Psychological Review

The Unified Tradeoff Model

Marc Scholten, Daniel J. Walters, Craig R. Fox, and Daniel Read

Online First Publication, March 21, 2024. https://dx.doi.org/10.1037/rev0000458

CITATION

Scholten, M., Walters, D. J., Fox, C. R., & Read, D. (2024). The unified tradeoff model.. *Psychological Review*. Advance online publication. https://dx.doi.org/10.1037/rev0000458

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

https://doi.org/10.1037/rev0000458

The Unified Tradeoff Model

Marc Scholten[1, 2], Daniel J. Walters[3], Craig R. Fox[4], and Daniel Read[5]

[1] Faculdade de Design, Tecnologia e Comunicação, Universidade Europeia

[2] Centro de Estudos em Gestão do Instituto Superior Técnico, Instituto Superior Técnico, Universidade de Lisboa

[3] Institut Européen d'Administration des Affaires

[4] Anderson School of Management, University of California, Los Angeles

[5] Warwick Business School, University of Warwick

Evidence is steadily mounting that attribute-based models offer a more accurate description of intertemporal choices than traditional alternative-based models. Among the attribute-based models, the tradeoff model offers the broadest coverage of research findings, but at the cost of considerable complexity: There now are various instantiations of the model dealing with partially overlapping universes of choice options and preference patterns. Moreover, there are reports of preference patterns in intertemporal decisions about monetary losses that contradict all attribute-based models proposed so far. Taking stock of these core challenges, and all other evidence, we develop an account of intertemporal choice, the unified tradeoff model, that is simpler, yet more comprehensive, than all currently available versions of the tradeoff model taken together. It borrows extensively from its predecessors, but it introduces a new element, *time bias*, that enables it to accommodate an extraordinarily broad range of preference patterns, and also generate new predictions that contradict all existing models of intertemporal choice. We report four studies that test and confirm its predictions regarding delay, interval, sign, and magnitude dependence in choices between single-dated outcomes, and a fifth study that tests and confirms its predictions regarding the relation between delay preference in choices that only involve single-dated payments and duration preference in choices that also involve sequences of payments. Having subjected the unified tradeoff model to an elevated risk of disconfirmation, we discuss its parsimony and scope in relation to yet other phenomena, most notably, preference patterns in consumption decisions, the final frontier for attribute-based models.

Keywords: intertemporal tradeoffs, time preference, and time bias; delay, interval, sign, and magnitude dependence; single-dated outcomes and outcome sequences; money and consumption; parsimony and scope

Most models of choice can be characterized as either alternative-based or attribute-based (Payne et al., 1988). In alternative-based models, each option is assigned an overall value as a function of its features, and the option with the highest overall value is chosen. In attribute-based models, the options are directly compared along their attributes, and the option that compares best with the other options is chosen. Models with risky/uncertain or delayed outcomes tend to be alternative-based, building on the notions of expected value (Bernoulli, 1954) and discounted value (Fisher, 1930) from economics and finance.

Although they have been less popular, attribute-based approaches also have a long history in the field of decision making, most notably in the domain of risky choice. For instance, Savage (1951) introduced the minimax regret criterion, in which an economic agent identifies the option that performs best under each possible state of the world, and chooses the option with the smallest

Marc Scholten https://orcid.org/0000-0003-1868-5928
Daniel J. Walters https://orcid.org/0000-0002-0121-7178
Craig R. Fox https://orcid.org/0000-0001-9057-112X
Daniel Read https://orcid.org/0000-0001-6672-9310

Marc Scholten received support from Grant PTDC/MHC-PCN/3805/2012 from the Fundação para a Ciência e Tecnologia, and Daniel Read received funding from Grant ES/K002201/1 from the Economic and Social Research Council and Grant RP2012-V-022 from the Leverhulme Trust.

The Pay Battery (used in Studies 1 and 5) and the first results on sign and magnitude dependence (reported as Study 3) were presented at the Network for Integrated Behavioural Science Workshop in 2015 (University of Nottingham, Nottingham, United Kingdom), and at the Foundations and Applications of Utility, Risk and Decision Theory Conference in 2016 (University of Warwick, Coventry, United Kingdom).

For survey contents, raw data, and a Web Appendix, see the Unified Tradeoff Model Project on the Open Science Framework Platform (https://osf.io/ga2bh/?view_only=03f8fbf726724cc08567364a62cfb4bb). The design, hypotheses, and analysis plan of the Study 2 were preregistered; for the preregistration, see Study 2 on the AsPredicted Platform (https://aspredicted.org/blind.php?x=91N_BYJ). The other studies were not preregistered.

Marc Scholten and Daniel Read contributed to funding acquisition. Marc Scholten, Daniel J. Walters, Craig R. Fox, and Daniel Read contributed to conceptualization, formal analysis, data collection and analysis, and writing–review and editing.

Correspondence concerning this article should be addressed to Marc Scholten, Faculdade de Design, Tecnologia e Comunicação, Universidade Europeia, Quinta do Bom Nome, Estrada da Correia 53, 1500-210 Lisbon, Portugal. Email: marc.scholten@universidadeeuropeia.pt

deviation from these best performances across all possible states. Similar notions were later incorporated into regret theory (Bell, 1982; Loomes & Sugden, 1982). Even original prospect theory (Kahneman & Tversky, 1979), which is primarily an alternative-based model, included attribute-based "simplification" operations as part of a preliminary "editing" stage, including elimination of dominated options, and elimination of small outcome or probability differences between the options. Fully attribute-based models of risky choice have been developed in the years since (e.g., González-Vallejo, 2002; Loomes, 2010). The first attribute-based model of intertemporal choice appeared much later (Scholten & Read, 2010), but, since its introduction, there has been a rapid development of attribute-based approaches to intertemporal choice (e.g., Cheng & González-Vallejo, 2016; Ericson et al., 2015; Read et al., 2013), and risky intertemporal choice (Luckman et al., 2020).

We will focus on intertemporal choice. Most experimental research in this domain has been restricted to rudimentary choices between pairs of single-dated monetary outcomes such as whether to receive $20 today or $40 in 1 year. These choices between a *S*maller amount *S*ooner (*SS*) and a *L*arger amount *L*ater (*LL*) are held to retain core features of real-world intertemporal decisions, such as whether to stay up late to watch a movie or wake up refreshed for work in the morning, or whether to spend money on a vacation today or save the money for a down payment on a future home. These decisions are usually conceived as a tradeoff between being better off sooner and being better off later. Moreover, even though the *SS*–*LL* paradigm may omit many features of the naturalistic decisions it is designed to represent, it still has exposed a large number of preference patterns that systematically violate the normative theory of investment decisions from economics, which holds that decision makers should maximize net present value given the prevailing interest rates they face (Fisher, 1930; Hirshleifer, 1958).

In response to these violations of the normative theory, there has been a proliferation of descriptive models, which have mostly taken an alternative-based approach to intertemporal choice (e.g., Loewenstein & Prelec, 1992). As it turns out, however, the attribute-based approach has consistently been shown to provide a more accurate description of intertemporal choice than the alternative-based approach, both qualitatively and quantitatively (Arfer & Luhmann, 2015; Cheng & González-Vallejo, 2016; Dai & Busemeyer, 2014; Ericson et al., 2015; Leland, 2002; Read et al., 2013; Roelofsma & Read, 2000; Rubinstein, 2003; Scholten & Read, 2010, 2013; Scholten et al., 2014, 2016).[1] Of course, models within either class differ in accuracy, substance, scope, and parsimony. We will place the emphasis on the scope and parsimony of attribute-based models.

Currently, the tradeoff model (Scholten & Read, 2010) offers the broadest coverage of reported preference patterns, but its scope has come at the cost of substantial complexity: There are now various instantiations of the model dealing with partially overlapping universes of choice options and preference patterns (Read & Scholten, 2012; Scholten & Read, 2010, 2013; Scholten et al., 2014, 2016). For instance, some instantiations only address choices between single-dated outcomes, more precisely, *SS*–*LL* choices (Scholten & Read, 2010, 2013; Scholten et al., 2014), while other instantiations also address choices involving outcome sequences (Read & Scholten, 2012; Scholten et al., 2016); yet, the models defined over single-dated outcomes are not special cases of the models defined over outcome sequences, even though single-dated outcomes are single-period outcome sequences. Moreover, when the universe of option pairs includes not only *SS*–*LL* choices, but also choices between a *Larger* amount *S*ooner (*LS*) and a *S*maller amount *L*ater (*SL*), as was the case in Hardisty et al.'s (2013) studies, intertemporal decisions about monetary losses exhibit preference patterns that contradict all attribute-based models proposed so far.

Taking stock of the core challenges, and all other evidence, we develop an account of intertemporal choice, the unified tradeoff model, that is simpler, yet more comprehensive, than all currently available versions of the tradeoff model taken together. An element that the unified tradeoff model and its predecessors share with most other models of intertemporal choice is *positive time preference* (Fisher, 1930), meaning a preference for obtaining positive outcomes as soon as possible, and incurring negative outcomes as late as possible. The unified tradeoff model, however, introduces a new element, *time bias*, that enables it to accommodate an extraordinarily broad range of preference patterns, and to generate new predictions that contradict all other models of intertemporal choice, attribute- and alternative-based. We put these distinctive predictions to the test, and see them confirmed.

The exposition is structured as follows. We first provide an intuitive description of time bias, as a prelude to the formal development of the unified tradeoff model. Initially, we develop the model for choices that only involve single-dated outcomes and derive testable implications from it, which we see confirmed in four studies. Subsequently, we develop the model for choices that also involve outcome sequences, and derive testable implications from it, which we see confirmed in a fifth study. Having subjected the unified tradeoff model to an elevated risk of disconfirmation, we discuss its parsimony and scope in relation to yet other phenomena, most notably, preference patterns in consumption decisions, that is, decisions about when to consume, and when to pay for consumption. Such decisions are of key interest to economists, but have so far been inaccessible to attribute-based models of intertemporal choice. Although our analysis is qualitative in nature, we close with a discussion of issues surrounding the specification of the unified tradeoff model for quantitative analyses, and the consequent need to replace scope by generalizability as the ultimate criterion of model validity.

Time Bias: An Intuitive Description

To capture the intuition of time bias, suppose you learn that a concert of your favorite rock band has been scheduled, but will only

[1] Process-tracing studies tend to encounter both intra– and inter-individual variability in whether information about the options is processed attribute- or alternative-wise, although attribute-based processing tends to be more common (Amasino et al., 2019; Arieli et al., 2011; Liu et al., 2021; Reeck et al., 2017). There appears to be widespread consensus in this research area that attribute- and alternative-based *processing* and *choice* are intimately connected, but that this is yet to be demonstrated. It is legitimate to have reservations about it. First, comparing amounts or delays directly is cognitively less demanding than combining amounts with delays, which may lead even those whose choices are alternative-based to wander off into attribute-wise processing. Second, people ultimately choose an option, meaning a given amount at a given delay, which may lead even those whose choices are attribute-based to perceive each option as a combination of its features.

be in 1 year. At first, you might think "oh no, I have to wait!" This primary response does not take the length of the wait into account. But then you might go on ruminating, thinking "and I even have to wait a whole year!" At that point, time length *is* taken into account. In our interpretation, the first response stands for time bias, the second for time preference. Time bias can be positive, in which case it reinforces positive time preference in contributing to a preference for good things sooner and bad things later, or it can be negative, in which case it countervails positive time preference in contributing to a preference for good things later and bad things sooner. It is particularly in reaction to bad events that negative time bias is no less common than positive time bias. For instance, if you learn that there is a need for a medical exam, and that it has been scheduled for 1 month from now, you might be thinking "at least it is not now" (positive time bias), or you might be asking yourself "can't we get it over with?" (negative time bias).

To clarify the distinction between time preference and time bias in more formal terms, consider the decision of whether to receive \$20 now or some amount later. Call the later amount \$$x$. The indifference point is where the value of the delayed \$$x$ equals the value of the immediate \$20. Positive time preference means that receiving \$20 now has a time advantage, the magnitude of which increases as a function of how much later \$$x$ is to be received. Positive time bias means that receiving \$20 now has an additional time advantage, the magnitude of which does not depend on how much later \$$x$ is to be received. At the indifference point, \$$x$ will equal \$20 *plus* the variable effect of positive time preference *plus* the fixed effect of positive time bias.

Time preference and time bias are theoretical constructs, to be distinguished from *delay preference*, an empirical observation driven by multiple factors, including time preference and time bias.[2] When the combined effect of these factors is to render longer delays undesirable, we refer to the observed delay preference as *delay aversion*; when their combined effect is to render longer delays desirable, we refer to the observed delay preference as *delay tolerance*.

We already mentioned that it is particularly in reaction to bad events that negative time bias is no less common than positive time bias. Specifically, studies on choices between different timings of the same monetary outcome have found that participants are virtually unanimous in their delay aversion for gains, but that they are divided in their delay preference for losses, with delay tolerance and delay aversion occurring in (almost) equal proportion (Stevens, 2016; Yates & Watts, 1975).

Consider the decision of whether to pay \$20 now or \$$x$ later. Among decision makers who combine positive time preference with positive time bias, the indifference point will be reached when \$$x$ equals \$20 *plus* the variable effect of time preference *plus* the fixed effect of time bias. Thus, they will be indifferent between paying \$20 now and paying more later. Among decision makers who combine positive time preference with negative time bias, however, the indifference point will be reached when \$$x$ equals \$20 *plus* the variable effect of time preference *minus* the fixed effect of time bias. With positive time preference and negative time bias going in opposite directions, the indifference point is less definite: If positive time preference outweighs negative time bias, these decision makers will be indifferent between paying \$20 now and paying more later, but, if positive time preference *is* outweighed by negative time bias, they will be indifferent between paying \$20 and paying *less* later. We will show that the combination of unanimity in positive time bias for gains and diversity in the sign of time bias for losses is a pervasive force behind preference patterns in intertemporal choice.

Parenthetically, the presumed *existence* of a bias does not imply any particular *explanation* of it (Keren & Teigen, 2004). The unified tradeoff model is agnostic about the causes of time bias, as it is about the causes of time preference (a rich source of speculation throughout the history of economic thought; see Loewenstein, 1992), or indeed about any of its components. For instance, negative time bias for monetary losses may expose debt aversion (Myerson et al., 2017), but people exhibit delay aversion for unpleasant *consumption* events as well (Loewenstein, 1987), which must be rooted in something else. More generally, delay aversion for an unpleasant event may be motivated by reducing cognitive demand (Rosenbaum et al., 2019), to "not have to think about it anymore," or averting dread (Hardisty & Weber, 2020; Loewenstein, 1987), to "not have it hanging over your head." Delay *tolerance* for an unpleasant event would indicate that such motives are not strong enough to outweigh positive time preference. Conversely, delay tolerance for a *pleasant* event would indicate that a desire to savor it outweighs positive time preference (Hardisty & Weber, 2020; Loewenstein, 1987), but this rarely happens, and, in the domain of monetary gains, delay aversion is ubiquitous. Anyhow, our concern is with the formal modeling of intertemporal choice, although we do convey the intuition behind every component of the unified tradeoff model when it is introduced.

Motivation and Development: Single-Dated Outcomes

The simplest plausible instantiation of the tradeoff model for choices that only involve single-dated outcomes, which we refer to as the baseline tradeoff model (for applications, see Ericson et al., 2015; He et al., 2022; Scholten & Read, 2013; Wulff & van den Bos, 2018), is motivated by four classic preference patterns that are incompatible with the normative theory from economics, which holds that investment decisions should be determined by prevailing interest rates, and be independent from consumption preferences (Fisher, 1930; Hirshleifer, 1958). We call the preference patterns "classic," not only because they were the first to be documented in experimental research (Thaler, 1981), but also because they have had tremendous impact on the descriptive modeling of intertemporal choice. However, the models developed in response to these classic preference patterns must now contend with Hardisty et al.'s (2013) core challenges, which will be key considerations in the development of the unified tradeoff model. Below, we develop our model for choices between single-dated outcomes, in five stages, each motivated by empirical findings.

[2] Hardisty et al. (2013) use the terms "pure time preference" and "time preference" for what we refer to as time preference and delay preference, respectively, while Frederick et al. (2002) use the terms "time preference" and "time discounting." The use of "delay preference" as designating an empirical observation can be traced back to early experimental research on intertemporal and risky choice (Mischel & Metzner, 1962), and appears to be reemerging (Luckman et al., 2017).

Motivation 1: Violations of Sign, Magnitude, and Delay Independence

A foundational construct in the normative analysis of intertemporal choice is positive time preference, which, as previously pointed out, generates delay aversion for gains (sooner is better), and delay tolerance for losses (later is better). Most evidence on intertemporal choice has been gathered on the presumption that people do exhibit these delay preferences, and depart from the normative theory in other respects.

Consider a choice between a smaller outcome, x_S, due after a shorter delay, t_S, and a larger outcome, x_L, due after a longer delay, t_L, and suppose that an individual is indifferent between these single-dated outcomes, that is, $(x_L, t_L) \sim (x_S, t_S)$. From the indifference point, we can compute the interest rate implied by the options, hereafter, the *implied interest rate* (Cohen et al., 2020):

$$\tilde{\imath} = \ln\left(\frac{x_L}{x_S}\right)^{1/(t_L - t_S)} = \frac{1}{t_L - t_S}(\ln|x_L| - \ln|x_S|). \quad (1)$$

Under the normative theory, $\tilde{\imath}$ should be constant; more precisely, it should be equal to the decision makers' best borrowing or savings rate, depending on whether they are currently a debtor or a creditor (Fisher, 1930; Hirshleifer, 1958; see Read & Scholten, 2017, for a compact explanation). A constant $\tilde{\imath}$, regardless of whether it corresponds to the market rate of interest or not, is widely adopted as the *benchmark model* in experimental research (Cohen et al., 2020). Four anomalies have dominated research on the constancy or nonconstancy of implied interest rates.

1. *The absolute magnitude effect*. A decision maker who is indifferent between a sooner and a later outcome will, when both outcomes are increased by a common multiplicative constant, strictly prefer the later gain (a reduced delay aversion) or the sooner loss (a reduced delay tolerance). For instance, someone who is indifferent between \$100 today and \$200 in 1 year will prefer \$2,000 in 1 year to \$1,000 today, and can, therefore, only be indifferent between \$1,000 today and *less than* \$2,000 in 1 year. This pattern violates the normative theory, which entails *magnitude independence*, because the multiplicative constant does not change the ratio between the outcomes (x_L/x_S), and, therefore, does not change the interest rate implied by the options.

