiently small ε. From the continuity of p, we find $p[X; Y] > \frac{1}{2}$ when there is no preference variation whereas $p[X; Y] < \frac{1}{2}$ when there is preference variation.

Proposition 2. *Consider choice options* $X = (x; t)$, $X_D = (x_D; t_D)$ *and* $Y = (y; s)$, *such that* $x \geq x_D$ *and* $t \leq t_D$ (*with at least one of these inequalities strict*), *but either* $x < y$ *or* $t > s$ (*not both*), *and suppose that in a choice between* X_D *and* Y, *we have* $p[Y; X_D] = p[X_D; Y] = 0.5$.

1. *If the distribution* $f(\delta)$ *in the variable exponential model* (*Eq.* (9)) *is degenerate* (*i.e., in the absence of preference variability*), *then in distinct binary choices,* $p[X; X_D] = p[X; Y]$.
2. *If the distribution of* $f(\delta)$ *has full support on* $[0, 1]$, *then*

$$\lim_{\theta\to\infty} p[X_D; X] = 0,$$

but

$$\lim_{\theta\to\infty} \Pr[Y; X] > 0,$$

i.e., for large enough θ, *we have* $p[Y; X] > p[X_D; X]$.

Proof. Part 1. In the absence of preference variability, the given $p[Y; X_D] = 0.5$ implies that $\delta^s y = \delta^{t_D} x_D$. Thus, $\delta^s y - \delta^t x = \delta^{t_D} x_D - \delta^t x$, and so $p[X_D; X] = p[Y; X]$.

Part 2. Regardless of the value of δ in (0,1), dominance implies that $\delta^{t_D} x_D - \delta^t x < 0$, so $\lim_{\theta\to\infty} p[X_D; X] = 0$. However, there exists some δ in (0, 1) such that $\delta^s y - \delta^t x > 0$, so as long as f has full support, $\lim_{\theta\to\infty} p[Y; X] > 0$.

Proposition 3. *Given any non-degenerate distribution* $f(\delta)$ *and* $\theta > 0$ *in the variable exponential model* (*Eq.* (9)), *there exist options* $X = (x; t)$, $X_D = (x_D; t)$ (*where* $x_D < x$), *and* $Y = (y; s)$ (*where* $y > x$ *and* $s > t$) *such that*:

1. $p[Y; X] \geq 0.5$ *and* $p[X; X_D] \geq 0.5$, *but* $p[Y; X_D] < \max\{p[Y; X], p[X; X_D]\}$ *or*
2. $p[X; X_D] \geq 0.5$ *and* $p[X_D; Y] \geq 0.5$, *but* $p[X; Y] < max\{p[X; X_D], p[X_D; Y]\}$.

Proof. Part 1. We will construct choice options such that $p[Y; X] = 0.5$ and $p[X; X_D] > 0.5$, but $p[Y; X_D] < p[X; X_D]$.

Consider choice options $X = (x; 0)$ and $X_D = (x_D; 0)$ offering immediate payoffs and choice option $Y = (y; s)$ such that $y > x > x_D$ with $x - x_D \gg 0$ and

$$p[Y; X] = \int f(\delta)\frac{1}{1 + e^{-\theta(\delta^S y - x)}}d\delta = 0.5.$$

(We can think of this equation as specifying the choice option Y that we will consider, given some choice option X. That is, we select values of $(y; s)$ so that it holds.)

We can immediately see that $p[X; X_D] > 0.5$ because X dominates X_D. We will now show that $p[Y; X_D] < p[X; X_D]$.