2. *The gain-loss asymmetry*. A decision maker who is indifferent between a sooner and a later gain will, when outcome magnitude is held constant, strictly prefer the sooner loss. Stated differently, delay tolerance for losses is weaker than delay aversion for gains. For instance, someone who is indifferent between receiving \$100 today and receiving \$200 in 1 year will prefer paying \$100 today to paying \$200 in 1 year, and can, therefore, only be indifferent between paying \$100 today and paying *less than* \$200 in 1 year. This pattern violates the normative theory, which entails *sign independence*, because the change of outcome sign does not change the ratio between the outcomes (x_L/x_S), and, therefore, does not change the interest rate implied by the options.

3. *The sign-magnitude asymmetry*. The gain-loss asymmetry is more pronounced for smaller outcomes than for larger ones, or, equivalently, the absolute magnitude effect is more pronounced for gains than for losses. The indifference points in Table 1 illustrate this anomaly: As outcome magnitude increases by a factor of 10, the implied interest rate decreases by 0.511 for gains, but by only 0.310 for losses; equivalently, as outcome sign changes from positive to negative, it decreases by 0.288 for small outcomes, but by only 0.087 for large ones.

4. *The common difference effect*. A decision maker who is indifferent between a sooner and a later outcome will, when the delays to both outcomes are increased by a common additive constant, strictly prefer the later gain (a reduced delay aversion) or the sooner loss (a reduced delay tolerance). For instance, someone who is indifferent between \$100 today and \$200 in 1 year will prefer \$200 in 11 years to \$100 in 10 years.[3] This pattern violates the normative theory, which implies *delay independence*, because the additive constant does not change the difference between the delays to the outcomes ($t_L - t_S$), and, therefore, does not change the interest rate implied by the options.

The baseline tradeoff model accommodates the four classic anomalies. Given a choice between a sooner and a later outcome, it proposes that the "effective differences" (Scholten & Read, 2010) between the options along the time and outcome attributes are traded off against one another, and that the option favored by the tradeoff is chosen. The *effective outcome difference*, *X*, is the difference between the values of the later and the sooner outcome, that is,

$$X = v(x_L) - v(x_S), \quad (2)$$

where either $x_L, x_S > 0$ or $x_L, x_S < 0$, and the value function v assigns a value to each outcome. The *effective time difference*, *T*, is the difference between the weights of the longer and shorter delay, that is,

$$T = w(t_L) - w(t_S), \quad (3)$$

where $t_L > t_S \geq 0$, and the time-weighing function w assigns a weight to each delay. At the point of indifference between the sooner and the later outcome,

$$\kappa\sigma T = X. \quad (4)$$

This expression has two elements that combine to create positive time preference. One element, $\kappa > 0$, is the strength of time preference, which, in the tradeoff model, is the rate at which effective time differences are traded off against effective outcome differences. The other element, σ, is an indicator for outcome sign. When the outcomes are of positive sign ($x_L, x_S > 0$), the indicator is set to $\sigma = 1$, in which case the sooner gain has the time advantage, so that the later gain must be larger than the sooner one ($x_L > x_S > 0$) for an indifference point to exist. Analogously, when the outcomes are of negative sign ($x_L, x_S < 0$), the indicator is set to $\sigma = -1$, in which case the sooner loss has a time disadvantage, so that the later

[3] Although it is widely held to be a core anomaly to the normative theory, the evidence on the common-difference effect is mixed, even in the restricted domain of monetary gains. We turn to this issue later.

Table 1
The Sign-Magnitude Asymmetry, and an Interpretation

	Outcome sign	
Outcome magnitude	Positive	Negative
Small	(\$100, 0) ~ (\$200, 1) $\tilde{t} = 0.693$ $v(x) = x$ $\kappa T = 200 - 100$	(−\$100, 0) ~ (−\$150, 1) $\tilde{t} = 0.405$ $v(-x) = 2(-x)$ $-\kappa T = 2(-150) - 2(-100)$
Large	(\$1,000, 0) ~ (\$1,200, 1) $\tilde{t} = 0.182$ $v(x) = 0.5x$ $\kappa T = 0.5(1{,}200) - 0.5(1{,}000)$	(−\$1,000, 0) ~ (−\$1,100, 1) $\tilde{t} = 0.095$ $v(-x) = -x$ $-\kappa T = -1{,}100 - (-1{,}000)$

Note. A simplified interpretation of the sign-magnitude asymmetry offered by the baseline tradeoff model: Losses are twice as impactful as gains, large amounts are half as impactful small amounts, and absolute sensitivity to outcomes is otherwise constant.

loss must be larger than the sooner one ($x_L < x_S < 0$) for an indifference point to exist.

The need for the σ-indicator reveals a deeper property of the attribute-based approach to intertemporal choice, which is that time has value, just as outcome does (Scholten et al., 2014). Specifically, under positive time preference, gaining sooner and losing later have positive value, or, *equivalently*, gaining later and losing sooner have negative value. The time-weighing function and the value function play a pivotal role in the construction of time and outcome value. Drawing on prospect theory (Tversky & Kahneman, 1991), and the original formulation of the tradeoff model (Scholten & Read, 2010), the baseline tradeoff model identifies four properties of the value function.

1. *Reference dependence*. Outcomes are evaluated as positive deviations (gains) and negative deviations (losses) from a reference point, typically one's current wealth, that is, $v(0) = 0$.

2. *Diminishing absolute sensitivity*. As outcome magnitude increases by constant absolute amounts, positive outcome value increases, and negative outcome value decreases, by decreasing absolute amounts, that is, $v''(x) \lesseqgtr 0$ for $x \gtreqless 0$. For instance, increasing \$100 by \$1 leads to a smaller increase in outcome value than increasing \$10 × \$1, that is, $v(101) - v(100) < v(11) - v(10)$. This property of the value function is indispensable for the tradeoff model to acquire a sense of proportionality (Scholten & Read, 2010): Under constant absolute sensitivity to outcomes, someone indifferent between \$100 today \$200 in 1 year would also be indifferent between \$10,100 today and \$10,200 in 1 year, a highly counterintuitive implication.

3. *Augmenting proportional sensitivity*. As outcome magnitude increases by constant proportional amounts, outcome value increases by increasing absolute amounts, that is, $(xv'(x))' > 0$. For instance, doubling \$100 leads to a larger increase in outcome value than doubling \$10, that is, $v(200) - v(100) > v(20) - v(10)$. This accounts for the absolute magnitude effect.

4. *Loss aversion*. Under the tradeoff model, loss aversion means that decision maker exhibits greater proportional sensitivity to losses than to gains, that is, $(-xv'(-x))' > (xv'(x))'$ for $x > 0$ (see Abdellaoui et al., 2007, for alternative definitions of loss aversion under prospect theory). For instance, doubling −\$100 leads to a larger absolute increase in outcome value than doubling \$100, that is, $v(-100) - v(-200) > v(200) - v(100)$. This accounts for the gain-loss asymmetry.

The properties of the value function also interact, and the sign-magnitude asymmetry is an indication that augmenting proportional sensitivity and loss aversion interact with diminishing absolute sensitivity. Table 1 provides a simplified illustration. Losses are twice as impactful as gains (*constant loss aversion*; see Tversky & Kahneman, 1991), and large amounts are half as impactful as small amounts (a discontinuous drop in absolute sensitivity to outcomes), while the decision maker otherwise exhibits constant absolute sensitivity, and, therefore, augmenting proportional sensitivity, to outcomes.[4] The four indifference points result from this specification.

The first three properties of the value function are shared by the time-weighing function: (a) Reference dependence, that is, $w(0) = 0$, (b) diminishing absolute sensitivity, that is, $w''(t) < 0$, and (c) augmenting proportional sensitivity, that is, $(tw'(t))' > 0$. It is diminishing absolute sensitivity that accounts for the common-difference effect. The value function and the time-weighing function also share the more basic property of *monotonicity*, meaning that both are increasing in their argument, that is, $v'(x) > 0$ and $w'(t) > 0$.

Importantly, available evidence on the classic preference patterns comes from studies in which researchers assumed that participants would act on positive time preference, and, therefore, felt there was no reason to even ask participants to decide whether to pay \$100 today or pay \$100 or less in 1 year. Thus, participants were not given an opportunity to express delay tolerance in gains or delay aversion in losses, meaning that most evidence on the classic preference patterns is distorted. This includes research on sign independence (Abdellaoui et al., 2010, 2013; Baker et al., 2003; Breuer et al., 2023; Estle et al., 2006; McAlvanah, 2010; Myerson et al., 2017; Read et al., 2017; Scholten & Read, 2013; Scholten et al., 2014), magnitude independence (Baker et al., 2003; Breuer et al.,

[4] If $v(x) = x$, then $mx_L - mx_S > x_L - x_S$ for $m > 1$ and $x_L > x_S$.

2023; Estle et al., 2006; Holt et al., 2008; McAlvanah, 2010; Myerson et al., 2017; Scholten & Read, 2013; Scholten et al., 2014), and delay independence (Holt et al., 2008; McAlvanah, 2010; Scholten et al., 2014). The very few studies that *did* allow participants to express delay preferences of either sign for outcomes of either sign, notably those conducted by Hardisty et al. (2013), promptly exposed anomalies that motivate the next two stages of model development.

Motivation 2: Universal Delay Aversion in Gains, Diversity of Delay Preference in Losses

As mentioned earlier, people are virtually unanimous in their delay aversion for monetary gains, but they are divided in their delay preference for monetary losses: Some prefer to pay later (delay tolerance), as they should under normative theory, the baseline tradeoff model, and most other models of intertemporal choice, but others prefer to pay sooner (delay aversion).

Universal delay aversion in monetary gains means that people want to be compensated for receiving money later, that is, to receive *more* later. When the compensation is *just* enough, they are at the *SS–LL* indifference point. Hardisty et al. (2013) indeed found that *SS–LL* indifference points were ubiquitous in gains, even though their participants could choose to receive *less* later.

Diversity of delay preferences in monetary losses means that some people want to be compensated for paying money *sooner* (delay tolerance), while others want to be compensated paying money *later* (delay aversion). When the compensation is *just* enough, the delay tolerant are at the *SS–LL* indifference point, whereas the delay averse are at an *LS–SL* indifference point. Hardisty et al. (2013) indeed found that *LS–SL* indifference points were very common in losses.

Table 2 provides illustrations of unconventional preference patterns in indifference points, which will serve as a reference throughout the rest of this article. Some rows in the table carry the letter E, meaning that the phenomenon in question lies in the *existence* of an indifference point, in this case, the existence of an *LS–SL* indifference point (rows 1, 2, 3, and 5). All other rows carry the letter C, meaning that the phenomenon in question lies in a *comparison* between indifference points.

One possible interpretation of Hardisty et al.'s (2013) evidence is that the sign of time preference corresponds to the sign of delay preference. Thus, Equation 4 describes decision makers who exhibit *SS–LL* indifference points in both gains and losses, whereas

$$\kappa T = X, \tag{5}$$

describes those who exhibit *SS–LL* indifference points in gains, but *LS–SL* indifference points in losses. Another interpretation, the one

Table 2
Illustration of Unconventional Preference Patterns in Intertemporal Choice

					Relative nonadditivity					
No.	Delay aversion for losses	Reverse magnitude effect	Magnitude-induced reversal of delay preference	Delay-induced reversal of delay preference	Time relative to outcome sign	Time relative to outcome magnitude	Choice triplet	t (years) 0	1	2
1	E	C	C					−$20.00	−$12.50	
2	E	C			C		Small losses		−$12.50	−$3.60
3	E	C	C					−$20.00		−$16.20
4		C	C	C				−$100.00	−$100.90	
5	E	C	C	C			Medium losses		−$100.90	−$98.30
6		C	C					−$100.00		−$117.40
7		C						−$500.00	−$542.90	
8		C	C				Large losses		−$542.90	−$572.10
9		C						−$500.00		−$623.30
10								$500.00	$621.40	
11						C	Large gains		$621.40	$724.20
12								$500.00		$812.00
13								$20.00	$35.20	
14					C	C	Small gains		$35.20	$50.90
15								$20.00		$44.00

Note. Preference patterns in indifference points between (x_S, t_S) and ($\tilde{x}_L$, t_L) generated by the unified tradeoff model under bipolar time bias, where E indicates that the phenomenon is one of existence (within rows), and C indicates that the phenomenon is one of comparison (between rows or triplets of rows).

advocated by the unified tradeoff model, is that (a) time preference is of universally positive sign across decision makers and decision outcomes, and consistent across the circumstances under which decisions are made, but (b) time preference operates in combination with time bias, the sign of which is universally positive for positive outcomes, but contingent on the decision maker, and the circumstances under which decisions are made, for negative ones.

Our approach bears resemblance to Loewenstein's (1987) model of delay preferences for consumption events, in which positive time preference operates on utility from consumption, whereas the decision maker's preoccupation with the future, and the vividness of a particular consumption event, promote utility from anticipation. These factors, both extraneous to time preference, and both also under the influence of extraneous factors, contribute to delay aversion for unpleasant events, and delay tolerance for pleasant ones (in case of short delays to vivid and fleeting consumption events). Analogously, the unified tradeoff model proposes that time bias may create an *appearance* of negative time preference.

Formally, the difference between time preference and time bias is that the former operates as a multiplicative constant that scales with time, while the latter operates as an additive constant that, therefore, does not scale with time. Actually, most attribute-based models of intertemporal choice developed since the introduction of the tradeoff model include an additive constant, or *bias parameter* (Luckman et al., 2020; Wulff & van den Bos, 2018), intended to capture person- and situation-specific influences on the overall preference between the available options (Cheng & González-Vallejo, 2016; Ericson et al., 2015; Read et al., 2013).[5] In our qualitative analysis of intertemporal choice, the sign of time bias is under the influence of person- and situation-specific factors.

The unanimity with which people exhibit delay aversion in monetary gains suggests that any time bias will strengthen the delay aversion promoted by positive time preference. Thus, it is a bias in favor of gaining sooner, or, equivalently, against gaining later. If we let b be time bias, the decision maker will be indifferent between a sooner and a later gain when

$$\kappa T + b = v(x_L) - v(x_S), \text{ or} \tag{6}$$

$$\kappa T = [v(x_L) - b] - v(x_S), \tag{7}$$

where $b > 0$ is positive time bias, and $x_L > x_S > 0$ is delay aversion. As just mentioned, positive time bias for gains can be conceived in two equivalent ways: As a bias in favor of gaining sooner, which is made explicit by treating b as a positive factor on the left-hand side of the equation (adding to the time advantage of *SS*), or as a bias against gaining later, which is made explicit by treating b as a negative factor on the right-hand side of the equation (subtracting from the outcome advantage of *LL*). As b becomes larger, x_L must become larger for indifference to be preserved, thus yielding a stronger delay aversion. A third description of the indifference point arises when we recall ourselves that the value of time is expressed in the same unit as the value of outcomes, so that time bias may be explicitly written as the value assigned to a monetary amount, that is,

$$\kappa T = [v(x_L) - v(c)] - v(x_S), \tag{8}$$

where $c > 0$. Since $v(c) = b$, or, conversely, $v^{-1}(b) = c$, Equation 8 is formally equivalent to the previous descriptions of the indifference point.

If the assumption of universally positive time preference is to be maintained, the diversity of delay preferences in losses *must* be a manifestation of a corresponding diversity in time bias for losses. We therefore introduce a structural condition to the operation of time bias, which is *unanimity in positive time bias for gains, but diversity in the sign of time bias for losses*. Under this condition, there are two groups of decision makers.

One group is characterized by *unipolar time bias*. In this group, time bias operates in the same direction for outcomes of opposite sign, reducing the absolute value of the later gain or the later loss. This favors the later loss, or, equivalently, disfavors the sooner one, that is,

$$-\kappa T = [v(x_L) + v(c)] - v(x_S), \tag{9}$$

where $c > 0$ is positive time bias, and $x_L < x_S < 0$ is delay tolerance. As c becomes larger, x_L must become larger for indifference to be preserved, thus yielding stronger delay tolerance.

Another group of decision makers is characterized by *bipolar time bias*. In this group, time bias operates in opposite directions for outcomes of opposite sign, reducing the absolute value of the later gain, but enhancing the absolute value of the later loss. This disfavors the later loss, or, equivalently, favors the sooner one, that is,

$$-\kappa T = [v(x_L) - v(c)] - v(x_S), \tag{10}$$

where $c > 0$ is negative time bias. Rearranging,

$$\phi = -\kappa T + v(c) = v(x_L) - v(x_S), \tag{11}$$

where ϕ is the time advantage or disadvantage of the sooner loss. When positive time preference outweighs negative time bias, the sooner loss has a time disadvantage, that is, $\phi < 0$, so it must have an outcome advantage for the indifference point to exist, meaning that it must be smaller than the later loss, that is, $x_L < x_S < 0$, which is delay tolerance. Conversely, when positive time preference is outweighed by negative time bias, the sooner loss has a time advantage, that is, $\phi > 0$, introducing two conditions for the indifference point to exist: First, the sooner loss must have an outcome disadvantage, meaning that it must be larger than the later loss, that is, $x_S < x_L < 0$, which is delay aversion, and, second, the time advantage of incurring the loss sooner must be smaller than the relief from foregoing the loss altogether, that is, $\phi < -v(x_S)$.

Time preference scales with T, meaning that its impact increases as T increases. There are two ways of doing this. One way is by increasing t_L, while holding t_S constant, in which case the interval separating the outcomes objectively increases, and T increases due to the monotonicity of the time-weighing function w. Thus, someone may be delay averse over a short interval, but delay tolerant over

[5] From a measurement perspective, it is quite sensible to treat an additive constant as a bias. To illustrate, consider balance K, which indicates weight in kilograms, and balances G and B, which indicate weight in grams. Whatever number K shows, G shows a number that is 1,000 times larger, while B shows a number that, only when 100 g is subtracted from it first, equals the number shown by G. The conclusion would be that balance B is off by 100 g, constituting a measurement bias.

longer ones, and the strength of delay tolerance should increase as the length of the interval continues to increase. Indications are that it does. For instance, in Thaler's (1981) study, median implied interest rates were very low when losses were separated by 3 months, but notably higher when the interval between the payments increased to 1 or 3 years. The other way of increasing T is by decreasing t_L and t_S, while holding $t_L - t_S$ constant, in which case T increases due to the concavity of the time-weighing function w. Thus, someone may be delay averse when both outcomes are remote, but then shift to delay tolerance when both outcomes are close by. This implication of the unified tradeoff model will be tested directly in Study 1.

While the unified tradeoff model as developed so far implies that the sign of delay preference for losses will generally depend on the delay to the losses, it also implies that it will *not* depend on the magnitude of the losses: When $\phi \gtrless 0$, it must be that $x_S \lessgtr x_L$, irrespective of whether x_S and x_L are large or small. The next step in the development of the unified tradeoff model is evidence showing that this may not be true for many people.