Observing that $\frac{1}{1+z}$ is a convex function of z, we apply Jensen's inequality, letting $z = e^{-\theta(\delta^S y - x)}$, to determine that

$$\int f(\delta)\frac{1}{1 + e^{-\theta(\delta^S y - x)}}d\delta > \frac{1}{1 + \int f(\delta)e^{-\theta(\delta^S y - x)}d\delta}.$$

This implies that

$$\int f(\delta)e^{-\theta(\delta^S y - x)}d\delta > 1 \tag{A1}$$

We have

$$p[X; X_D] = \frac{1}{1 + e^{-\theta(x-x_D)}}.$$

Using a Taylor series around $\varepsilon = e^{-\theta(x-x_D)} \approx 0$, we can write this as

$$p[X; X_D] = 1 - \varepsilon + O(\varepsilon^2).$$

We also have

$$p[Y; X_D] = \int f(\delta)\frac{1}{1 + e^{-\theta(\delta^s y - x_D)}}d\delta$$

$$= \int f(\delta)\frac{1}{1 + e^{-\theta(x-x_D)}e^{-\theta(\delta^s y - x)}}d\delta.$$

Using a Taylor series around $\varepsilon = e^{-\theta(x-x_D)} \approx 0$, we can write this as

$$p[Y; X_D] = 1 - \left(\int f(\delta)e^{-\theta(\delta^s y - x)}d\delta\right)\varepsilon + O(\varepsilon^2).$$

Using the previous two equations and Eq. A(1), we find that

$$p[Y; X_D] < p[X; X_D]$$

as long as $x - x_D$ is sufficiently large. This violates strong stochastic transitivity.

Part 2. We will construct choice options such that $p[X; X_D] > 0.5$ and $p[X_D; Y] = 0.5$, but $p[X; Y] < p[X; X_D]$.

Consider choice options $X = (x; 0)$ and $X_D = (x_D; 0)$ offering immediate payoffs and choice option $Y = (y; s)$ such that $y > x > x_D$ with $x - x_D \ll 1$ and

$$p[X_D; Y] = \int f(\delta)\frac{1}{1 + e^{-\theta(x_D - \delta^s y)}}d\delta = 0.5. \tag{A2}$$

Again we have $p[X; X_D] > 0.5$ because X dominates X_D. We will now show that $p[X; Y] < p[X; X_D]$.

We have

$$p[X; X_D] = \frac{1}{1 + e^{-\theta(x-x_D)}}.$$

Using a Taylor series around $x - x_D \approx 0$, we can write this as

$$p[X; X_D] = \frac{1}{2} + \frac{1}{4}\theta(x - x_D) + O((x - x_D)^2).$$

But we also have

$$p[X; Y] = \int f(\delta)\frac{1}{1 + e^{-\theta(x - \delta^s y)}}d\delta$$

$$= \int f(\delta)\frac{1}{1 + e^{-\theta(x-x_D)}e^{-\theta(x_D - \delta^s y)}}d\delta.$$

Using a Taylor series around $x - x_D \approx 0$, we can write this as

$$p[X; Y] = \int f(\delta)\frac{1}{1 + e^{-\theta(x_D - \delta^s y)}}d\delta + \int f(\delta)\frac{e^{-\theta(x_D - \delta^s y)}}{(1 + e^{-\theta(x_D - \delta^s y)})^2}d\delta\theta(x - x_D) + O((x - x_D)^2).$$

Algebraic manipulation of the second term allows us to rewrite this as

$$p[X; Y] = \int f(\delta)\frac{1}{1 + e^{-\theta(x_D - \delta^s y)}}d\delta + \int f(\delta)\frac{1}{1 + e^{-\theta(x_D - \delta^s y)}} - \frac{1}{(1 + e^{-\theta(x_D - \delta^s y)})^2}d\delta\theta(x - x_D) + O((x - x_D)^2). \tag{A3}$$

Observing that z^2 is a convex function of z, we apply Jensen's inequality, letting $z = \frac{1}{1 + e^{-\theta(x_D - \delta^s y)}}$, to determine that

$$\int f(\delta)\frac{1}{(1 + e^{-\theta(x_D - \delta^s y)})^2}d\delta > \left(\int f(\delta)\frac{1}{1 + e^{-\theta(x_D - \delta^s y)}}d\delta\right)^2.$$

Table A1
Space partitioning results from the ten independent repetitions.

Repetition	Proportion of the space (%)		
	Decreasing impatience	Increasing impatience	Neither
1	0.73	0	99.27
2	0.73	0	99.27
3	0.74	0	99.26
4	0.73	0	99.27
5	0.75	0	99.25
6	0.74	0	99.26
7	0.73	0	99.27
8	0.72	0	99.28
9	0.75	0	99.25
10	0.72	0	99.28

Inserting this inequality into Eq. A(3) and retaining only the constant and linear terms in the Taylor approximation, we get $p[X;Y] < \int f(\delta)\frac{1}{1+e^{-\theta(x_D-\delta^s y)}}d\delta + \left[\int f(\delta)\frac{1}{1+e^{-\theta(x_D-\delta^s y)}}d\delta - \left(\int f(\delta)\frac{1}{1+e^{-\theta(x_D-\delta^s y)}}d\delta\right)^2\right]\theta(x-x_D)$.

We now use Eq. A(2), which gives us these integrals, to simplify this as

$$p[X;Y] < \frac{1}{2} + \frac{1}{4}\theta(x-x_D).$$

We have thus found that $p[X;Y] < p[X;X_D]$ as long as $x-x_D$ is sufficiently small. This violates strong stochastic transitivity.

Appendix B. Parameter space partitioning test

We applied the Parameter Space Partitioning tool to test whether the variable exponential model generally displays decreasing impatience (see Pitt et al., 2006 for technical details). We embedded the variable exponential model in a logistic choice rule for stochastic choice, and assumed a variable discount factor $\delta \sim \text{Uniform}[\bar{\delta}-\eta, \bar{\delta}+\eta]$ for preference variability.

We set up both a parameter space and a reasonable stimulus space for the partitioning search. The parameter space had three dimensions, representing the upper bound of the variable discount factor ($\delta_+ = \bar{\delta}+\eta$, $\delta_+ \in [0, 1]$), the ratio of the lower bound discount factor to the upper bound ($\lambda = \frac{\bar{\delta}-\eta}{\bar{\delta}+\eta}$, $\lambda \in [0, 1]$), and the choice stochasticity ($\theta \in [0, 100]$).

The stimulus space had five dimensions, representing the sooner delay ($t \in [0, 100]$), the sooner payoff ($x \in [1, 100]$), the ratio of the sooner delay to the later delay ($\mu = \frac{t}{s}$, $\mu \in [0, 1]$), the ratio of the sooner payoff to the later payoff ($\omega = \frac{x}{y}$, $\omega \in [0, 1]$), and an additional front-end delay k. With values on these five dimensions, we created two choice problems: (1) A choice between $X_P = (x; t)$ and $Y_P = (y; s)$ and (2) A choice between $X_R = (x; t+l)$ and $Y_R = (y; s+l)$, where $y > x > 0$, $s > t \geq 0$ and $l > 0$. For each sample of parameters and stimuli, we calculated $p[X_P; Y_P]$ and $p[X_R; Y_R]$, which can yield decreasing impatience ($p[X_P; Y_P] \geq 0.5$ and $p[X_R; Y_R] < 0.5$), increasing impatience ($p[X_P; Y_P] < 0.5$ and $p[X_R; Y_R] \geq 0.5$) or neither. The computational method for calculating choice probabilities was the same as for model fitting in Section 5.1.2 in the main article.

We ran Parameter Space Partitioning search for 10 independent repetitions to obtain reliable results. In each repetition, this search tool implements a sophisticated two-stage Markov chain Monte Carlo (MCMC) algorithm to identify data patterns. To further ensure accuracy in pattern detection, we increased the number of MCMC samples to 100,000 for Stage-1 burn-in adaptation and 200,000 for Stage-2 formal adaptation within each cycle (the algorithm runs a maximum of six cycles by default). We also used the "HitMiss" method for volume estimation with an additional 100,000 MCMC samples, as recommended by Pitt et al. (2006). We found the results were highly consistent across repetitions, with approximately 0.73% displaying decreasing impatience and none increasing impatience. Table A1 shows the proportions of the partitioned space obtained from the 10 repetitions.