Motivation 3: Magnitude-Induced Sign Reversals of Delay Preference in Losses

Hardisty et al. (2013) conducted a series of studies that gave participants the opportunity to exhibit delay preferences of either sign for outcomes of either sign. They found that average implied interest rates for losses increased, from negative to positive, as the magnitude of the losses increased, hereafter, referred to as a *magnitude-induced sign reversal of delay preference* for losses (see Table 2, rows [1, 4], [3, 6] and [5, 8]). The unified tradeoff model interprets the observed reversal as an indication that the impact of time bias decreases as outcome magnitude increases, which it, in turn, interprets as an indication that time bias is evaluated *in proportion to* the outcomes under consideration.

As mentioned earlier, the tradeoff model relies on diminishing absolute sensitivity to outcomes for treating outcomes in proportion to one another. Thus, it would be natural if the model relied on the same principle for treating time bias in proportion to the outcomes under consideration. In Equations 8–10, time bias is associated with the later outcome, but is nonetheless evaluated independently from it (*bias segregation*); for time bias to be evaluated in proportion to the outcomes, it must operate on the later outcome *within* the value function (*bias integration*), which we, therefore, introduce as a structural condition to the unified tradeoff model.

Incorporating time bias within the value function, and reintroducing the σ-indicator from Equation 4 for a more compact statement of the model, decision makers with unipolar time bias will be indifferent between a sooner and a later outcome when

$$\kappa\sigma T = v(x_L - \sigma c) - v(x_S), \tag{12}$$

provided that $\sigma x_L > c > 0$, whereas decision makers with bipolar time bias will be indifferent between a sooner and a later outcome when

$$\kappa\sigma T = v(x_L - c) - v(x_S), \tag{13}$$

provided that, in the case of gains, $x_L > c > 0$, and, in the case of losses, $-\kappa T - v(-c) < -v(x_S)$. As was the purpose of bias integration, decision makers with bipolar time bias may exhibit magnitude-induced sign reversals of delay preference in losses. Solving Equation 13 for x_L,

$$v^{-1}[v(x_S) - \kappa T] + c = x_L, \tag{14}$$

we obtain delay aversion ($0 > x_L > x_S$) or delay tolerance ($x_L < x_S < 0$) when

$$v^{-1}[v(x_S) - \kappa T] + c \gtrless x_S, \text{ or} \tag{15}$$

$$v(x_S) - v(x_S - c) \gtrless \kappa T. \tag{16}$$

As x_S approaches 0, a decision maker will exhibit delay aversion when negative time bias outweighs positive time preference, that is, $-v(-c) > \kappa T$, and, as x_S approaches infinity, the same decision maker will exhibit delay tolerance as long as time preference is of positive sign, that is, $0 < \kappa T$, which is one of our core premises.

In sum, the unified tradeoff model interprets magnitude-induced sign reversals of delay preference in losses as an indication that time bias is treated in proportion to outcomes insofar as outcomes are treated in proportion to one another: Just as sensitivity to $x_L - x_S$ decreases as x_S increases, so does sensitivity to c.

Bias integration also ensures a more transparent interpretation of time bias. For instance, a person who is entitled to receive $1 now, and who demands $5 for accepting any longer wait, can only be indifferent between $1 now and an amount larger than $6 later, no matter how short the wait. More generally, under Equations 12 and 13, which invokes bias integration, the existence of an indifference point between a pair of single-dated gains requires that

$$x_L - c > x_S, \text{ or} \tag{17}$$

$$x_L - x_S > c. \tag{18}$$

Thus, the critical compensation is the constant c, independent of the amount the person is entitled to receive now. However, under Equation 8, which invokes bias segregation, the existence of an indifference point requires that

$$v(x_L) - v(c) > v(x_S), \text{ or} \tag{19}$$

$$x_L - x_S > v^{-1}[v(c) + v(x_S)] - x_S. \tag{20}$$

By diminishing absolute sensitivity to outcomes, the critical compensation is larger than c, and increases with the amount the person is entitled to receive now. For instance, the person from our example, who truly demands $5 for accepting any longer wait, may effectively be unable to reach indifference between $1 now and amounts up to $9 later. Moreover, the gap between $6 and the critical compensation for the existence of an indifference point becomes wider the more money is to be received now. These undesirable implications of bias segregation are instantly eliminated by imposing the structural condition of bias integration.

Motivation 4: The Delay-Speedup Asymmetry

Many intertemporal decisions are not only about the timing of outcomes, but also about *changes in* the timing of outcomes, and the question for the decision maker is to accept or reject the change proposed. Consider the following items from a study conducted by Scholten and Read (2013), in which x is an amount stated by the experimenter, and $\tilde{x}$ is an amount elicited from the participant.

Delay: *I am* [entitled to receive/obliged to pay] €x_0 *today. For me, it would be just as* [good/bad] *to delay the transaction, and* [receive/pay] €$\tilde{x}_t$ *in* t *years instead.*

Speedup: *I am* [entitled to receive/obliged to pay] €x_t *in* t *years. For me, it would be just as* [good/bad] *to speed up the transaction, and* [receive/pay] €$\tilde{x}_0$ *today instead.*

Equations 12 and 13 predict a symmetry between the scenarios: If the stated delayed amount in the speedup scenario (x_t) were set to the elicited delayed amount in the delay scenario ($\tilde{x}_t$), the elicited immediate amount in the speedup scenario ($\tilde{x}_0$) would be equal to the stated immediate amount in delay scenario (x_0), that is, if $(x_0, 0) \rightarrow (\tilde{x}_t, t)$ and $(\tilde{x}_0, 0) \leftarrow (\tilde{x}_t, t)$, then $x_0 = \tilde{x}_0$. However, the implied symmetry between delay and speedup scenarios does not hold, as first documented by Loewenstein (1988).

Perhaps the cleanest demonstration of what Loewenstein and Prelec (1992) refer to as the *delay-speedup asymmetry* comes from a study conducted by Appelt et al. (2011, Figure 1, and panel *a* of Figure 4). On average, their participants (a) demanded more compensation (a higher $\tilde{\imath} > 0$) for delaying a gain than they offered for speeding it up, (b) demanded less compensation (a lower $\tilde{\imath} > 0$) for speeding up a loss than they offered for speeding up a gain, and (c) had a slight tendency to *demand* compensation ($\tilde{\imath} \leq 0$) for delaying a loss.

The unified tradeoff model identifies three drivers of the delay-speedup asymmetry. The first is loss aversion, which yields a lower $\tilde{\imath}$ for losses than for gains overall. The second driver is time bias. Unipolar time bias reinforces positive time preference in promoting a higher $\tilde{\imath}$ for both gains and losses. Bipolar time bias reinforces positive time preference in promoting a higher $\tilde{\imath}$ for gains but countervails positive time preference in promoting a lower or even negative $\tilde{\imath}$ for losses. Combining these two effects is why, at the group level, $\tilde{\imath}$ is lower for losses than for gains. The third driver is *status-quo bias*, or a bias against changing plans, as originally discussed by W. Samuelson and Zeckhauser (1988). By the status-quo bias, a change in timing has opposite effects for gains and losses, in that the delay scenario yields a higher $\tilde{\imath}$ for gains, whereas the speedup scenario yields a higher $\tilde{\imath}$ in losses.

In the unified tradeoff model, status-quo bias puts a constant premium or penalty on the later outcome, as does time bias. However, whereas time bias depends on the decision maker and the sign of the outcomes, status-quo bias depends on the sign of the outcomes and the direction of any change in timing. With status-quo bias incorporated into the model, the indifference point is

$$\kappa\sigma T = v(x_L - \sigma c - \delta q) - v(x_S), \tag{21}$$

$$\kappa\sigma T = v(x_L - c - \delta q) - v(x_S), \tag{22}$$

where $q > 0$ is status-quo bias, and δ is an indicator for which option, if any, is the status quo. For choices between single-dated outcomes, $\delta = 1$ when the sooner outcome is the status quo (delay), $\delta = 0$ when neither option is the status quo (no change in timing), and $\delta = -1$ when the later outcome is the status quo (speedup).[6] Status-quo bias thus generates a higher $\tilde{\imath}$ when x_L and δq are of the same sign (delay a gain, or speed up a loss), and a lower $\tilde{\imath}$ when x_L and δq are of opposite sign (speed up a gain, or delay a loss).

Motivation 5: Superadditivity of Intervals in Gains

Suppose you are indifferent between \$100 today and \$200 in 1 year, and between \$200 in 1 year and \$300 in 2 years. By transitivity, you should be indifferent between \$100 today and \$300 in 2 years as well. However, the first controlled test of transitive indifference relations (Read, 2001) showed that participants were less delay averse over the undivided interval than over the divided one (*subadditivity*). Thus, you would prefer \$300 in 2 years to \$100 today. The unified tradeoff model ascribes this to positive time bias in gains: It operates in the same direction as positive time preference, toward delay aversion, by putting a constant penalty on the later gain *over each interval considered*, meaning *once* over the undivided interval, but *twice* over the two adjacent subintervals, so delay aversion will be weaker over the undivided interval than over the divided one.

A subsequent test of transitive indifference relations (Scholten & Read, 2006) confirmed the initial result, but also showed that, when a subinterval was itself divided into yet shorter subintervals, participants were *more* delay averse over the undivided subinterval than over the divided one (*superadditivity*). Thus, the overall pattern was a progression from superadditivity to subadditivity over intervals of increasing length.

To accommodate superadditivity, we follow the original statement of the tradeoff model by including a threshold in the evaluation of effective time differences. It is akin to the notion of "discriminatory thresholds," or a "deliberate disregard of 'small' but perceptible differences" between options along an attribute (Fishburn, 1974, p. 1446). In the tradeoff model, however, subthreshold differences are not disregarded entirely, but are given disproportionately less weight than suprathreshold differences, hereafter referred to as *subthreshold hyposensitivity*. This is accomplished by a bipartite tradeoff function over effective time differences (T), increasing linearly in T above and below a threshold, but more slowly below than above it (Scholten & Read, 2010, Equation 23, Figure 3). Letting Q denote the tradeoff function, and letting $\varepsilon > 0$ be the threshold in that function, the indifference point is

$$Q(\sigma T) = v(x_L - \sigma c - \delta q) - v(x_S), \tag{23}$$

$$Q(\sigma T) = v(x_L - c - \delta q) - v(x_S), \tag{24}$$

where,

$$Q(\sigma T) = \begin{cases} \kappa\sigma T, & T \geq \epsilon, \\ \kappa\sigma T/\vartheta, & T < \epsilon, \end{cases} \tag{25}$$

and $\vartheta > 1$ is subthreshold hyposensitivity. Thus formulated, the unified tradeoff model is able to generate the progression from superadditivity to subadditivity over intervals of increasing length. Specifically, if an undivided suprathreshold interval (T) is divided

[6] Equation 21 has a close precursor in Benzion et al.'s (1989) added-compensation approach, "which asserts that individuals require compensation for a change in their financial position" (p. 270). Under this approach, the indifference point is described as $x_S = d(t_L)(x_L - \delta q)$, where d is a discount function, i.e., $0 < d(t_L) \leq 1$ for $t_L \geq 0$.

into two subintervals (T_1 and T_2), and each of these is itself divided into yet shorter subintervals ($T_{\cdot 1}$ and $T_{\cdot 2}$), the following requirements must be met: First, $T_{\cdot 1}$ and $T_{\cdot 2}$ must be subthreshold, *and* hyposensitivity to these subintervals must outweigh positive time bias; second, *either* it must be the case that T_1 and T_2 are subthreshold, *and* hyposensitivity to these subintervals is outweighed by positive time bias, *or* it must be the case that T_1 and T_2 are suprathreshold, so that positive time bias is the only force left.

This concludes the development of the unified tradeoff model for choices between single-dated outcomes. We next report four studies on such choices to test several implications of our model for the effects of delay length, interval length, outcome sign, and outcome magnitude. As mentioned earlier, however, most studies have presented participants with *SS–LL* pairs only, thus *imposing* delay aversion for gains, and delay tolerance for losses, but we know now that many participants, if they had been given the opportunity to do so, would have exhibited delay aversion for losses instead. We will thoroughly reexamine delay, interval, sign, and magnitude dependence in intertemporal choice, to obtain an undistorted picture of choices between single-dated outcomes.

Study 1: Delay Dependence

Among the classic preference patterns is the common-difference effect: Given two outcomes and the delays to the outcomes, increasing both delays by a common additive constant changes indifference between a sooner and a later gain or loss into a preference for the later gain or the sooner loss. This is at odds with *stationarity*, by which preference relations depend only on the interval separating the outcomes, and not on the delay to the onset of the interval.

Let a be the common additive constant, let $(a + t_L) - (a + t_S) = t_L - t_S$ be the interval separating two gains, and let $a + t_S$ be the delay to the onset of the interval, hereafter the *front-end delay.* The most common stationarity violation is that increasing the front-end delay changes indifference between *SS* and *LL* into a preference for *LL* over *SS*. To restore indifference, some feature of the options must change. If the outcomes are adjusted, either the sooner gain (x_S) must be increased, or the later gain (x_L) must be decreased. If indifference can be restored without changing the ordinal relation between the gains ($x_L > x_S > 0$), the front-end delay has reduced the strength of delay preference, from stronger to weaker delay aversion. However, if indifference cannot be restored without changing the ordinal relation between the gains ($x_S \geq x_L > 0$), the front-end delay has reversed the *sign* of delay preference, meaning that the person exhibits indifference between *LS* and *SL*.

We can thus distinguish three scenarios. The first, *strong stationarity*, occurs when adding the front-end delay neither reduces the strength nor reverses the sign of delay preference. This is stationarity as conventionally conceived. The second scenario, *strong nonstationarity*, occurs when adding the front-end delay reduces not only the strength, but also the sign, of delay preference. We also will refer to this as a *delay-induced sign reversal of delay preference*. The third scenario, *weak stationarity*, is an intermediate scenario that occurs when adding the front-end delay reduces the strength, but not the sign, of delay preference.

Which of these scenarios is to be expected critically depends on outcome sign. The unified tradeoff model unconditionally generates strong stationarity in the domain of gains, since positive time preference and positive time bias for gains both contribute to delay aversion. However, the model allows for all three scenarios in the domain of losses. In particular, decision makers who combine positive time preference with negative time bias for losses can exhibit weak stationarity or strong nonstationarity, since positive time preference pushes them toward delay tolerance, while negative time bias pushes them toward delay aversion.

Below, we consider three models, and their implications for strong nonstationarity in the domain of losses. One is the unified tradeoff model, which combines unipolar or bipolar time bias with *positive* time preference. Another model is the mirror image of the unified tradeoff model, hereafter the *mirror-image model*, which combines unipolar and bipolar time bias with *negative* time preference. Pitting these two models against each other in our study on delay dependence will provide for a critical test between positive and negative time preference under the overarching assumption of time bias. The third model comes from the literature and will be considered throughout our studies on delay, interval, sign, and magnitude dependence. This is Benhabib et al.'s (2010) *fixed-cost discounting model*, the only existing model that, like the time-biased tradeoff models, is able to generate delay- and magnitude-induced sign reversals of delay preference in the domain of losses.

Delay-Induced Reversals of Delay Preference in Losses

In Study 1, participants chose between different timings of the same monetary loss, a payment of £100. As in research on the *immediacy effect* (Prelec & Loewenstein, 1991), the choices were between an immediate and a delayed payment (*present-future choices*) or, upon introducing a front-end delay, between a less delayed and a more delayed payment (*future-future choices*).

The Unified Tradeoff Model

Decision makers with unipolar time bias combine positive time preference (against the sooner payment) with positive time bias (in favor of the later payment), and are, therefore, unconflicted: They will exhibit delay tolerance in both present-future and future-future choices, since

$$-\kappa T < 0 < v(x + c) - v(x). \tag{26}$$

We henceforth refer to this variant of weak stationarity as *unconditional TT stationarity*.

Decision makers with bipolar time bias, however, combine positive time preference (against the sooner payment) with negative time bias (against the later payment), and, therefore, *are* conflicted: They will exhibit delay tolerance or delay aversion when

$$-\kappa T \lessgtr v(x - c) - v(x) < 0, \tag{27}$$

which depends not only the relative strength of time preference and time bias, but also on the effective time difference. By diminishing absolute sensitivity to delays, $w(a + t_L) - w(a + t_S) < w(t_L) - w(t_S)$, where $a > 0$ is the front-end delay. Thus, the effective time difference is larger for present-future choices than for future-future choices, and so these decision makers may exhibit delay tolerance in present-future choices, but delay aversion in future-future choices. We henceforth refer to this variant of strong nonstationarity as *TA nonstationarity*: "Pay sooner, but not now" (Table 2, rows [4, 5]).

It also possible that changes in the effective time difference never reverse the sign of delay preference, in which case we can obtain either variant of weak stationarity, that is, *conditional TT stationarity* (Table 2, rows [7, 8]) or *conditional AA stationarity* (Table 2, rows [1, 2]). The fourth possible pattern, *AT nonstationarity*, is ruled out by the unified tradeoff model.

The Mirror-Image Model

This model interprets delay aversion in losses at least in part as a manifestation of negative time preference. Decision makers with bipolar time bias combine negative time preference (in favor of the sooner payment) with negative time bias (against the later payment), and are, therefore, unconflicted: They will exhibit delay aversion in both present-future and future-future choices, since

$$\kappa T > 0 > v(x - c) - v(x). \tag{28}$$

We henceforth refer to this variant of weak stationarity as *unconditional AA stationarity*.

Decision makers with unipolar time bias, however, combine negative time preference (in favor of the sooner payment) with positive time bias (in favor of the later payment), and, therefore, *are* conflicted: They will exhibit delay aversion or delay tolerance when

$$\kappa T \gtrless v(x + c) - v(x) > 0. \tag{29}$$

Since the effective time difference is larger for present-future choices than for future-future choices, these decision makers may exhibit delay aversion in present-future choices, but delay tolerance in future-future choices. We henceforth refer to this variant of strong nonstationarity as *AT nonstationarity*: "Pay later, or else now." It also possible that changes in the effective time difference never reverse the sign of delay preference, in which case we can obtain either variant of weak stationarity, that is, *conditional AA stationarity* or *conditional TT stationarity*. The fourth possible pattern, *AT nonstationarity*, is ruled out by the mirror-image model.

To summarize, the critical contrast between the time-biased tradeoff models lies in the variant of strong nonstationarity that they generate: The unified tradeoff model rules out that people would want to pay later, or else now (*AT nonstationarity*), whereas the mirror-image model rules out that people would want to pay sooner, but not now (*TA nonstationarity*).