We ran an additional parameter space partitioning test to determine whether the prediction of decreasing was restricted to a limited region of the parameters. To this end, we divided the whole parameter space into ten subspaces by evenly dividing the upper bound of the variable discount factor (δ_+) into ten sub-intervals, such as 0–0.1 and 0.4–0.5. We found that there are instances of decreasing impatience predictions in all of the ten subspaces, although the majority of them gathered in subspaces with $\delta_+ > 0.8$ (see Fig. A1).

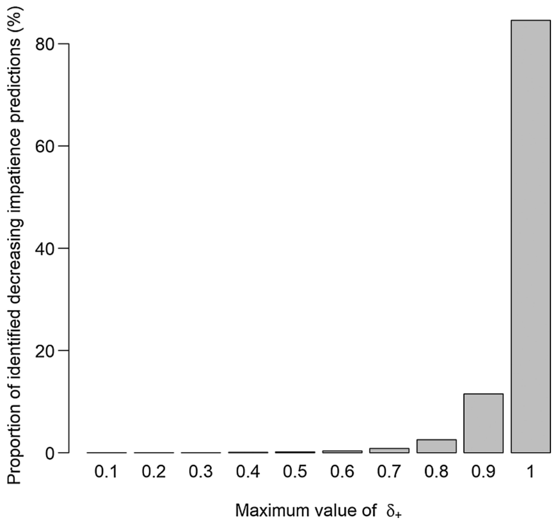

Fig. A1. The distribution of decreasing impatience predictions over sub-parameter spaces. The labels on the *x*-axis represent the maximum value of δ_+ for each subspace. The corresponding minimum δ_+ value is always the maximum value minus 0.1.

Appendix C. Experimental designs

See Tables A2–A4.

Table A2
Experimental design in Experiment 1: Sixty choice problems between sooner (*X*) and later (*Y*) options, along with the annual interest rates implied by the indifference between the options.

Sooner option (*X*)		Later option (*Y*)		Implied annual interest rate
Amount ($)	Delay (months)	Amount ($)	Delay (months)	
100	0	84	3	−50%
100	0	111	3	50%
100	0	119	3	100%
100	0	157	3	500%
100	0	182	3	1000%
100	0	71	6	−50%
100	0	122	6	50%
100	0	141	6	100%
100	0	245	6	500%
100	0	332	6	1000%
100	0	59	9	−50%
100	0	136	9	50%
100	0	168	9	100%
100	0	383	9	500%
100	0	604	9	1000%
100	3	84	6	−50%
100	3	111	6	50%
100	3	119	6	100%
100	3	157	6	500%
100	3	182	6	1000%
100	3	71	9	−50%
100	3	122	9	50%
100	3	141	9	100%
100	3	245	9	500%
100	3	332	9	1000%
100	3	59	12	−50%
100	3	136	12	50%
100	3	168	12	100%
100	3	383	12	500%
100	3	604	12	1000%
100	6	84	9	−50%
100	6	111	9	50%

(*continued on next page*)

Table A2 (*continued*)

Sooner option (*X*)		Later option (*Y*)		Implied annual interest rate
Amount ($)	Delay (months)	Amount ($)	Delay (months)	
100	6	119	9	100%
100	6	157	9	500%
100	6	182	9	1000%
100	6	71	12	−50%
100	6	122	12	50%
100	6	141	12	100%
100	6	245	12	500%
100	6	332	12	1000%
100	6	59	15	−50%
100	6	136	15	50%
100	6	168	15	100%
100	6	383	15	500%
100	6	604	15	1000%
100	9	84	12	−50%
100	9	111	12	50%
100	9	119	12	100%
100	9	157	12	500%
100	9	182	12	1000%
100	9	71	15	−50%
100	9	122	15	50%
100	9	141	15	100%
100	9	245	15	500%
100	9	332	15	1000%
100	9	59	18	−50%
100	9	136	18	50%
100	9	168	18	100%
100	9	383	18	500%
100	9	604	18	1000%

Table A3
Experimental design in Experiment 2: Sixty choice problems between sooner (*X*) and later (*Y*) options, along with the annual interest rates implied by the indifference between the options.

Sooner option (*X*)		Later option (*Y*)		Implied annual interest rate
Amount ($)	Delay (months)	Amount ($)	Delay (months)	
100	0	93	3	−25%
100	0	106	3	25%
100	0	115	3	75%
100	0	122	3	125%
100	0	129	3	175%
100	0	87	6	−25%
100	0	112	6	25%
100	0	132	6	75%
100	0	150	6	125%
100	0	166	6	175%
100	0	81	9	−25%
100	0	118	9	25%
100	0	152	9	75%
100	0	184	9	125%
100	0	214	9	175%
100	3	93	6	−25%
100	3	106	6	25%
100	3	115	6	75%
100	3	122	6	125%
100	3	129	6	175%
100	3	87	9	−25%
100	3	112	9	25%
100	3	132	9	75%
100	3	150	9	125%
100	3	166	9	175%
100	3	81	12	−25%

(*continued on next page*)

Table A3 (*continued*)

Sooner option (X)		Later option (Y)		Implied annual interest rate
Amount ($)	Delay (months)	Amount ($)	Delay (months)	
100	3	118	12	25%
100	3	152	12	75%
100	3	184	12	125%
100	3	214	12	175%
100	6	93	9	−25%
100	6	106	9	25%
100	6	115	9	75%
100	6	122	9	125%
100	6	129	9	175%
100	6	87	12	−25%
100	6	112	12	25%
100	6	132	12	75%
100	6	150	12	125%
100	6	166	12	175%
100	6	81	15	−25%
100	6	118	15	25%
100	6	152	15	75%
100	6	184	15	125%
100	6	214	15	175%
100	9	93	12	−25%
100	9	106	12	25%
100	9	115	12	75%
100	9	122	12	125%
100	9	129	12	175%
100	9	87	15	−25%
100	9	112	15	25%
100	9	132	15	75%
100	9	150	15	125%
100	9	166	15	175%
100	9	81	18	−25%
100	9	118	18	25%
100	9	152	18	75%
100	9	184	18	125%
100	9	214	18	175%

Table A4
Experimental design in Experiment 3: Four triplets of options X, Y and X_D.

Sooner option (X)		Later option (Y)		Dominated option (X_D)	
Amount ($)	Delay (months)	Amount ($)	Delay (months)	Amount ($)	Delay (months)
100	0	120	3	95	0
100	3	120	6	95	3
100	6	120	9	95	6
100	9	120	12	95	9

Appendix D. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2019.03.003.

References

Adams, A., Cherchye, L., De Rock, B., & Verriest, E. (2014). Consume now or later? Time inconsistency, collective choice, and revealed preference. *American Economic Review, 104*(12), 4147–4183.
Ainslie, G., & Herrnstein, R. J. (1981). Preference reversal and delayed reinforcement. *Animal Learning & Behavior, 9*(4), 476–482.
Apesteguia, J., & Ballester, M. A. (2016). Stochastic representative agent. UPF economics working paper no. 1536.
Attema, A. E., Bleichrodt, H., Rohde, K. I., & Wakker, P. P. (2010). Time-tradeoff sequences for analyzing discounting and time inconsistency. *Management Science, 56*(11), 2015–2030.
Becker, G. M., DeGroot, M. H., & Marschak, J. (1963). An experimental study of some stochastic models for wagers. *Behavioral Science, 8*(3), 199–202.