The Fixed-Cost Discounting Model

In Benhabib et al.'s (2010) model, an outcome is first discounted as a function of its delay, and then a fixed cost is put on the discounted outcome for *any* delay. In a present-future choice, the decision maker will be indifferent between the immediate and the delayed outcome when

$$x_S = d(t_L)x_L - c, \tag{30}$$

where $0 < d(t_L) < 1$ and $d'(t_L) < 0$ is delay discounting, $c > 0$ is the fixed cost. Additionally, to preclude indifference between a delayed gain and an immediate *loss*, or between a gain later and *nothing* now, the fixed cost must be smaller than the discounted value of the delayed gain, that is, $0 < c < d(t_L)x_L$. The fixed cost operates as a constant *discount* for a delayed gain, and as a constant *markup* for a delayed loss, analogous to bipolar time bias. In choices between different timings of the same payment ($x_L = x_S = x < 0$), the decision maker will be conflicted, since delay discounting contributes to a preference for paying later, whereas the fixed cost contributes to a preference for paying sooner. The decision maker will exhibit delay aversion or delay tolerance when the fixed-cost outweighs, or is outweighed by, delay discounting, that is,

$$x \gtrless d(t_L)x - c, \text{ or} \tag{31}$$

$$c \gtrless -[1 - d(t_L)]x. \tag{32}$$

The implications of the fixed-cost discounting model for future-future choices have so far gone unexamined. However, because any outcome is discounted as a function of its delay, and because the outcome, once discounted, is charged an additional cost for being delayed at all, the point of indifference between a less delayed and more delayed outcome would be described as

$$d(a)x_S - c = d(a + t_L)x_L - c, \text{ or} \tag{33}$$

$$d(a)x_S = d(a + t_L)x_L, \tag{34}$$

where $a > 0$ is a front-end delay. Thus, the fixed cost drops out, and the fixed-cost discounting model reduces to a conventional discounting model. In choices between different timings of the same payment ($x_L = x_S = x < 0$), the decision makers will be unconflicted, since delay discounting is the only force left, generating a preference for paying later, that is,

$$d(a)x < d(a + t_L)x, \text{ or} \tag{35}$$

$$d(a) > d(a + t_L). \tag{36}$$

To summarize, the fixed-cost discounting model will generate either *AT nonstationarity* or *conditional TT stationarity*, depending on whether the fixed cost outweighs, or is outweighed by, delay discounting in present-future choices.

In Study 1, we test the unified tradeoff model on its implications for delay dependence in the domain of losses. The focus is on strong nonstationarity: The unified tradeoff model has *TA nonstationarity* as its distinctive implication, whereas the fixed-cost discounting model and the mirror-image model have *AT nonstationarity* as their shared implication. A predominance of *TA nonstationarity* over *AT nonstationarity* would support our argument that choice is attribute-based (by the weakened position of the fixed-cost discounting model), and that delay aversion for losses is a manifestation of negative time bias outweighing positive time preference (by the weakened position of the mirror-image model).

Method

We examine delay preferences in a set of seven decisions about the timing of a single payment, as depicted in Table 3. We call it the *Pay Battery*. It is composed of four present-future choices, in which the sooner payment is due today, while the delay to the later payment ranges from 1 to 4 years, and three future-future choices, in which the later payment is due in 4 years, while the delay to the sooner payment ranges from 1 to 3 years. Patterns of weak stationarity and strong nonstationarity can be evaluated for three matched interval

Table 3
A Schematic Representation of the Seven Choice Tasks on the Pay Battery, Aggregate Proportions (p) of Deferral Choices, and Their 95% Confidence Intervals (95% CI_p)

		t (years)							
Timing of sooner payment	Interval	0	1	2	3	4		95% CI_p	
Immediate (present-future)	{0, 1}	−\$100	−\$100				.61	.64	.68
	{0, 2}	−\$100		−\$100			.62	.66	.69
	{0, 3}	−\$100			−\$100		.62	.65	.69
	{0, 4}	−\$100				−\$100	.60	.63	.67
Delayed (future-future)	{1, 4}		−\$100			−\$100	.54	.57	.61
	{2, 4}			−\$100		−\$100	.51	.55	.59
	{3, 4}				−\$100	−\$100	.52	.56	.60

pairs in the Pay Battery, that is, ({0, 1}, {3, 4}), ({0, 2}, {2, 4}), and ({0, 3}, {1, 4}).

For each participant, and each choice on the Pay Battery, we code a "defer" choice as 1, and an "advance" choice as 0. Then, for each of the three matched interval pairs, we compute a *difference score*, which is the difference between the coded choice on the sooner interval and the coded choice on the later interval. Finally, we compute a *nonstationarity score*, which is the sum of difference scores across the three matched interval pairs, analogous to Tversky and Russo's (1969) *M*-values. The nonstationarity scores range from −3 (*full AT nonstationarity*), through 0 (*TT stationarity* or *AA stationarity*), to 3 (*full TA nonstationarity*). For participants with a zero nonstationarity score, we also compute a *deferral score*, which is the sum of coded choices on the Pay Battery, and separate these participants into the delay tolerant, whose deferral score ranges from 4 to 7 (a majority score), and the delay averse, whose deferral score ranges from 0 to 3 (a minority score).

Upon listing the set of $2^7 = 128$ possible choice patterns on the Pay Battery, we compute the nonstationarity score for each choice pattern, and determine the frequency (f) with which each score occurs across the 128 patterns, while we subdivide zero scores and their frequencies into those associated with a majority deferral score and those associated with a minority deferral score. The relative frequency ($f/128$) is the occurrence rate that would be observed if choice behavior were entirely random, that is, a coin toss for each of the seven choices on the Pay Battery. We compare the predicted (P) with the observed (O) occurrence rate by taking $z = (O - P)/\sqrt{P}$, and evaluating the z-value under the standard normal distribution.

We state the distinctive implication of the unified tradeoff model for delay dependence as follows: Among the variants of strong nonstationarity, *full TA nonstationarity* (an indication that bipolar time bias is outweighed by positive time preference in present-future choices, but not in future-future choices) should occur more frequently than *full TA nonstationarity* (an indication that unipolar time bias is outweighed by negative time preference in present-future choices, but not in future-future choices), and *net TA nonstationarity* (scores 3, 2, and 1) should occur more frequently than *net AT nonstationarity* (scores −1, −2, −3).

Sample

We recruited 707 British residents through Bilendi, an internet service in which members earn points for completing surveys. We offered £0.75 worth of points (see Footnote 7). Participants averaged 40 years of age; 57% were female, 99% held an academic degree (70% a bachelor's degree), and 87% were employed (65% full time). The combined income of all household members in the previous year was less than £20,000 for 14% of participants, between £20,000 and £59,999 for 51%, and £60,000 or more for 22%; the remaining 12% preferred not to state their income.

Survey

Participants completed the Pay Battery as part of a larger survey on intertemporal decisions about monetary losses. It was introduced as follows:

> In the next seven choices, you are asked to imagine that you have an obligation to pay £100, but that you are given an opportunity to decide when you make that payment. The time until the payment will change from one choice to the next. In each case, please state your personal preference.

Each choice task was displayed on a separate page, with the sooner outcome always presented above the later outcome. The order of the tasks, either from {0, 1} to {3, 4}, as in Table 3, or from {3, 4} to {0, 1}, was counterbalanced across participants. Survey contents and raw data can be found at https://osf.io/ga2bh/?view_only=03f8fbf726724cc08567364a62cfb4bb. This study was not preregistered.[8]

Results

Group Level

Table 3 shows the proportions of participants choosing to defer payment for each choice task on the Pay Battery, and their 95% confidence intervals. The results are consistent with *TA nonstationarity*, the distinctive implication of the unified tradeoff

[7] Real monetary losses are almost impossible to implement in the laboratory, so we followed the tradition of using hypothetical losses. See Abdellaoui et al. (2010), Appelt et al. (2011), Hardisty et al. (2013), and McAlvanah (2010), for many more arguments against incentivizing loss tasks.

[8] The studies reported in this paper were conducted over a span of 10 years. Only Study 2 was preregistered, as it was the only study we ran after preregistration had become a widespread research practice.

model, in that preference for deferring the payment is reliably more common in present-future choices than in future-future choices: For all three matched interval pairs, the 95% confidence intervals for the proportion of deferral choices are nonoverlapping.

Individual Level

As reported in Table 4, the relative frequency of *TT stationarity* (46.7%), *AA stationarity* (24.8%), and *full TA nonstationarity* (7.5%) reliably exceed chance level, whereas the relative frequency of *full AT nonstationarity* (2.0%) does not. Thus, as expected, *full TA nonstationarity* occurs more often than *full AT nonstationarity*, a reliable result, $\chi^2(1) = 22.70$, $p < .001$. Also, *net TA nonstationarity* (20.6%) occurs reliably more often than *net AT nonstationarity* (7.9%), $\chi^2(1) = 40.10$, $p < .0001$.

Conclusion

The modal choice pattern, exhibited by almost 50% of the sample, was *TT stationarity*, to which both variants of time bias contribute. The second most common choice pattern, exhibited by almost 25%, was *AA stationarity*, to which only bipolar time bias contributes. Bipolar time bias also contributes to *TA nonstationarity*, the third most common choice pattern, with *net TA nonstationarity* being exhibited by over 20% of the sample, and *full TA nonstationarity* by 7.5%. The case for *TA nonstationarity* is stronger than these relative frequencies may suggest. First, *full TA nonstationarity* occurred at 4.7 times the rate it would be expected to occur merely by chance, whereas *TT stationarity* occurred at 3.0 times, and *AA stationarity* at 1.6 times, that rate. Second, *AT nonstationarity*, a pattern ruled out by the unified tradeoff model, but a shared implication of the mirror-image model and Benhabib et al.'s (2010) fixed-cost discounting model, was the least common choice pattern, with only 7.9% of the sample exhibiting *net AT nonstationarity*, and 2.0% exhibiting *full AT nonstationarity*, 1.3 times the rate it would be expected to occur by chance. In sum, the unified tradeoff model has convincingly survived its first critical test.

We compared the frequencies of *TA* and *AT nonstationarity* under the null hypothesis that, a priori, the two choice patterns would be equally likely to occur, which they would be if choice behavior were entirely random. We note that the null hypothesis of equiprobability holds under broader conditions than random choice behavior. For instance, when the "true" preference states in a universe of decision makers are the prototypical variants of *TT stationarity* (TTTT.TTT for the four present-future choices and the three future-future choices, respectively), *AA stationarity* (AAAA.AAA), *TA nonstationarity* (TTTT.AAA), and *AT nonstationarity* (AAAA.TTT), and, even when decision makers can deviate from these preference states by committing "errors," a constant error rate across choice tasks is sufficient for the null hypothesis of equiprobability to hold under our classification of choice patterns in Table 4. Of course, the error rate may not be constant, and, in the Web Appendix available at https://osf.io/ga2bh/?view_only=03f8fbf726724cc08567364a62cfb4bb, we conduct a true-and-error analysis that relaxes the null hypothesis of equiprobability. Consistent with the results from our qualitative analysis, *TA nonstationarity* is estimated to be more prevalent than *AT nonstationarity*.

Study 2: Interval Dependence

Alternative-based models of intertemporal choice entail *additivity of intervals*, as shown in Section 1 of the Appendix for indifference relations between options: If $(x_0, 0) \sim (x_t, t)$, and $(x_t, t) \sim (x_{2t}, 2t)$, then $(x_0, 0) \sim (x_{2t}, 2t)$. Additivity of intervals has been tested on indifference data from matching (McAlvanah, 2010; Read & Roelofsma, 2003; Zauberman et al., 2009) and choice-based matching (Kinari et al., 2009; McAlvanah, 2010; Read, 2001; Read & Roelofsma, 2003; Scholten & Read, 2006), but mostly using monetary gains as outcomes.[9] The one exception is McAlvanah (2010), who tested additivity of intervals as a null hypothesis using monetary gains *and* losses as outcomes. However, he presented participants with *SS–LL* pairs only, thus imposing delay aversion for gains, and delay tolerance for losses. As a result, additivity of intervals has not yet been properly tested in the domain of monetary losses.

A major contrast between the variants of time bias emerges regarding their implications for the nonadditivity of intervals in monetary losses, a technical derivation of which is provided in Section 2 of the Appendix. Under either variant of time bias, positive time preference operates toward delay tolerance for losses by putting a *proportional* weight on effective time differences. If all subintervals are suprathreshold, they are additive, in that $\kappa[w(2t) - w(t)] + \kappa[w(t) - w(0)] = \kappa[w(2t) - w(0)]$, so the impact of time preference does not change when an interval is divided into subintervals. Time bias, however, puts a *constant* premium or penalty on the later loss *over each interval considered*, so that the impact of time bias increases when an interval is divided into subintervals.

Unipolar time bias puts a constant *premium* on the later loss over each interval considered, that is, n premia over n subintervals, and one premium over the undivided interval. It therefore reinforces positive time preference in contributing to delay tolerance, while it also promotes *subadditivity* of intervals, or a weaker delay tolerance over the undivided interval than over the divided one. That is, someone indifferent between payments over a series of subintervals will prefer the sooner to the later payment over the undivided interval, and so the later payment must become smaller for indifference to be restored:

$$\text{If } (x_0, 0) \sim (x_t, t) \text{ and } (x_t, t) \sim (x_{2t}, 2t) \quad \text{for } x_{2t} < x_t < x_0 < 0, \tag{37}$$

$$\text{then } (x_0, 0) \succ (x_{2t}, 2t) \text{ and } (x_0, 0) \sim (y_{2t}, 2t) \quad \text{for } x_{2t} < y_{2t} < x_0 < 0. \tag{38}$$

However, there is subthreshold hyposensitivity as a potential third factor: If the effective subintervals are subthreshold but the effective undivided interval is not, this third factor countervails time bias in promoting *superadditivity* of intervals, or a stronger delay tolerance

[9] Additivity of intervals has also been tested on preference data from repeated choice (Dai, 2017) and nonrepeated choice (Cheng & González-Vallejo, 2016; Roelofsma & Read, 2000; Scholten & Read, 2006, 2010; Scholten et al., 2014), but mostly with monetary gains as outcomes. However, the interpretation of choice data also depends on how a deterministic model of choice, such as the tradeoff model, is translated into a probabilistic model. We turn to this issue later.

Table 4
Weak Stationarity and Strong Nonstationarity on the Pay Battery (Study 1)

Delay preference group	Deferral score	Nonstationarity score	Predicted	Observed	z
"Pay later or else now" (AT)	[3; 4]	−3	11.0 1.6%	14 2.0%	0.89
	[2; 5]	−2	66.3 9.4%	20 2.8%	−5.68
	[1; 6]	−1	165.7 23.4%	22 3.1%	−11.16
"Pay later" (TT)	[4; 7]	0	110.5 15.6%	330 46.7%	20.89
"Pay sooner" (AA)	[0; 3]	0	110.5 15.6%	175 24.8%	6.14
"Pay sooner but not now" (TA)	[1; 6]	1	165.7 23.4%	51 7.2%	−8.91
	[2; 5]	2	66.3 9.4%	42 5.9%	−2.98
	[3; 4]	3	11.0 1.6%	53 7.5%	12.62

Note. Column-wise percentages. The z-scores are evaluated under the standard normal curve, with $p(z < -1.96 \vee z > 1.96) = .05$.

over the undivided interval than over the divided one. Therefore, under unipolar time bias, the unified tradeoff model generates either subadditivity or superadditivity of intervals, depending on whether positive time bias outweighs, or is outweighed by, subthreshold hyposensitivity.

Bipolar time bias, in contrast, puts a constant *penalty* on the later loss over each interval considered, that is, n penalties over n subintervals, and one penalty over the undivided interval. It therefore countervails positive time preference in contributing to delay aversion, while it also promotes *superadditivity* of intervals. That is, someone indifferent between payments over a series of subintervals will prefer the later to the sooner payment over the undivided interval:

$$\text{If } (x_0, 0) \sim (x_t, t), \text{ and } (x_t, t) \sim (x_{2t}, 2t), \tag{39}$$

$$\text{then } (x_0, 0) \prec (x_{2t}, 2t). \tag{40}$$

How this translates into strength of delay preference depends on the sign of delay preference, which, in turn, depends on which of the countervailing forces is prevalent. When positive time preference is prevalent, as manifested by delay tolerance, superadditivity translates into a stronger tolerance to deferring a payment over the undivided interval than over the divided one (see Table 2, rows [7, 8, 9]):

$$\begin{gathered}\text{If } (x_0, 0) \sim (x_t, t), \text{ and } (x_t, t) \sim (x_{2t}, 2t)\\ \text{for } x_{2t} < x_t < x_0 < 0,\end{gathered} \tag{41}$$

$$\begin{gathered}\text{then } (x_0, 0) \prec (x_{2t}, 2t) \text{ and } (x_0, 0) \sim (y_{2t}, 2t)\\ \text{for } y_{2t} < x_{2t} < x_0 < 0.\end{gathered} \tag{42}$$

However, when negative time bias is prevalent, as manifested by delay aversion, superadditivity translates into a *weaker* aversion to deferring a payment over the undivided interval than over the divided one (see Table 2, rows [1, 2, 3]):

$$\begin{gathered}\text{If } (x_0, 0) \sim (x_t, t), \text{ and } (x_t, t) \sim (x_{2t}, 2t)\\ \text{for } x_0 < x_t < x_{2t} < 0,\end{gathered} \tag{43}$$

$$\begin{gathered}\text{then } (x_0, 0) \prec (x_{2t}, 2t) \text{ and } (x_0, 0) \sim (y_{2t}, 2t)\\ \text{for } x_0 < y_{2t} < x_{2t} < 0.\end{gathered} \tag{44}$$

Regardless of whether superadditivity translates into a stronger tolerance or a weaker delay aversion over the undivided interval, the later payment must become larger for indifference to be restored. Moreover, there is again subthreshold hyposensitivity as a potential third factor: If the effective subintervals are subthreshold but the effective undivided interval is not, this third factor reinforces negative time bias in promoting superadditivity of intervals. Therefore, under bipolar time bias, the unified tradeoff model unconditionally generates superadditivity of intervals.

Benhabib et al.'s (2010) fixed-cost discounting model is not strictly an alternative-based model, because, although each option is assigned an overall value, and the option with the highest overall value is chosen, the assignment of a fixed cost to a delayed outcome depends on a direct comparison with the other outcome along the time attribute: The fixed cost is assigned to the delayed outcome when the other outcome is a *sooner* outcome, and, more specifically, an *immediate* outcome. Despite this comparative process, however, the fixed-cost discounting model entails additivity of intervals, as shown in Section 1 of the Appendix. The reason is that the fixed cost drops out from the description of future-future choices, meaning that there is no propagation of fixed costs across subintervals. Indeed, in the absence of immediate outcomes, the fixed-cost discounting model reduces to a conventional discounting model: Strictly alternative-based, and, therefore, additive in intervals.