Becker, G. M., DeGroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science, 9*(3), 226–232.
Bhatia, S., & Loomes, G. (2017). Noisy preferences in risky choice: A cautionary note. *Psychological Review, 124*(5), 678–687.
Blavatskyy, P. R. (2014). Stronger utility. *Theory and Decision, 76*(2), 265–286.
Blavatskyy, P. R., & Maafi, H. (2018). Estimating representations of time preferences and models of probabilistic intertemporal choice on experimental data. *Journal of Risk and Uncertainty, 56*(3), 259–287.
Blavatskyy, P. R., & Pogrebna, G. (2010). Models of stochastic choice and decision theories: Why both are important for analyzing decisions. *Journal of Applied Econometrics, 25*(6), 963–986.
Boehm, U., Steingroever, H., & Wagenmakers, E.-J. (2018). Using Bayesian regression to test hypotheses about relationships between parameters and covariates in cognitive models. *Behavior Research Methods, 50*, 1–22.
Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*(4), 434–455.
Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153–178.
Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432–459.
Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology, 44*(1), 171–189.
Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1–32.
Cavagnaro, D. R., Aranovich, G. J., McClure, S. M., Pitt, M. A., & Myung, J. I. (2016). On the functional form of temporal discounting: An optimized adaptive test. *Journal of Risk and Uncertainty, 52*(3), 233–254.
Cheremukhin, A., Popova, A., & Tutino, A. (2015). A theory of discrete choice with information cost. *Journal of Economic Behavior & Organization, 113*, 34–50.
Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review, 121*(3), 463–480.
Dai, J. (2017). Are intertemporal preferences transitive? A Bayesian analysis of repeated individual intertemporal choices. *Decision, 4*(1), 1–24.
Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General, 143*(4), 1489–1514.
Dai, J., Pleskac, T. J., & Pachur, T. (2018). Dynamic cognitive models of intertemporal choice. *Cognitive Psychology, 104*, 29–56.
Denrell, J. C. (2015). Reference-dependent risk sensitivity as rational inference. *Psychological Review, 122*(3), 461–484.
Diamond, P., & Köszegi, B. (2003). Quasi-hyperbolic discounting and retirement. *Journal of Public Economics, 87*(9–10), 1839–1872.
Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic-dominance principle of decision making. *Psychological Science, 10*(4), 353–359.
Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and under-confidence: The role of error in judgment processes. *Psychological Review, 101*(3), 519–527.
Fiedler, K., & Unkelbach, C. (2014). Regressive judgment: Implications of a universal property of the empirical world. *Current Directions in Psychological Science, 23*(5), 361–367.
Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature, 40*(2), 351–401.
Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis.* Chapman and Hall/CRC.
Gollier, C., & Zeckhauser, R. (2005). Aggregation of heterogeneous time preferences. *Journal of Political Economy, 113*(4), 878–896.
Golman, R. (2011). Quantal response equilibria with heterogeneous agents. *Journal of Economic Theory, 146*(5), 2013–2028.
Golman, R. (2012). Homogeneity bias in models of discrete choice with bounded rationality. *Journal of Economic Behavior & Organization, 82*(1), 1–11.
Goodrich, B., Gabry, J., Ali, I., & Brilleman S. (2018). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.17.4. < http://mc-stan.org/ > .
Green, L., Fisher, E. B., Perlow, S., & Sherman, L. (1981). Preference reversal and self-control: Choice as a function of reward amount and delay. *Behaviour Analysis Letters, 1*(1), 43–51.
Green, L., Fristoe, N., & Myerson, J. (1994). Temporal discounting and preference reversals in choice between delayed outcomes. *Psychonomic Bulletin & Review, 1*(3), 383–389.
Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition, 25*(5), 715–723.
Gul, F., & Pesendorfer, W. (2006). Random expected utility. *Econometrica, 74*(1), 121–146.
Han, R., & Takahashi, T. (2012). Psychophysics of time perception and valuation in temporal discounting of gain and loss. *Physica A: Statistical Mechanics and Its Applications, 391*(24), 6568–6576.
Hilbert, M. (2012). Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making. *Psychological Bulletin, 138*(2), 211–237.
Howes, A., Warren, P. A., Farmer, G., El-Deredy, W., & Lewis, R. L. (2016). Why contextual preference reversals maximize expected value. *Psychological Review, 123*(4), 368–391.
Jackson, M. O. & Yariv, L. (2018). The non-existence of representative agents. Available at SSRN: < https://ssrn.com/abstract=2684776 > or < https://doi.org/10.2139/ssrn.2684776 > .
Jackson, M. O., & Yariv, L. (2014). Present bias and collective dynamic choice in the lab. *American Economic Review, 104*(12), 4184–4204.
Jackson, M. O., & Yariv, L. (2015). Collective dynamic choice: The necessity of time inconsistency. *American Economic Journal: Microeconomics, 7*(4), 150–178.
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*(5), 2513–2531.
Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292.
Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.
Killeen, P. R., & Taylor, T. J. (unpublished manuscript). Averaged exponential discount functions are hyperbolic: Implications for behavioral economics and theories of self-control.
Kinari, Y., Ohtake, F., & Tsutsui, Y. (2009). Time discounting: Declining impatience and interval effect. *Journal of Risk and Uncertainty, 39*(1), 87–112.
Kirby, K. N., & Herrnstein, R. J. (1995). Preference reversals due to myopic discounting of delayed reward. *Psychological Science, 6*(2), 83–89.
Kirby, K. N., & Maraković, N. N. (1995). Modelling myopic decisions: Evidence for hyperbolic delay-discounting within subjects and amounts. *Organizational Behavior and Human Decision Processes, 64*(1), 22–30.
Koopmans, T. C. (1960). Stationary ordinal utility and impatience. *Econometrica, 28*(2), 287–309.
Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press.
Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics, 112*(2), 443–477.
Lewandowsky, S., Freeman, M. C., & Mann, M. E. (2017). Harnessing the uncertainty monster: Putting quantitative constraints on the intergenerational social discount rate. *Global and Planetary Change, 156*, 155–166.
Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *Quarterly Journal of Economics, 107*(2), 573–597.
Loomes, G. (2005). Modelling the stochastic component of behaviour in experiments: Some issues for the interpretation of data. *Experimental Economics, 8*(4), 301–323.
Loomes, G., Moffatt, P. G., & Sugden, R. (2002). A microeconometric test of alternative stochastic theories of risky choice. *Journal of Risk and Uncertainty, 24*(2), 103–130.
Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39*, 641–648.
Loomes, G., & Sugden, R. (1998). Testing different stochastic specifications of risky choice. *Economica, 65*, 581–598.
Lu, J., & Saito, K. (2018). Random intertemporal choice. *Journal of Economic Theory, 177*, 780–815.
Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York, NY: Wiley.
Luce, R. D., & Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, & R. B. Bush (Eds.). *Handbook of mathematical psychology*. New York, NY: Wiley.