In Study 2, we identify two groups of participants. One group consistently exhibits delay aversion over three adjacent subintervals and over the undivided interval, hereafter *full delay aversion*, and denoted AAA.A, where delay preference over the three adjacent subintervals and delay preference over the undivided interval are separated by the full stop. Another group of participants consistently exhibits delay tolerance, hereafter *full delay tolerance*, and denoted TTT.T. The unified tradeoff model associates these groups with distinct patterns of interval effects.

The fully delay averse operate under bipolar time bias, and their negative time bias for losses outweighs positive time preference. Negative time bias not only contributes to delay aversion, it also promotes superadditivity of intervals, which, in this group of participants, means a weaker aversion to deferring a payment over an undivided interval than over a divided one. The unified tradeoff model therefore expects the fully delay averse to be unanimous in exhibiting superadditivity of intervals.

The fully delay tolerant include two subgroups of participants. One subgroup operates under bipolar time bias, as the fully delay averse do, but *their* negative time bias for losses *is* outweighed by positive time preference. Negative time bias, even if outweighed by positive time preference, promotes superadditivity of intervals, which, in this subgroup, means a stronger tolerance to deferring a payment over an undivided interval than over a divided one. The other subgroup operates under unipolar time bias, and their positive time bias for losses reinforces positive time preference in contributing to delay tolerance, while it also promotes subadditivity of intervals, meaning a weaker tolerance to deferring a payment over an undivided interval than over a divided one. In sum, the unified tradeoff model expects the fully delay tolerant to disperse into two subgroups, one exhibiting superadditivity, the other subadditivity, of intervals.

Drawing on the above deductions, we state the distinctive implications of the unified tradeoff model for interval dependence as follows: A majority of the fully delay averse should exhibit superadditivity of intervals, and the fully delay tolerant should exhibit subadditivity and superadditivity of interval effects in more equal proportion than the fully delay averse.

Method

We presented participants with three adjacent subintervals ($0 \to 1 \to 2 \to 3$), and one undivided interval ($0 \to 3$), with delays given in years. We elicited indifference points by means of a *chained* choice-based matching procedure: If $\tilde{x}_{L_{01}}$ is the later payment eliciting indifference over the subinterval $0 \to 1$, the sooner payment for the adjacent subinterval $1 \to 2$ is $x_{S_{12}} = \tilde{x}_{L_{01}}$. Similarly, if $\tilde{x}_{L_{12}}$ is the later payment eliciting indifference over the subinterval $1 \to 2$, the sooner payment for the adjacent subinterval $2 \to 3$ is $x_{S_{23}} = \tilde{x}_{L_{12}}$. Finally, if $\tilde{x}_{L_{23}}$ and $\tilde{x}_{L_{03}}$ are the later payments eliciting indifference over the subinterval $2 \to 3$ and the undivided interval $0 \to 3$, respectively, the participant exhibits subadditivity/additivity/superadditivity of intervals when $\tilde{x}_{L_{03}}$ is smaller than/equal to/larger than $\tilde{x}_{L_{23}}$.[10]

Choice Lists

The sooner payment (x_S) over the soonest subinterval ($0 \to 1$) was set at −£200. In each choice list, there were nine items. At the midpoint of the list $x_{L_5} = x_S$. Payments smaller than x_S were computed as $x_{L_{i-1}} = \delta x_{L_i}$ for $2 < i \leq 5$, and payments larger than x_S were computed as $x_{L_{i+1}} = x_{L_i}/\delta$ for $5 \leq i < 8$, using $\delta = 0.94425$, a value selected to ensure that amounts for the undivided interval would not be smaller than −100 (at the low end) or larger than −400 (at the high end). The resulting amounts were rounded to the nearest £5.

Indifference Points

If a participant always preferred the sooner payment on a list, we set $\tilde{x}_L$ at $x_{L_0} = \delta x_{L_1}$. If a participant always preferred the later payment on a list, we set $\tilde{x}_L$ at x_{L_9}. If a participant shifted once, from the later payment at position i on the list to the sooner payment at position $i + 1$, we set $\tilde{x}_L$ at x_{L_i}. If a participant shifted once, but in the reverse direction, we set the sooner payment on the next choice list at $x_{L_{i+1}}$, even though the shift constitutes a violation of *monotonicity*, since the participant shifts from rejecting a smaller payment to accepting a larger one. If a participant shifted more than once, and the last shift was at position i, we set the sooner payment on the next choice list at x_{L_i} when the shift was from the later to the sooner payment, and at $x_{L_{i+1}}$ when the shift was in the reverse direction. Participants who violated monotonicity and/or made multiple switches on at least one choice list were eliminated from the analysis. However, their survey was not terminated prematurely, because we wanted to get a complete picture of anomalous behavior in the chained matching procedure.

Sample

We recruited 1,001 British residents through Prolific Academic, with each participant earning £0.34 (amounting to about £7 an hour). Participants averaged 46 years of age; 52% were female, 59% held an academic degree (42% a bachelor's degree), and 81% were employed (56% full time). The combined income of all household members in the previous year was less than £20,000 for 14% of participants, between £20,000 and £59,999 for 54% of participants, £60,000 or more for 29% of participants, and 3% preferred not to state their income.

Survey

Participants were asked to imagine that they had to pay some amount of money, and that they had to decide whether to pay the amount sooner in time or else pay a different amount later in time. They first completed the three choice lists for the subintervals, and in their natural temporal order ($0 \to 1 \to 2 \to 3$), followed by the choice list for the undivided interval ($0 \to 3$). Each choice list was displayed on a separate page. Before reaching that page, participants were informed that they would be asked to consider paying the money [today/in 1 year/in 2 years/today] or [in 1 year/in 2 years/in 3 years/in 3 years]. On the page itself, they received a short instruction. For instance, the instruction for the undivided interval read as follows:

> Your choice is to either pay £200 *today* or instead pay an alternative amount *in 3 years*. Please state your preference for each pair of options.

The magnitudes of the delayed payment were presented in ascending order, with the sooner payment on the left, and the

[10] Busemeyer and Townsend (1993, Footnote 1) would consider the chained matching procedure an instantiation of *dynamic decision tasks*, which are "tasks that involve a sequence of decisions in which choices and outcomes available at later stages depend on choices and outcomes that occur at earlier stages," as opposed to *static decision tasks*, which "involve only a single stage—one decision followed by one outcome" (p. 432).

later payment on the right. Survey contents and raw data can be found at https://osf.io/ga2bh/?view_only=03f8fbf726724cc0856 7364a62cfb4bb). The design, the hypotheses, and the analysis plan of Study 2 were preregistered. For a record of the preregistration, see https://aspredicted.org/91N_BYJ.

Results

Across the four choice lists, 4.2% of participants violated monotonicity, 8.5% shifted more than once, and 0.7% did both, leaving us with 86.6% valid cases, or 867 from the original sample of 1,001 cases. Of the valid cases, 51.7% exhibited additivity of intervals, 12.8% shifted between delay aversion and delay tolerance across the four intervals, and 9.6% did both. We excluded these participants from the analysis as well, leaving us with 26.0% of the valid cases, 225 in total. Among these focal cases, that is, cases identified by our hypotheses, 68 were fully delay averse (AAA.A), and 157 were fully delay tolerant (TTT.T).

Among the fully delay averse, a very large majority (82.4%) exhibited superadditivity of intervals, $\chi^2(1) = 28.47$, $p < .0001$, as expected. Among the fully delay tolerant, a more modest, yet reliable, majority (61.8%) exhibited subadditivity of intervals, $\chi^2(1) = 8.72$, $p < .005$. The difference in the size of the majorities is reliable as well, $\chi^2(1) = 9.23$, $p < .005$, meaning that the fully delay averse were more cohesive, or the fully delay tolerant more diverse, in the direction of the interval effect, as expected. In sum, the unified tradeoff model has convincingly survived its second critical test.

Conclusion

We have so far derived four distinctive implications from the unified tradeoff model, two for the effects of delay length (Study 1), and two for the effects of interval length (Study 2), on decisions about monetary losses. All four implications were tested and confirmed, lending very strong support to our model. Next, we test the implications of the unified tradeoff model for sign and magnitude dependence.

Study 3: Sign and Magnitude Dependence in Present-Future Choices

Unipolar time bias is positive time bias for negative as well as positive outcomes. It increases delay aversion in gains and delay tolerance in losses, and it increases the absolute magnitude effect in both outcome domains. Moreover, it operates identically for gains and for losses, and so it attenuates the gain-loss and sign-magnitude asymmetries. To show this, we solve the indifference point in the domain of gains, as described by Equation 12, for the later gain ($x_L > 0$), that is,

$$v^{-1}[\kappa T + v(x_S)] + c = x_L, \tag{45}$$

and solve the indifference point in the domain of losses, as described by Equation 13, for the later loss (on this particular occasion given as $-x_L < 0$), that is,

$$v^{-1}[-\kappa T + v(-x_S)] - c = -x_L. \tag{46}$$

When invoking constant loss aversion, by which the value of losses is magnified relative to the value of gains by a multiplicative constant, that is, $v(-x) = -\lambda v(x)$ for $x > 0$ and $\lambda > 1$,

$$v^{-1}[\kappa T/\lambda + v(x_S)] + c = x_L, \text{for } x_S, x_L > 0. \tag{47}$$

Combining Equations 45 and 47 into a single expression by letting $\Lambda = 1$ for gains, and $\Lambda = \lambda > 1$ for losses, the interest rate implied by the point of indifference between two outcomes separated by one unit of time is

$$\begin{aligned} \tilde{\imath} &= \frac{v^{-1}[\kappa T/\Lambda + v(x_S)] + c}{x_S} - 1 \\ &= \frac{v^{-1}[\kappa T/\Lambda + v(x_S)]}{x_S} + \frac{c}{x_S} - 1. \end{aligned} \tag{48}$$

As can be seen, c contributes both to a higher $\tilde{\imath}$ and a decrease of $\tilde{\imath}$ with the magnitude of x_S, identically for gains and losses. Therefore, unipolar time bias has a homogenizing effect on $\tilde{\imath}$ for gains and losses, superimposed on the asymmetries generated by outcome valuation (v) and time weighing (w, located in T).

Bipolar time bias is positive time bias for positive outcomes, and negative time bias for negative outcomes, and this yields major discrepancies between gains and losses, since negative time bias contributes to a *lower* or even *negative* $\tilde{\imath}$, and an *increase* of $\tilde{\imath}$ with the magnitude of x_S, that is, a *reverse* magnitude effect (See Table 2, rows [1, 4, 7], [2, 5, 8], [3, 6, 9]). The interest rate implied by the indifference point for two losses separated by one unit of time is

$$\tilde{\imath} = \frac{v^{-1}[\kappa T/\lambda + v(x_S)]}{x_S} - \frac{c}{x_S} - 1. \tag{49}$$

The model generates either a reverse magnitude effect ($\partial\tilde{\imath}/\partial x_S > 0$) or an absolute magnitude effect ($\partial\tilde{\imath}/\partial x_S < 0$), depending on whether the sooner loss has a time advantage or disadvantage, that is,

$$-\kappa T + \lambda v(c) \gtrless 0. \tag{50}$$

Furthermore, and in accord with Inequality Equation 16, the model generates delay aversion ($\tilde{\imath} < 0$) or delay tolerance ($\tilde{\imath} > 0$) in losses when

$$\lambda[v(c + x_S) - v(x_S)] \gtrless \kappa T. \tag{51}$$

Thus, diminishing absolute sensitivity to outcomes is an additional factor in determining delay preference. As it vanishes, the model generates delay aversion whenever the sooner loss has a time advantage over the later loss, that is, $-\kappa T + \lambda v(c) > 0$. Otherwise, sensitivity to c decreases as the magnitude of x_S increases, and so the model generates delay tolerance over a broader range of losses. Moreover, decision makers who share a bipolar time bias will nonetheless differ in the strength of time bias, as they will in the strength of time preference, and in the rate at which absolute sensitivity to outcomes diminishes with outcome magnitude. Therefore, some may consistently exhibit delay aversion over a given range of losses (the fully delay averse), others may exhibit the magnitude-induced sign reversal of delay preference, from delay aversion to delay tolerance, and still others may consistently exhibit delay tolerance (the fully delay tolerant). However, they should *all* exhibit a reverse magnitude effect, hereafter referred as the *diversity-proof reverse magnitude effect*.

Benhabib et al. (2010) developed the fixed-cost discounting model as an account of the absolute magnitude effect in present-future choices between gains, but Hardisty et al. (2013) understood that, in present-future choices between losses, it would generate a reverse magnitude effect, and a magnitude-induced sign reversal of delay preference. Solving Equation 30 for the later outcome,

$$(x_S + c)/d(t_L) = x_L, \tag{52}$$

the interest rate implied by the indifference point for two outcomes separated by one unit of time is

$$\tilde{\imath} = (1 + c/x_S)/d(t_L) - 1. \tag{53}$$

Thus, the fixed-cost discounting model will generate an absolute magnitude effect for gains, but a reverse magnitude effect for losses, as long as $c > 0$. Furthermore, the later outcome will be better or worse than the sooner one, that is, $x_L \gtrless x_S$, when

$$c \gtrless -[1 - d(t_L)]x_S. \tag{54}$$

For gains, the right-hand side is of negative sign, and so the model unconditionally generates delay aversion. For losses, however, the right-hand side is of positive sign, and so the model generates delay aversion when the sooner loss is small, but delay tolerance when the sooner loss is large, which is the magnitude-induced sign reversal of delay preference. Individual differences in the magnitude of the fixed cost and the rate of delay discounting may create diversity in the sign of delay preferences, but not in the direction of the magnitude effect: Regardless of delay preference, individuals should exhibit a reverse magnitude effect, which is the diversity-proof reverse magnitude effect.

Hardisty et al.'s (2013) studies were a major step forward for research on sign and magnitude dependence in intertemporal choice, showing how a straightforward, and long awaited, modification in experimental procedure, that is, giving participants the opportunity to exhibit delay preference of either sign in either outcome domain, can lead to drastically different results. However, the role of *diversity* in sign and magnitude dependence cannot be evaluated on their data, since they only manipulated outcome sign within participant, and did not do the same for outcome magnitude.

Method

We closely follow Hardisty et al.'s (2013, Study 3) method, but we manipulate outcome magnitude as well as outcome sign within participant, so that we can conduct our data analysis at three levels: (a) The group level, at which we observe the average participant, or "representative agent" (Camerer & Ho, 1994, p. 186); (b) the subgroup level, at which we select and compare two representative agents, that is, those who are fully delay averse and those who are fully delay tolerant regarding losses; and (c) the individual level, at which we let each participant be his or her own agent.

Group Level

At the group level, we examined average $\tilde{\imath}$ as a function of outcome sign and outcome magnitude.[11] We would replicate Hardisty et al.'s (2013) group-level results if we found $\tilde{\imath}$ to decrease with the magnitude of gains (absolute magnitude effect), but to increase (reverse magnitude effect), from negative to positive (magnitude-induced sign reversal of delay preference), with the magnitude of losses.

Subgroup Level

There were three outcome magnitudes: Small, medium, and large, so that, for each participant, and each outcome sign, there were $2^3 = 8$ possible choice patterns by which the sign of $\tilde{\imath}$ could vary across the three outcome magnitudes, with six including at least one sign reversal, and the remaining two including no sign reversals, referred to as full delay aversion (AAA) and full delay tolerance (TTT). For gains, the unified tradeoff model precludes seven of the eight choice patterns, leaving full delay aversion (AAA) as the only possibility. For losses, the model generates four choice patterns, including full delay aversion (AAA), which is a manifestation of bipolar time bias, and full delay tolerance (TTT), which can be a manifestation of bipolar *or* unipolar time bias. Only participants exhibiting a nonreversal for losses were admitted to the analysis at the subgroup level.

We examined average $\tilde{\imath}$ as a function of outcome sign, outcome magnitude, and delay preference for losses (AAA vs. TTT). Among the fully delay averse (AAA), we expected an absolute magnitude effect for gains, but a reverse magnitude effect for losses, as manifestations of bipolar time bias. Among the fully tolerant (TTT), we expected an absolute magnitude effect for gains, but *diversity in the direction of the magnitude effect for losses*: An absolute magnitude effect as a manifestation of unipolar time bias, and a reverse magnitude effect as a manifestation of bipolar time bias.

Individual Level

For losses, the unified tradeoff model generates not only the two nonreversals (AAA and TTT), but also two *monotonic reversals,* that is, shifts from delay aversion to delay tolerance (ATT, AAT), which are magnitude-induced sign reversals of delay preference at the individual level, and associated with bipolar time bias. The unified tradeoff model precludes the remaining four choice patterns, that is, the two monotonic reversals in the opposite direction (TAA, TTA), and the two *nonmonotic reversals* (ATA, TAT).

For each participant and outcome sign, there were $3^3 = 27$ possible patterns by which $\tilde{\imath}$ could vary across the three outcome magnitudes. A *weakly decreasing/increasing* pattern was recorded when $\tilde{\imath}$ decreased/increased at least once, but never increased/decreased; a *constant* pattern was recorded when $\tilde{\imath}$ never changed; and a *nonmonotonic* pattern was recorded otherwise. Among the 27 possible patterns, three are weakly decreasing (absolute magnitude effect), three are weakly increasing (reverse magnitude effect), one is constant (null effect), and the remaining 20 are nonmonotonic.

We examined the (relative) frequencies with which the different patterns occurred as a function of participants' delay preference for losses. Our predictions at the individual level echo those at the

[11] Continuously compounded implied interest rates preserve the functional relation with outcomes in the group-level data, because the arithmetic mean of individual-level $\tilde{\imath}$s yields the same result as taking the arithmetic means of $\ln|x_L|$ and $\ln|x_L|$, and computing the group-level $\tilde{\imath}$ from there, which is not true for discretely compounded implied interest rates (Scholten & Read, 2013).

subgroup level: Among the fully delay averse (AAA), we should see an absolute magnitude effect for gains, but a reverse magnitude effect for losses; among the fully delay tolerant (TTT), we should see an absolute magnitude effect for gains, but diversity in the direction of the magnitude effect for losses.

The Diversity-Proof Reverse Magnitude Effect in Losses

Across all levels of analysis, we should see the diversity-proof reverse magnitude effect in losses: (a) At the group level, a reverse magnitude effect and a magnitude-induced sign reversal; (b) at the subgroup level, a reverse magnitude effect among the fully delay averse (AAA); and (c) at the individual level, a reverse magnitude effect among the fully delay tolerant (TTT) as well as the fully delay averse (AAA), and magnitude-induced sign reversals (ATT, AAT).

Choice Lists

Participants completed a series of choice lists, closely resembling those administered by Hardisty et al. (2013). There were three lists for each outcome sign, corresponding to three magnitudes of the immediate outcome (x_S): small (£20), medium (£100), and large (£500). As in Study 2 and Hardisty et al. (2013), there were nine items in each list, across which the magnitude of x_L increased. The magnitude of x_L was computed using $\delta = 0.9$, and, for x_S of magnitude £20, £100, and £500, rounded to the nearest £1, £5, and £25, respectively (i.e., to the nearest 5%).

Implied Interest Rates

If a participant always preferred the sooner loss/later gain, we set $\tilde{x}_L$ to the smallest (first) delayed outcome in the list, and computed $\tilde{\imath}$ from there. If the participant always preferred the later loss/sooner gain, we set $\tilde{x}_L$ at the largest (ninth) delayed outcome in the list, and computed $\tilde{\imath}$ from there. If the participant shifted once across the list, we took $\tilde{\imath}$ to be the arithmetic average of i before and after the shift, unless the shift violated monotonicity, in which case we set $\tilde{\imath}$ to a missing value. If the participant shifted more than once, we set $\tilde{\imath}$ to a missing value as well.

Preference Patterns

Letting $\tilde{\imath}_j$ be the implied interest rate obtained for choice list j, we computed, for each participant, (a) $\tilde{\imath}_g = (\tilde{\imath}_{20} + \tilde{\imath}_{100} + \tilde{\imath}_{500})/3$ and $\tilde{\imath}_l = (\tilde{\imath}_{-20} + \tilde{\imath}_{-100} + \tilde{\imath}_{-500})/3$, for the overall deviation of the $\tilde{\imath}$s from zero; (b) $\tilde{\imath}_g - \tilde{\imath}_l$, for the gain-loss asymmetry; (c) two linear trends, $\tilde{\imath}_{500} - \tilde{\imath}_{20}$ and $\tilde{\imath}_{-500} - \tilde{\imath}_{-20}$, for the (absolute or reverse) magnitude effect; and, orthogonally, (d) the interactive trend, $(\tilde{\imath}_{-20} + \tilde{\imath}_{500})/2 - (\tilde{\imath}_{20} + \tilde{\imath}_{-500})/2$, for the sign-magnitude asymmetry. Also, we recorded an absolute magnitude effect for gains when $\tilde{\imath}_j$ was weakly decreasing in the magnitude of gains (both $\tilde{\imath}_{500} \leq \tilde{\imath}_{100}$ and $\tilde{\imath}_{100} < \tilde{\imath}_{20}$, or both $\tilde{\imath}_{500} < \tilde{\imath}_{100}$ and $\tilde{\imath}_{100} \leq \tilde{\imath}_{20}$), and we recorded a reverse magnitude effect for gains when $\tilde{\imath}_j$ was weakly increasing in the magnitude of gains (both $\tilde{\imath}_{500} \geq \tilde{\imath}_{100}$ and $\tilde{\imath}_{100} > \tilde{\imath}_{20}$, or both $\tilde{\imath}_{500} > \tilde{\imath}_{100}$ and $\tilde{\imath}_{100} \geq \tilde{\imath}_{20}$). We did the same for losses.

Sample

We recruited 406 British residents through Bilendi, and offered £0.75 worth of points.[12] Participants averaged 40 years of age; 46% were female, 38% held an academic degree (27% a Bachelor's degree), and 69% were employed (47% full time). The combined income of all household members in the previous year was less than £20,000 for 30% of participants, between £20,000 and £59,999 for 49%, and £60,000 or more for 9%; the remaining 12% preferred not to state their income.

Survey

Participants first completed the choice lists for losses, followed by the corresponding lists for gains. We administered loss and gain trials in this order because we expected that participants would spontaneously demand compensation for deferring gains, which could encourage them to believe that, to be consistent, they should offer compensation for deferring losses (carry-over effect). This decision was motivated by Hardisty et al.'s (2013, Study 3) observation that implied interest rates for losses were higher when the loss tasks were completed after the gain tasks, whereas those for gains did not vary with the order of the tasks. The choice lists for [losses/gains] were introduced as follows:

> In the next six questions, you are asked to imagine that you *[have an obligation to pay/are entitled to receive]* an amount of money, but that you are given an opportunity to *[pay/receive]* the amount today or else a different amount in the future. In each case, please state your personal preference.

Each choice list was displayed on a separate page, and, as in Hardisty et al.'s (2013) Study 3 and our Study 2, the magnitudes of x_L were presented in ascending order, with the sooner outcome (e.g., "Receive £100 today") on the left, and the later outcome (e.g., "Receive £100 in 1 year") on the right. Survey contents and raw data can be found at https://osf.io/ga2bh/?view_only=03f8fbf726724cc08567364a62cfb4bb. This study was not preregistered.

Results

Across the six choice lists, 5.7% of participants violated monotonicity, 5.4% shifted more than once, and 2.7% did both, leaving us with 86.2% valid cases.

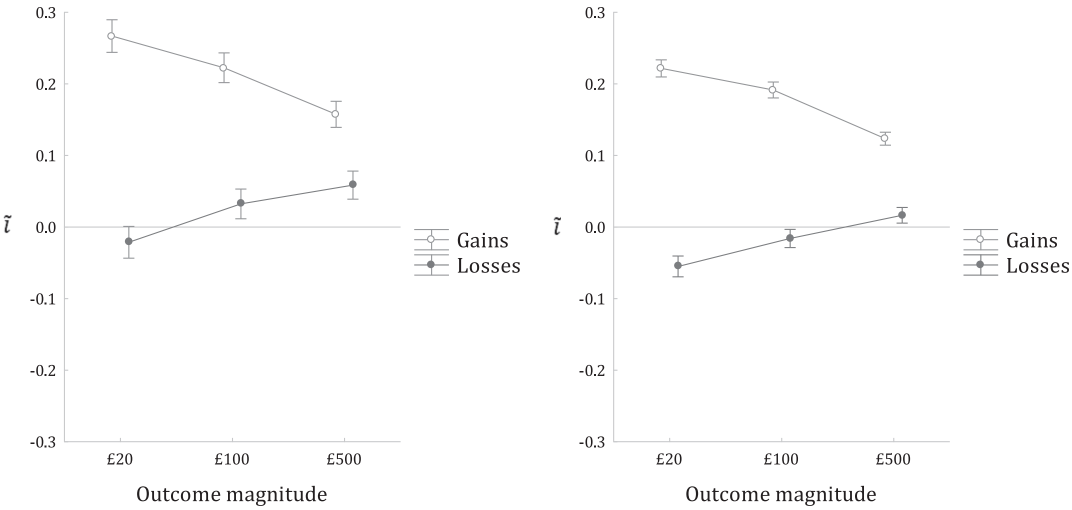
Group Level

The left panel of Figure 1 shows average $\tilde{\imath}$s as a function of outcome sign and magnitude. We replicate Hardisty et al.'s (2013) group-level results: For gains, average $\tilde{\imath}$s are positive, $t(368) = 23.42$, $p < .0001$ (2-tailed), Cohen's $d = 1.22$, 95% CI_d [1.08, 1.35], and exhibit the absolute magnitude effect, $t(373) = -11.93$, $p < .0001$, $d = -0.62$, 95% CI_d [−0.73, −0.51], whereas, for losses, average $\tilde{\imath}$s exhibit the reverse magnitude effect, $t(372) = 6.81$, $p < .0001$, $d = 0.35$, 95% CI_d [0.25, 0.46], and, although positive overall, $t(372) = 2.66$, $p < .01$, $d = 0.14$, 95% CI_d [0.04, 0.24], exhibit the magnitude-induced sign reversal, with confidence

[12] Because we intend a clean comparison between losses and gains, we treated loss and gain tasks in rigorously the same way, also in incentivizing neither (see Footnote 7). It should also be noted that numerous comparisons between incentivized and nonincentivized choices between single-dated gains have shown no systematic differences (see the review in Brañas-Garza et al., 2023, and their lab, field, and online experiments with over 2,000 participants).

Figure 1

Group-Level Results for Present-Future Choices (Study 3, Left Panel) and Future-Future Choices (Study 4, Right Panel)



Note. Average implied interest rates ($\tilde{\iota}$s), and their 95% confidence intervals, as a function of outcome sign and magnitude.

intervals shifting location from below the zero line for small losses to above the zero line for medium and large losses.

Subgroup Level

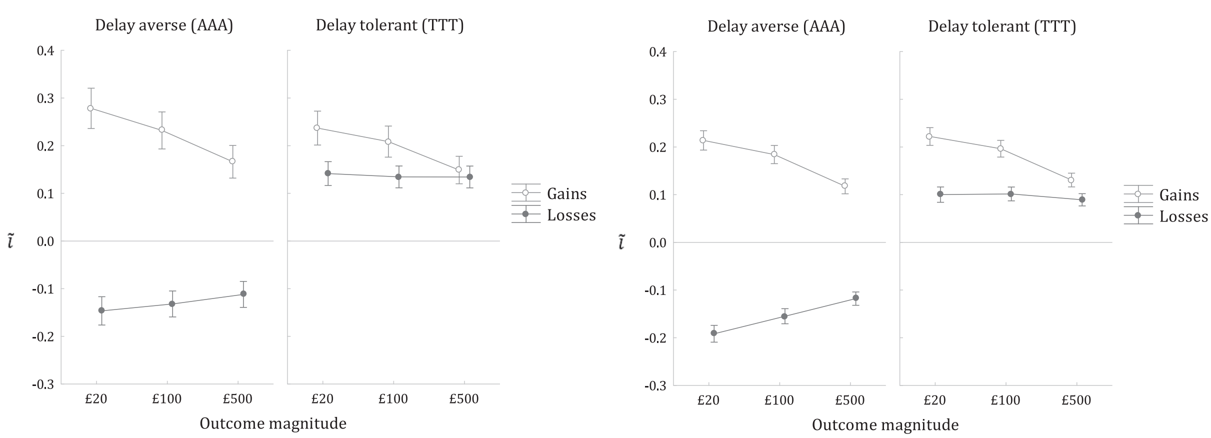
The top panel of Figure 2 shows average $\tilde{\iota}$s as a function of outcome sign and magnitude, separately for those who, in the domain of losses, are fully delay averse (AAA) and those who are fully delay tolerant (TTT). The fully delay averse exhibit the absolute magnitude effect for gains, $t(104) = -5.85$, $p < .0001$, $d = -0.57$, 95% CI_d [−0.78, −0.36], and the reverse magnitude effect for losses, $t(109) = 4.11$, $p < .0001$, $d = 0.39$, 95% CI_d [0.20, 0.58]. The fully delay tolerant exhibit the three classic outcome-related preference patterns: The absolute magnitude effect, $t(146) = -5.65$, $p < .0001$, $d = -0.47$, 95% CI_d [−0.64, −0.30], the gain-loss asymmetry, $t(145) = 3.89$, $p < .0005$, $d = 0.32$, 95% CI_d [0.15, 0.49], and the sign-magnitude asymmetry, $t(146) = -5.10$, $p < .0001$, $d = -0.42$, 95% CI_d [−0.59, −0.25], consisting in a strong and reliable absolute magnitude effect for gains, $t(146) = -7.93$, $p < .0001$, $d = -0.65$, 95% CI_d [−0.83, −0.47], but a weak and unreliable absolute magnitude effect for losses, $t(154) = -0.82$, $p = .41$, $d = -0.07$, 95% CI_d [−0.22, 0.09].

Individual Level

The top panel of Table 5 shows the eight sign patterns, and the (relative) frequencies with which they occurred, for each outcome domain (gains and losses). In the domain of gains, the overwhelming

Figure 2

Subgroup-Level Results for Present-Future Choices (Study 3, Top Panel) and Future-Future Choices (Study 4, Bottom Panel)



Note. Average implied interest rates ($\tilde{\iota}$s), and their 95% confidence intervals, as a function of outcome sign and magnitude, among the fully delay averse (AAA) and the fully delay tolerant (TTT).

Table 5
Magnitude-Induced Sign Reversals and Nonreversals at the Individual Level in Present-Future Choices (Study 3) and Future-Future Choices (Study 4)

		Outcome domain			
		Gains		Losses	
Timing of sooner outcome	Delay preference pattern	Frequency	Percentage	Frequency	Percentage
Immediate (present-future)	AAA	353	95.66	110	29.49
	TTT	5	1.36	155	41.55
	ATT	0	0.00	55	14.75
	AAT	2	0.54	35	9.38
	TAA	3	0.81	4	1.07
	TTA	1	0.27	4	1.07
	ATA	4	1.08	4	1.07
	TAT	1	0.27	6	1.61
Delayed (future-future)	AAA	743	97.38	253	34.38
	TTT	5	0.66	295	40.08
	ATT	4	0.52	68	9.24
	AAT	3	0.39	87	11.82
	TAA	5	0.66	10	1.36
	TTA	1	0.13	6	0.82
	ATA	1	0.13	9	1.22
	TAT	1	0.13	8	1.09

Note. Column-wise percentages.

majority of participants exhibit full delay aversion (AAA). In the domain of losses, participants disperse into those exhibiting full delay tolerance (TTT), full delay aversion (AAA), and reversals from delay aversion to delay tolerance (ATT, AAT). These reversals occur much more frequently than reversals in the opposite direction (TAA, TTA) and nonmonotonic reversals (ATA, TAT) combined, $\chi^2(1) = 48.00$, $p < .0001$.

The top panel of Table 6 shows the (relative) frequencies with which magnitude effects (absolute, null, reverse, or nonmonotonic) occurred for each delay preference (averse or tolerant) in each outcome domain (gains or losses). Null effects were abundant, making up almost half the cases recorded in gains, and over two-thirds of the cases recorded in losses. Otherwise, we see greater uniformity for gains than for losses.

In their nearly unanimous preference for advancing gains (AAA), participants exhibit absolute magnitude effects at a much higher rate than reverse and nonmonotonic magnitude effects combined, $\chi^2(1) = 80.90$, $p < .0001$. Analogously, participants who consistently prefer advancing losses (AAA) exhibit reverse magnitude effects at a much higher rate than absolute and nonmonotonic magnitude effects combined, $\chi^2(1) = 11.56$, $p < .005$. However, participants who consistently prefer deferring losses (TTT) exhibit reverse magnitude effects at a *lower* rate than absolute and nonmonotonic magnitude effects combined, $\chi^2(1) = 3.95$, $p < .05$. Moreover, those participants

Table 6
Absolute, Null, Reverse, and Nonmonotonic Magnitude Effects at the Individual Level in Present-Future Choices (Study 3) and Future-Future Choices (Study 4)

			Magnitude effect			
Timing of sooner outcome	Outcome domain	Delay preference pattern	Absolute	Null	Reverse	Nonmonotonic
Immediate (present-future)	Gains	AAA	155 43.91%	166 47.03%	4 1.13%	28 7.93%
		TTT	0 0.00%	4 80.00%	1 20.00%	0 0.00%
	Losses	TTT	24 15.48%	98 63.23%	21 13.55%	12 7.74%
		AAA	1 0.91%	85 77.27%	21 19.09%	3 2.73%
Delayed (future-future)	Gains	AAA	377 50.74%	299 40.24%	20 2.69%	47 6.33%
		TTT	0 0.00%	3 60.00%	1 20.00%	1 20.00%
	Losses	TTT	52 17.63%	196 66.44%	34 11.53%	13 4.41%
		AAA	3 1.19%	169 66.80%	70 27.67%	11 4.35%

Note. Row-wise percentages.

exhibit absolute and reverse magnitude effects at similar rates, $\chi^2(1) = 0.20$, $p = .35$, meaning that the sign-magnitude asymmetry, as "confirmed" at the subgroup level, is, at least in part, an artifact of diversity in the direction of the magnitude effect.

Of course, the net result of averaging opposite effects depends not only on the frequencies with which they occur, but also on their sizes. The top panel of Figure 3 shows average $\tilde{\imath}$s as a function of outcome sign and magnitude, only for participants who consistently prefer deferring losses (TTT), separately for those who, in the domain of losses, exhibit an absolute magnitude effect, $t(23) = -8.07$, $p < .0001$, $d = -1.65$, 95% CI_d [−2.26, −1.02], and those who exhibit a reverse magnitude effect, $t(20) = 6.81$, $p < .0001$, $d = 1.49$, 95% CI_d [0.85, 2.10]. The opposite effects are about the same size, meaning that the "sign-magnitude asymmetry" encountered at the subgroup level is, in large part, an aggregation artifact.

Conclusion

In Study 3, we replicated Hardisty et al.'s (2013) group-level results: The absolute magnitude effect for gains, and a reverse magnitude effect for losses, including a magnitude-induced sign reversal of delay preference, from delay aversion to delay tolerance.

Most diagnostic for the purpose of evaluating the unified tradeoff model, however, are the individual-level results. In the domain of gains, nearly the entire sample exhibited full delay aversion (AAA), and, null effects aside, the absolute magnitude effect was predominant. In the domain of losses, almost 30% of the sample exhibited full delay aversion (AAA); over 20% exhibited magnitude-induced reversals from delay aversion to delay tolerance (AAT, ATT); and over 40% exhibited full delay tolerance (TTT). Not counting null effects, which were even more abundant in losses than in gains, the reverse magnitude effect was predominant among the fully delay averse (AAA), but was as common (and as strong) as the absolute magnitude effect among the fully delay tolerant (TTT). These fine-grained results are entirely consistent with the unified tradeoff model.

One could argue, however, that the results from Study 3 are consistent not only with the unified tradeoff model, but also with the fixed-cost discounting model, if only it were given the freedom to assume that one group of decision makers imputes a fixed cost to delayed losses as well as delayed gains (analogous to bipolar time bias), while another group of decision makers imputes a fixed cost to delayed gains, but a fixed *benefit* to delayed losses (analogous to unipolar time bias). Even then, however, the fixed-cost discounting model would generate the full array sign and magnitude effects only for present-future choices, because, when applied to future-future choices, it reduces to a conventional discounting model, generating universally positive values of $\tilde{\imath}$, and a constant value of $\tilde{\imath}$ for each individual decision maker. Thus, an examination of sign and magnitude in future-future choices will provide for a critical test between the unified tradeoff model and the fixed-cost discounting model.

Study 4: Sign and Magnitude Dependence in Future-Future Choices

In Study 4, we examine whether the results for present-future choices, as obtained in Study 3, extend to future-future choices, as the unified tradeoff model implies. For that purpose, we increased the delays to the outcomes by 1 year, that is, from $0 \rightarrow 1$ (present-future choices) to $1 \rightarrow 2$ (future-future choices).