Luhmann, C. C. (2013). Discounting of delayed rewards is not hyperbolic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1274–1279.

Marley, A. A. J., & Regenwetter, M. (2017). Choice, preference, and utility: Probabilistic and deterministic representations. In W. Batchelder, H. Colonius, E. Dzhafarov, & J. Myung (Eds.). *New handbook of mathematical psychology. Volume 1: Measurement and methodology* (pp. 375–453). Cambridge: Cambridge University Press.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.). *The effect of delay and of intervening events on reinforcement value: Quantitative analyses of behavior* (pp. 55–73). Hillsdale, NJ: Lawrence Erlbaum.

McAlvanah, P. (2010). Subadditivity, patience, and utility: The effects of dividing time intervals. *Journal of Economic Behavior & Organization, 76*(2), 325–337.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.). *Frontiers in econometrics* (pp. 105–142). New York, NY: Academic Press.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics, 15*(5), 447–470.

Mellers, B. A., & Biagini, K. (1994). Similarity and choice. *Psychological Review, 101*(3), 505–518.

Millar, A., & Navarick, D. J. (1984). Self-control and choice in humans: Effects of video game playing as a positive reinforcer. *Learning and Motivation, 15*(2), 203–218.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... Petersen, S. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review, 89*(1), 103–124.

Peters, J., Miedl, S. F., & Büchel, C. (2012). Formal comparison of dual-parameter temporal discounting models in controls and pathological gamblers. *PLoS One, 7*(11), 1–12.

Phelps, E. S., & Pollak, R. A. (1968). On second-best national saving and game-equilibrium growth. *The Review of Economic Studies, 35*(2), 185–199.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review, 113*(1), 57–83.

R Core Team (2018). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing https://www.R-project.org/.

Raftery, A. E. (1996). Approximate Bayes factors and accounting for model uncertainty in generalised linear models. *Biometrika, 83*(2), 251–266.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347–356.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty, 23*(1), 5–32.

Read, D., McDonald, R., & He, L. (2018). Intertemporal choice: Choosing for the future. In A. Lewis (Ed.). *The Cambridge handbook of psychology and economic behavior* (pp. 167–197). New York, NY: Cambridge University Press.

Read, D., & Roelofsma, P. (2003). Subadditive versus hyperbolic discounting: A comparison of choice and matching. *Organizational Behavior and Human Decision Processes, 91*(2), 140–153.

Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H., & Stevens, J. R. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision, 5*(2), 63–94.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review, 118*(1), 42–56.

Regenwetter, M., Davis-Stober, C. P., Lim, S. H., Guo, Y., Popova, A., Zwilling, C., ... Messner, W. (2014). QTest: Quantitative testing of theories of binary choice. *Decision, 1*(1), 2–34.

Regenwetter, M., & Marley, A. A. J. (2001). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology, 45*(6), 864–912.

Regenwetter, M., & Robinson, M. M. (2017). The construct–behavior gap in behavioral decision research: A challenge beyond replicability. *Psychological Review, 124*(5), 533–550.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature, 44*(3), 631–661.

Rodriguez, C. A., Turner, B. M., & McClure, S. M. (2014). Intertemporal choice as discounted value accumulation. *PLoS One, 9*(2), e90138.

Rodriguez, C. A., Turner, B. M., Van Zandt, T., & McClure, S. M. (2015). The neural basis of value accumulation in intertemporal choice. *European Journal of Neuroscience, 42*(5), 2179–2189.

Samuelson, P. A. (1937). A note on measurement of utility. *The Review of Economic Studies, 4*(2), 155–161.

Sayman, S., & Öncüler, A. (2009). An investigation of time inconsistency. *Management Science, 55*(3), 470–482.

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review, 117*(3), 925–944.

Scholten, M., Read, D., & Sanborn, A. (2014). Weighing outcomes by time or against time? Evaluation rules in intertemporal choice. *Cognitive Science, 38*(3), 399–438.

Scholten, M., Read, D., & Sanborn, A. (2016). Cumulative weighing of time in intertemporal tradeoffs. *Journal of Experimental Psychology: General, 145*(9), 1177–1205.

Solnick, J. V., Kannenberg, C. H., Eckerman, D. A., & Waller, M. B. (1980). An experimental analysis of impulsivity and impulse control in humans. *Learning and Motivation, 11*(1), 61–77.

Sopher, B., & Sheth, A. (2006). A deeper look at hyperbolic discounting. *Theory and Decision, 60*(2), 219–255.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64*(3), 153–181.

Swait, J., & Marley, A. A. J. (2013). Probabilistic choice (models) as a result of balancing multiple goals. *Journal of Mathematical Psychology, 57*(1–2), 1–14.

Takahashi, T. (2005). Loss of self-control in intertemporal choice may be attributable to logarithmic time-perception. *Medical Hypotheses, 65*(4), 691–693.

Takahashi, T., Oono, H., & Radford, M. H. (2008). Psychophysics of time perception and intertemporal choice models. *Physica A: Statistical Mechanics and its Applications, 387*(8), 2066–2074.

Thaler, R. H. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters, 8*(3), 201–207.

Thurstone, L. L. (1959). *The measurement of values.* Chicago: University of Chicago Press.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making. *Proceedings of the National Academy of Sciences, 113*(11), 3102–3107.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review, 76*(1), 31–48.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Weitzman, M. L. (2001). Gamma discounting. *American Economic Review, 91*(1), 260–271.

Wilcox, N. T. (2008). Stochastic models for binary discrete choice under risk: A critical primer and econometric comparison. *Research in Experimental Economics, 12*, 197–292.

Wulff, D. U., & van den Bos, W. (2018). Modeling choices in delay discounting. *Psychological Science, 29*(11), 1890–1894.

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*(4), 543–556.

Zhao, W. J., Diederich, A., Trueblood, J. S., & Bhatia, S. (2019). Automatic biases in intertemporal choices. *Psychonomic Bulletin and Review* (in press).