Method

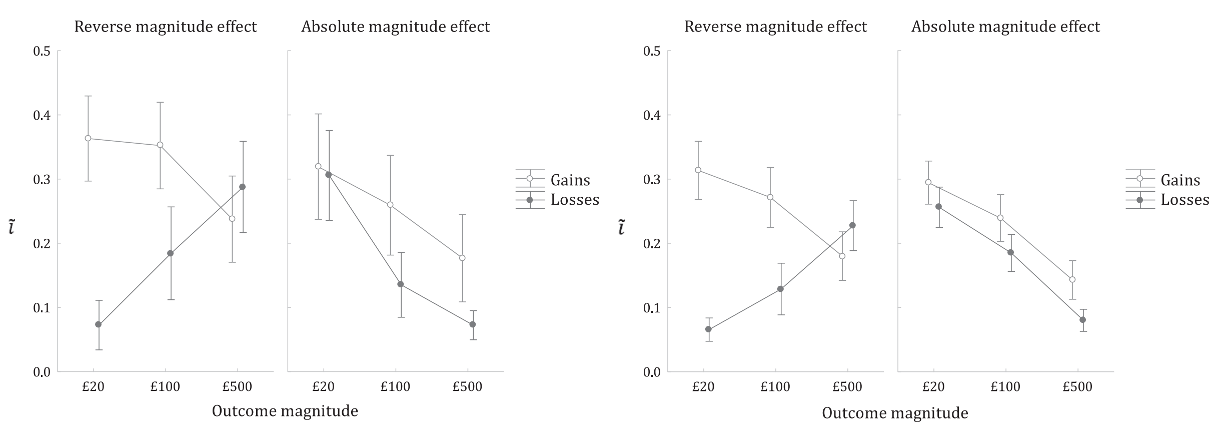
In Study 4, we followed the same data collection and data analysis protocols as in Study 3. However, noting that null effects of outcome magnitude were abundant (top panel of Table 6), we doubled the sample size.

Sample

We recruited 803 British residents through Prolific Academic, with each participant earning £0.60 (amounting to about £9 an hour). Participants averaged 38 years of age; 72% were female, 53%

Figure 3

Individual-Level Results for Present-Future Choices (Study 3, Top Panel) and Future-Future Choices (Study 4, Bottom Panel) by the Fully Delay Tolerant (TTT)



Note. Average implied interest rates ($\tilde{\imath}$s), and their 95% confidence intervals, as a function of outcome sign and magnitude, among those whose strength of delay tolerance for losses decreases (absolute magnitude effect) or increases (reverse magnitude effect) with the magnitude of the losses.

held an academic degree (40% a bachelor's degree), and 75% were employed (45% full time). The combined income of all household members in the previous year was less than £20,000 for 20% of participants, between £20,000 and £59,999 for 62% of participants, £60,000 or more for 13% of participants, and 5% preferred not to state their income.

Survey

We administered the tasks in the same order as in Study 3, and introduced the choice lists for losses as follows:

> In the next three questions, you are asked to imagine that you have an obligation to pay an amount of money, *but not as yet*. Rather, you must either pay £[20/100/500] *in 1 year* or else pay another amount *in 2 years*. For each amount in the lists below, please state whether you would rather pay £[20/100/500] in 1 year or the alternative amount in 2 years.

Analogously, we introduced the choice lists for gains as follows:

> In the next three questions, you are asked to imagine that you are entitled to receive an amount of money, *but not as yet*. Rather, you can either receive £[20/100/500] *in 1 year* or else receive another amount *in 2 years*. For each amount in the lists below, please state whether you would rather receive £[20/100/500] in 1 year or the alternative amount in 2 years.

In each task, the sooner outcome was presented on the left, and the later outcome on the right. Survey contents and raw data can be found at https://osf.io/ga2bh/?view_only=03f8fbf726724cc08567364a62cfb4bb. This study was not preregistered.

Results

Across the six choice lists, 3.6% of participants violated monotonicity, 8.2% shifted more than once, and 0.3% did both, leaving us with 87.9% valid cases.

Group Level

The right panel of Figure 1 shows average $\tilde{\tau}$s as a function of outcome sign and magnitude. For gains, average $\tilde{\tau}$s are positive, $t(763) = 38.5$, $p < .0001$, $d = 1.39$, 95% CI_d [1.29, 1.49], and exhibit an absolute magnitude effect, $t(765) = -20.17$, $p < .0001$, $d = -0.73$, 95% CI_d [−0.81, −0.65]. For losses, average $\tilde{\tau}$s exhibit a reverse magnitude effect, $t(747) = 10.73$, $p < .0001$, $d = 0.39$, 95% CI_d [0.32, 0.47], and, although negative overall, $t(736) = -3.25$, $p < .01$, $d = -0.12$, 95% CI_d [−0.19, −0.05], they exhibit a magnitude-induced sign reversal, with confidence intervals shifting location from below the zero line for small and medium losses to above the zero line for large losses.

Subgroup Level

The bottom panel of Figure 2 shows average $\tilde{\tau}$s as a function of outcome sign and magnitude, separately for those who, in the domain of losses, are fully delay averse (AAA) and those who are fully delay tolerant (TTT). The fully delay averse exhibit the absolute magnitude effect for gains, $t(241) = -10.09$, $p < .0001$, $d = -0.65$, 95% CI_d [−0.79, −0.51], and the reverse magnitude effect for losses, $t(253) = 8.39$, $p < .0001$, $d = 0.53$, 95% CI_d [0.40, 0.66]. The fully delay tolerant exhibit the three classic outcome-related preference patterns: The absolute magnitude effect, $t(286) = -10.37$, $p < .0001$, $d = -0.61$, 95% CI_d [−0.74, −0.49], the gain-loss asymmetry, $t(286) = 12.50$, $p < .0001$, $d = 0.74$, 95% CI_d [0.61, 0.87], and the sign-magnitude asymmetry, $t(287) = -9.50$, $p < .0001$, $d = -0.56$, 95% CI_d [−0.68, −0.44], consisting in a strong and reliable absolute magnitude effect for gains, $t(287) = -13.16$, $p < .0001$, $d = -0.78$, 95% CI_d [−0.91, −0.64], but a weak and marginally reliable absolute magnitude effect for losses, $t(295) = -1.71$, $p < .10$, $d = -0.10$, 95% CI_d [−0.21, 0.01].

Individual Level

The bottom panel of Table 5 shows the eight sign patterns, and the (relative) frequencies with which they occurred, for each outcome domain (gains and losses). In the domain of gains, the overwhelming majority of participants exhibit full delay aversion (AAA). In the domain of losses, participants disperse into those exhibiting full delay tolerance (TTT), full delay aversion (AAA), and reversals from delay aversion to delay tolerance (ATT, AAT). These reversals occur much more frequently than reversals in the opposite direction (TAA, TTA) and nonmonotonic reversals (ATA, TAT) combined, $\chi^2(1) = 79.17$, $p < .0001$.

The bottom panel of Table 6 shows the (relative) frequencies with which magnitude effects (absolute, null, reverse, or nonmonotonic) occurred for each delay preference (averse or tolerant) in each outcome domain (gains or losses). Null effects are abundant: Two out of every five cases recorded in gains, and two out of every three cases recorded in losses. Otherwise, we see greater uniformity for gains than for losses.

In their nearly unanimous preference for advancing gains (TTT), participants exhibit absolute magnitude effects at a much higher rate than reverse and nonmonotonic magnitude effects combined, $\chi^2(1) = 216.44$, $p < .0001$. Analogously, participants who consistently prefer advancing losses (AAA) exhibit reverse magnitude effects at a much higher rate than absolute and nonmonotonic magnitude effects combined, $\chi^2(1) = 37.33$, $p < .001$. However, participants who consistently prefer deferring losses (TTT) exhibit reverse magnitude effects at a lower rate than absolute and nonmonotonic magnitude effects combined, $\chi^2(1) = 9.71$, $p < .005$, and at a lower rate than absolute magnitude effects alone, a marginally reliable result, $\chi^2(1) = 3.77$, $p < .10$.

The bottom panel of Figure 3 shows average $\tilde{\tau}$s as a function of outcome sign and magnitude, only for participants who consistently prefer deferring losses (TTT), separately for those who, in the domain of losses, exhibit an absolute magnitude effect, $t(51) = -13.07$, $p < .0001$, $d = -1.89$, 95% CI_d [−2.34, −1.43], and those who exhibit a reverse magnitude effect, $t(33) = 10.56$, $p < .0001$, $d = 1.81$, 95% CI_d [1.26, 2.36]. The effects in opposite direction are about the same size.

Conclusion

Participants exhibited essentially the same patterns of sign and magnitude dependence in future-future choices as they did in present-future choices, and at all levels of analysis. Thus, the unified tradeoff model has convincingly survived its third critical test, whereas the fixed-cost discounting model faces yet another insurmountable challenge. Yet, the unified tradeoff model as developed so far only addresses choices between single-dated outcomes. We next expand it to also address choices involving outcome sequences.

Motivation and Development: Outcome Sequences

At this point, we have two different generalizations of the baseline tradeoff model. One is the unified tradeoff model for choices that involve only single-dated outcomes (Equations 23–25), which introduces three constants (time bias, status-quo bias, and subthreshold hyposensitivity), and two structural conditions (unanimity in positive time bias for gains, but diversity in the sign of time bias for losses, and bias integration). The other generalization is the cumulative tradeoff model (Scholten et al., 2016) for choices that also involve outcome sequences. In this section, we draw on both to arrive at a full statement of the unified tradeoff model.

In the cumulative tradeoff model, options are compared on their overall outcome value as it accumulates over time, and on the average time it takes for outcome value to accumulate, called the *duration of value accumulation*. The option favored by the comparisons is chosen. Overall outcome value is simply the sum of the values assigned to the outcomes in a sequence, that is, $\sum_{i=1}^{n} v(x_i)$, where either $x_i \geq 0$ for all i, and $x_i > 0$ for at least one i (a nonnegative outcome sequence), or $x_i \leq 0$ for all i, and $x_i < 0$ for at least one i (a nonpositive outcome sequence). The duration of value accumulation, denoted $\hat{t}$, is a weighted average of the delays to the outcomes in a sequence, obtained through a *cumulative weighing of time*, meaning that delay t_i is weighted by the value that has accumulated up to period i, that is,

$$\hat{t} = \frac{\sum_{i=1}^{n} a_i t_i}{\sum_{i=1}^{n} a_i}, \tag{55}$$

where

$$a_i = \sum_{k=1}^{i} v(x_k). \tag{56}$$

By the cumulative weighing of time, longer delays receive greater weight in the calculation of duration than shorter ones. The effective outcome difference is the difference between the overall outcome values of the outcome sequence with the longer duration and the outcome sequence with the shorter duration, that is,

$$X = \sum_{i=1}^{n_L} v(x_{L_i}) - \sum_{i=1}^{n_S} v(x_{S_i}). \tag{57}$$

For choices between single-dated outcomes, $n_L = n_S = 1$, in which case the effective outcome difference reduces to Equation 2. The effective time difference is the difference between the weights of the longer and the shorter duration, that is,

$$T = w(\hat{t}_L) - w(\hat{t}_S). \tag{58}$$

The duration of a single-dated outcome equals the delay to the outcome, that is, $\hat{t} = t$, because a single-dated outcome is an outcome sequence that ends where it begins. Thus, for choices between single-dated outcomes, the effective time difference reduces to Equation 3. The indifference point is

$$\kappa\sigma T = X, \tag{59}$$

where $\sigma = 1$ for nonnegative outcome sequences, and $\sigma = -1$ for nonpositive outcome sequences, which, in combination with $\kappa > 0$, yields positive time preference. For choices between single-dated outcomes, the indifference point reduces to Equation 4. We next develop the unified tradeoff model for the general context of choices involving outcome sequences, across three stages, each motivated by empirical findings.

Motivation 6: The Front-End Amount Effect

Consider the option pairs in Table 7. The Base pair is a choice between single-dated gains. The Alternate pair is obtained by adding the same amount to both options at the shorter delay, referred to as a *front-end amount*. This manipulation has been reported to yield a preference shift from L to S_F, more precisely, $.5 \leq P_{L_F S_F} < P_{LS}$ (Rao & Li, 2011; Read & Scholten, 2012; Table 7).

The benchmark model (i.e., the approximation of normative theory in experimental research) predicts that the front-end amount will not change preference, because it does not change the total interest to be earned (200% over the 2-month period). Moreover, the cumulative tradeoff model, as specified above, predicts that the front-end amount will yield a preference shift in the *opposite* direction, that is, from S to L_F. First, L_F has a larger outcome advantage than L, because, by diminishing absolute sensitivity,

$$X_F = [v(500) + v(30)] - v(510) > v(30) - v(10) = X, \text{ or} \tag{60}$$

$$v(10) - [v(510) - v(500)] > 0. \tag{61}$$

Second, L_F has a smaller time disadvantage than L, because, by the cumulative weighing of time, the duration of the former ($1 < \hat{t}_{L_F} < 2$) is shorter than the duration of the latter ($t_L = 2$), so that $\kappa T_F < \kappa T$. On both counts, the cumulative tradeoff model generates a reverse front-end amount effect.

In contrast, an earlier specification of the tradeoff model for two-period outcome sequences (Read & Scholten, 2012) *does* generate the front-end amount effect. It does so by defining overall outcome value as the value assigned to the sum of the outcomes, that is, $v(\sum_{i=1}^{n} x_i)$, rather than the sum of the values assigned to the outcomes, that is, $\sum_{i=1}^{n} v(x_i)$, hereafter referred to as *outcome integration* and *outcome segregation*, respectively (see also Thaler, 1985). With overall outcome value thus defined, L_F has a *smaller* outcome advantage than L, because, by diminishing absolute sensitivity,

$$X_F = v(530) - v(510) < v(30) - v(10) = X. \tag{62}$$

Table 7
Illustration of the Front-End Amount Effect

Base pair	Alternate pair
S: \$10 today	S_F: \$510 today
L: \$30 in 2 months	L_F: \$500 today and \$30 in 2 months

If this reduction in the outcome advantage of L_F outweighs the reduction in its time disadvantage, the front-end amount will reduce the probability of L_F being chosen. Thus, the unified tradeoff model interprets the front-end amount effect as a manifestation of *outcome valuation by outcome integration*, which it, therefore, introduces as a structural condition.[13]

Motivation 7: Sign Irrelevance of Common Consequences

Consider the option pairs in Table 8. The Base pair is a choice between single-dated outcomes, S and L, whereas the Alternate pair is a choice between two outcome sequences, S_C and L_C, which are obtained by introducing a *common consequence*, denoted C, to the Base pair, that is, an identical outcome at an identical delay. This manipulation has been reported to yield a preference shift from S to L_C, that is, $P_{L_C S_C} > P_{LS}$, irrespective of whether C is a gain (Rao & Li, 2011, Experiment 3; Read & Scholten, 2012, Experiment 1; Sun & Jiang, 2015, Experiment 1), or a loss (Rao & Li, 2011, Experiment 1; Sun & Jiang, 2015, Experiment 3A; Table 8).

The cumulative tradeoff model only addresses cases in which the common consequence, C, is of the same sign as the other outcomes: With C inserted between x_S and x_L, the duration of S_C is longer than the delay to x_S, whereas the duration of L_C is shorter than the delay to x_L, so L_C has a smaller time disadvantage than L, and has, therefore, a greater chance of being chosen. The unified tradeoff model also addresses cases in which the common consequence differs in sign from the other outcomes. It does so simply by incorporating the σ-indicator in the assessment of duration, that is,

$$a_i = \sigma v\left(\sum_{k=1}^{i} x_k\right), \tag{63}$$

thus introducing a structural condition henceforth referred to as *cumulative time weighing by absolute accumulated outcome value*. Under this condition, the sign of the common consequence is irrelevant, at least for the direction of the preference shift. It will not be irrelevant for the size of the shift, which is expected to be larger when C is a loss than when it is a gain: First, by diminishing absolute sensitivity, a common subtractive constant increases the value difference between two gains, whereas a common additive constant decreases it, that is, $v(30 - C) - v(10 - C) > v(30 + C) - v(10 + C)$; second, by loss aversion, the accumulated outcome value of L_C in period 1 will have a greater weight in the assessment of duration when C is a loss than when it is a gain, that is, $\sigma v(-C) > v(C)$ for $C > 0$, which is conducive to a larger decrease in duration.

Table 8
An Illustration of Sign Irrelevance of Common Consequences

Base pair	Alternate pair
S: Receive \$10 today	S_C: Receive \$10 today and [receive/pay] \$5 in 1 month
L: Receive \$30 in 2 months	L_C: [Receive/pay] \$5 in 1 month and receive \$30 in 2 months

Motivation 8: Desire for Spreading

People tend to exhibit a preference for segregating outcomes over time, that is, a *desire for spreading* (Loewenstein & Prelec, 1993), whether the outcomes are pleasant or unpleasant events (Linville & Fischer, 1991), monetary gains or losses (Guyse & Simon, 2011; Thaler & Johnson, 1990), or schedules of payments for consumption events (Prelec & Loewenstein, 1998). The cumulative tradeoff model treats the desire for spreading as a factor that, in combination with diminishing absolute sensitivity, determines the shape of the value function over *isolated* outcomes, with both factors contributing to a concave value function over gains, that is, $v(x) + v(x) > v(2x)$ for $x > 0$. However, the factors countervail one another in determining the shape of value function over losses, which will be either convex or concave, that is, $v(x) + v(x) \lesseqgtr v(2x)$ for $x < 0$, depending on whether diminishing absolute sensitivity outweighs, or is outweighed by, desire for spreading. In the unified tradeoff model, however, the value function is defined over *integrated* outcomes: Since $v(x + x) = v(2x)$, the shape of the value function has no bearing on preference for segregating or integrating outcomes over time, and so the desire for spreading must be incorporated as a separate component.

Technically, and drawing on Loewenstein and Prelec (1993), the desire for spreading is treated as distaste for departures from a uniform outcome distribution over time.[14] The departure of an option from uniformity, denoted D, is defined over deviations between accumulated outcomes. A record is kept of outcome accumulation until each consecutive period i under the actual outcome distribution, that is,

$$A_i = \sum_{k=1}^{i} x_k, \tag{64}$$

and is compared to the corresponding outcome accumulation under a uniform outcome distribution, that is,

$$U_i = \frac{i}{n}\sum_{k=1}^{n} x_k. \tag{65}$$

Taking the value of the absolute deviation between A_i and U_i, and summing across the n periods, yields the overall value of the departure from uniformity, that is,

$$D = \textstyle\sum_{i=1}^{n} v(|A_i - U_i|) \tag{66}$$

A distaste for departures from uniformity is given as $-\varsigma D$, where $\varsigma > 0$ is the desire for spreading.

Motivation 9: The Relocation Effect

Loewenstein and Prelec (1993) gave no indications concerning which periods are to be included in the calculation of the departure from uniformity, D, and they did not need to, because they presented

[13] The full statement of the cumulative tradeoff model ascribes the front-end amount effect to the influence of similarity on intertemporal choice. We turn to the issue of similarity later.

[14] The difference is that Loewenstein and Prelec (1993) obtain cumulative value under outcome segregation, whereas we obtain it under outcome integration.

their participants with options that had a "consumption event" in each consecutive period. If this is not the case, however, it becomes apparent that D can be calculated in different ways, as we will illustrate with reference to the option pairs in Table 9.

We identify two methods of calculating D. One, the *restrictive method*, is alternative-based, and includes only those periods in which an option has explicitly stated outcomes. For the above options, the number of periods to be included in the calculation is $n_A = 4$, $n_B = 2$, and $n_C = 1$, respectively, so that none of the options deviates from a uniform distribution. This result is counterintuitive, since A spreads out the total amount of $4x$ more evenly than B, and B does it more evenly than C.

Then there is the *comparative method*, which is attribute-based, and includes all periods in which either the option itself *or* the other options under consideration has an explicitly stated outcome. For pairwise comparisons, the number of periods to be included in the calculation is $n = n_L + n_S - n_{LS}$, where n_{LS} is the number of periods in which outcomes are explicitly stated for both options. Thus, $n = 4 + 2 - 2 = 4$ for the option pair $\{A, B\}$, and $n = 4 + 1 - 1 = 4$ for option pair $\{A, C\}$ meaning that both B and C deviate from a uniform distribution, and C deviates more from a uniform distribution than B. Furthermore, $n = 2 + 1 - 1 = 2$ for the option pair $\{B, C\}$, meaning that B does *not* deviate from a uniform distribution, but C does.

We advocate the comparative method for the calculation of D, thus introducing a *comparative assessment of spreading* as a structural condition to the unified tradeoff model. This is motivated not only by its intuitive appeal, but also by a preference pattern referred to as the *relocation effect* (Scholten et al., 2016). Consider the option pair in Table 10. The Alternate pair is obtained from the Base pair by relocating the common amount of \$500 from the front end to the back end of the time interval spanned by the options. This manipulation has been reported to decrease the probability of participants choosing L instead of S, that is, $P_{L_BS_B} < P_{L_FS_F}$.

By outcome integration, the relocation should leave the preference between the options unaffected, because the outcome advantage of L over S is $v(530) - v(510)$ in both option pairs. Moreover, by the cumulative weighing of time, L_B has a smaller time disadvantage than L_F, that is, $T_B < T_F$, so that the relocation should *increase* the probability of participants choosing L instead of S, that is, $P_{L_BS_B} > P_{L_FS_F}$.[15] Worse, the restrictive method of calculating D would *reinforce* the cumulative weighing of time in promoting a preference shift from S_F to L_B, because these are the options that would not deviate from a uniform distribution. However, by the comparative method, S_F deviates more from a uniform distribution than L_F does, and, analogously, L_B deviates more from a uniform distribution than S_B does, so the unified tradeoff model generates the relocation effect as long as the desire for spreading outweighs the cumulative weighing of time.

Table 9
Options for Calculating Departures From Uniformity

Option	t (years) 1	2	3	4
A	x	x	x	x
B	$2x$		$2x$	
C	$4x$			

Table 10
An Illustration of the Relocation Effect

Base pair	Alternate pair
S_F: \$510 today	S_B: \$10 today and \$500 in 2 months
L_F: \$500 today and \$30 in 2 months	L_B: \$530 in 2 months

This concludes the development of the unified tradeoff model, a full statement of which is provided in Table 11. The expansion to choices involving outcome sequences has introduced only one constant (desire for spreading) and three structural conditions (outcome valuation by outcome integration, cumulative time weighing by absolute accumulated outcome value, and comparative assessment of spreading). We next conduct a fifth study to test our model on the implied correspondence between delay preference and duration preference.

Study 5: Delay Preference and Duration Preference

Consider the option pairs in Table 10. Scholten et al. (2016) conducted a series of studies using the gain pair, and setting x at 300 or 400 U.K. pounds; 10 years earlier, Rubinstein (2006) had conducted a study using both pairs, and setting x at 500 U.S. dollars. The gain sequence and the single-dated loss are the options with the highest present value in the respective option pairs under any nonzero interest rate i (see Scholten et al., 2016, for a derivation). Nonetheless, Rubinstein (2006) found that majorities of participants favored the single-dated gain and the loss sequence. Scholten et al. (2016) also found persistent majorities in favor of the single-dated gain (Table 12).

The explanation that the unified tradeoff model offers for the majority choice of the single-dated gain is that the single-dated gain has a shorter duration than the gain sequence, that is, $t_S = 2 < 2⅓ = \tilde{t}_L$, and that the combination of positive time preference and positive time bias favors the option with the shortest duration. Moreover, the majority choice of the single-dated gain is an indication that the combination of positive time preference and positive time bias outweighed a desire for spreading out gains over time, because, by the comparative method of calculating the departure from uniformity, D, the single-dated gain deviates more from a uniform distribution than the gain sequence.

The majority choice of the loss sequence is a more intricate issue. The loss sequence is favored by a desire for spreading out losses over time, but the loss sequence has a longer duration than the single-dated loss, meaning that it is favored by positive time preference and positive time bias, but *disfavored* by negative time bias. Whether the loss sequence *is* the majority choice or not in any particular setting, the diversity in the sign of time bias for losses leads the unified tradeoff model to predict individual differences in the likelihood of the loss sequence being chosen.

In Study 5, we use the Pay Battery from Study 1 to identify two groups of participants: Those who, under changes in front-end delay,

[15] By a straightforward derivation, $T_B = \frac{v(10)+v(10)}{v(10)+v(510)} < \frac{v(530)+v(530)}{v(500)+v(530)} = T_F$.

Table 11
Full Statement of the Unified Tradeoff Model

Designation	Specification[a]
Effective outcome difference	$X = (V_L - V_S) - \varsigma(D_L - D_S)$
Overall outcome value	$V_L = v\left(\sum_{i=1}^{n} x_{L_i} - \beta c - \delta q\right)$ $V_S = v\left(\sum_{i=1}^{n} x_{S_i}\right)$
Overall value of the departure from uniformity	$D = \begin{cases} 0, & n = n_L = n_S = 1 \\ \sum_{i=1}^{n} v(d_i), & n = n_L + n_S - n_{LS} \end{cases}$
Departure from uniformity in period i	$d_i = \left\lvert \sum_{k=1}^{i} x_k - \frac{i}{n}\sum_{k=1}^{n} x_k \right\rvert$
Desire for spreading	$\varsigma > 0$
Unipolar time bias	$c > 0; \beta = \sigma = \begin{cases} 1, & \sum_{i=1}^{n} x_{L_i} > 0 \\ -1, & \sum_{i=1}^{n} x_{L_i} < 0 \end{cases}$
Bipolar time bias	$c > 0; \beta = 1$
Status-quo bias[b]	$q > 0; \delta = \begin{cases} 1, & S \rightarrow L \\ 0, & S \rightleftarrows L \\ -1, & S \leftarrow L \end{cases}$
Effective time difference	$T = w(\hat{t}_L) - w(\hat{t}_S)$
Duration of value accumulation	$\hat{t} = \frac{\sum_{i=1}^{n} a_i t_i}{\sum_{i=1}^{n} a_i}$
Cumulative weight of delay	$a_i = \sigma v(\sum_{k=1}^{i} x_k); \sigma = \begin{cases} 1, & \sum_{k=1}^{i} x_k > 0 \\ -1, & \sum_{k=1}^{i} x_k < 0 \end{cases}$
Indifference point	$Q(\sigma T) = X$
Tradeoff function	$Q(\sigma T) = \begin{cases} \kappa\sigma T, & T \geq \varepsilon \\ \kappa\sigma T/\vartheta, & T < \varepsilon \end{cases}$
Threshold	$\varepsilon > 0$
Subthreshold hyposensitivity	$\vartheta > 1$
Positive time preference	$\kappa > 0; \sigma = \begin{cases} 1, & \sum_{i=1}^{n} x_i > 0 \\ -1, & \sum_{i=1}^{n} x_i < 0 \end{cases}$

[a] The value function v and time-weighing function w exhibit reference dependence, diminishing absolute sensitivity, and augmenting proportional sensitivity; the value function also exhibits loss aversion. [b] The option from which an arrow originates is the status quo. The reciprocal arrows indicate that neither option is the status quo, but that one of the options may nonetheless be anchored on *as if it were* the status quo.

consistently choose a shorter delay to a payment (the *AA stationary*), and those who consistently choose a longer delay to the payment (the *TT stationary*). As interpreted by the unified tradeoff model, the *AA stationary* have bipolar time bias (positive for gains, negative for losses), and their negative time bias for losses outweighs their positive time preference, whereas the *TT stationary* include two unidentifiable subgroups: Those who have bipolar time bias as well, but whose negative time bias for losses *is* outweighed by positive time preference (the *conditional TT stationary*), and those who have unipolar time bias (positive for losses as well as gains), and whose positive time bias for losses reinforces positive time preference in promoting delay tolerance (the *unconditional TT stationary*).

Table 12
Rubinstein's (2006) Option Pairs

Option	t (years) 1	2	3	Option	t (years) 1	2	3
Gain sequence	x		x	Single-dated loss		$-2x$	
Single-dated gain		$2x$		Loss sequence	$-x$		$-x$

As to which of the identifiable groups, the *TT stationary* or the *AA stationary*, is more likely to choose the loss sequence, the unified tradeoff model arrives at a definite prediction, based on three considerations, First, the *AA stationary* choose the loss sequence only when the combination of desire for spreading and positive time preference outweighs negative time bias. Second, the *conditional TT stationary* is more likely to choose the loss sequence than the *AA stationary*, because their positive time preference is stronger and/or their negative time bias is weaker. Third, the *unconditional TT stationary* is most likely to choose the loss sequence, because, in this subgroup, positive time bias reinforces positive time preference in promoting delay tolerance. Thus, the prediction is that the *TT stationary* will be more likely to choose the loss sequence than the *AA stationary*.

It is a remarkable prediction, because, from an economic perspective, choosing to defer a monetary loss is the right thing to do under any positive rate of market interest, while choosing the loss sequence from Rubinstein's (2006) option pair is the wrong thing to do under any nonzero rate of interest, positive *or* negative. Should the "role reversal" of the *TT stationary* be confirmed, we can conclude that even economically sensible decisions to defer a monetary loss may not be based on economic considerations.

While the *AA stationary* should be least likely to choose the loss sequence, the unified tradeoff model does predict that even these decision makers will become more likely to choose it as the absolute magnitude of the losses increases, since its spreading advantage over the single-dated loss increases as the departures from uniformity increase in absolute magnitude. We test this prediction as well.

Method

Sample

We recruited 713 British residents through Bilendi, and offered £0.75 worth of points. Participants averaged 45 years of age; 53% were female, 76% held an academic degree (52% a Bachelor's degree), and 73% were employed (50% full time). The combined income of all household members in the previous year was less than £20,000 for 17% of participants, between £20,000 and £59,999 for 54%, and £60,000 or more for 17%; the remaining 11% preferred not to state their income.

Survey

The participants completed the Pay Battery, but, before that, they made 28 choices involving payment sequences. Two option pairs had Rubinstein's (2006) format, with x set at £200 in one pair, and at £400 in the other. There were 17 variations on these option pairs, obtained by inserting common amounts in common time periods, as when manipulating front-end amounts. These 19 options pairs will be analyzed below. The payment sequence choices were presented as choices between "payment plans," and in matrix format (analogous to Scholten et al., 2016, Figure 1). The left–right position of the options was counterbalanced across participants. Survey contents and raw data can be found at https://osf.io/ga2bh/?view_only=03f8fbf726724cc08567364a62cfb4bb. This study was not preregistered.

Results

The distribution of participants across delay preference groups on the Pay Battery was similar to that from Study 1. Almost half the sample (47.5%) exhibited *TT stationarity*, and almost a third of the sample (30.3%) exhibited *AA stationarity*. The remaining participants (22.6%) exhibited a variant of strong nonstationarity, with *full TA nonstationarity* (5.8%) outnumbering *full AT nonstationarity* (0.42%) by a ratio of more than 10:1, and *net TA nonstationarity* (16.3%) outnumbering net *AT nonstationarity* (6.3%) by a ratio of about 5:2.

When x was £200, the payment sequence was chosen by a minority of the *AA stationary* (44.4%), but by a large majority of the *TT stationary* (80.4%), a reliable difference, $\chi^2(1) = 75.90$, $p < .0001$. When x was £400, there was, among the *AA stationary*, a shift toward the payment sequence, $\chi^2(1) = 49.78$, $p < .0001$, now chosen by a majority of the *AA stationary* (60.2%), yet, a far smaller majority than the majority of *TT stationary* choosing the payment sequence (83.04%), a reliable difference, $\chi^2(1) = 35.77$, $p < .0001$.

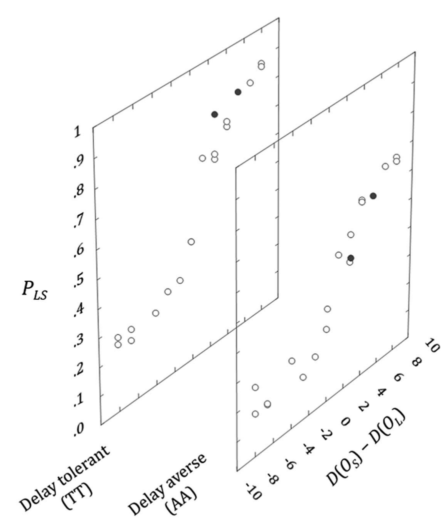
Turning to all 19 option pairs, Figure 4 shows the probability of choosing the option with the longest duration (L), instead of the option with the shortest duration (S), as a function of delay preference, and the difference between L and S in their departure from a uniform outcome distribution. The solid dots are Rubinstein's (2006) option pairs. We see that the delay tolerant are more likely to choose L than the delay averse, and that, in each delay preference group, the probability of choosing L steadily increases as it compares more favorably with S in spreading out the payments over time.

Conclusion

We confirmed both predictions of the unified tradeoff model. First, those who preferred longer delays to a payment (the *TT stationary*) were more likely to prefer a longer duration of payment accumulation than those who preferred shorter delays to a payment (the *AA stationary*). The factor responsible for this relation between delay preference and duration preference is time bias, with negative time bias being indispensable for generating a preference for shorter delays and durations.

Second, increasing the absolute magnitude of the payments yielded a preference shift from the single-dated payment to the

Figure 4
Results of Study 5



Note. Probability of choosing the option with the longest duration (L) instead of the one with the shortest duration (S), as a function of delay preference, and the difference between L and S in their departure from a uniform outcome distribution (calculated under an identity value function). The solid dots are Rubinstein's (2006) option pairs.

payment sequence. The factor responsible for this shift is the desire for spreading. The shift was very modest among the *TT stationary*, who exhibited a strong preference for the payment sequence anyway. This should come as no surprise, because, in this group, the payment sequence was favored by a preference for longer durations as well as a desire for spreading.

In sum, the unified tradeoff model has convincingly survived its fourth critical test. It can be concluded that our model offers an exceptionally broad coverage of preference patterns in intertemporal choice, many of which would be incomprehensible and irreconcilable otherwise. We next discuss the scope and parsimony of our model in relation to yet other phenomena.

Scope and Parsimony

In this section, we show how the unified tradeoff model is able to accomplish broader coverage with greater parsimony than all of its predecessors taken together. Moreover, we show how it accounts for preference patterns in consumption decisions, which is the final frontier for attribute-based models of intertemporal choice.

Relative Nonadditivity in Gains

For choices between single-dated outcomes, the unified tradeoff model invokes bias integration, meaning that time bias operates on the later outcome within the value function. This assumption was motivated by existing evidence of magnitude-induced sign reversals of delay preferences for losses in present-future choices (Hardisty et al., 2013), which we replicated (Study 3), and demonstrated to occur in future-future choices as well (Study 4). Bias integration has broad implications, also for nonadditivity of intervals. We derived implications from the unified tradeoff model for nonadditivity of intervals in losses, which we tested and confirmed (Study 2), whereas past research on this topic has by and large been restricted to gains. In that research, it has been found that the direction of nonadditivity, whether it is subadditivity or superadditivity, depends on the magnitude of the gains under consideration, a result that we will reexamine now.

People may exhibit a progression from superadditivity to subadditivity over intervals of increasing length, as Scholten and Read (2006) observed in indifference data from choice-based matching. In retrospect, their study was actually the first, and has so far been the last, to provide indifference data on a more intricate preference pattern, referred to as *relative nonadditivity*: Whether people exhibit superadditivity or subadditivity of intervals depends on how large or small gains are relative to their delays. Specifically, Scholten and Read's (2006) participants exhibited superadditivity when large amounts were distributed over short subintervals, but subadditivity when large amounts were distributed over a long undivided interval. Later, Scholten and Read (2010) and Scholten et al. (2014) observed relative nonadditivity in preference data from single-shot choices (see also Cheng & González-Vallejo, 2016). Specifically, their group-level data exhibited superadditivity or subadditivity *over the same set of intervals*, depending on how large or small the outcomes were relative to their delays.

Scholten et al. (2014) equated additivity of intervals in preference data with Luce's (1959) *product rule*, under which the odds of choosing the later outcome over an undivided interval ($t_S \rightarrow t_L$) should be equal to the product of the odds of choosing the later outcome over successive subintervals ($t_S \rightarrow t_M$ and $t_M \rightarrow t_L$):

$$\Omega_{LS} = \Omega_{LM} \cdot \Omega_{MS}. \tag{67}$$

For instance, if the odds of choosing \$200 in 1 year instead of \$100 today are 2:1, and if the odds of choosing \$300 in 2 years instead of \$200 in 1 year are 3:2, then the odds of choosing \$300 in 2 years instead of \$100 today should be 3:1. Strength of preference for monetary gains is subadditive in intervals when $\Omega_{LS} > \Omega_{LM} \cdot \Omega_{MS}$, and superadditive in intervals when $\Omega_{LS} < \Omega_{LM} \cdot \Omega_{MS}$. Relative nonadditivity is a shift from subadditivity to superadditivity when increasing the magnitude of the gains by a common multiplicative constant.

Relative nonadditivity has previously served as a key consideration in developments of the tradeoff model. Scholten et al. (2014) took it as an indication that the odds of choosing *LL* instead *SS* derive from the ratio between the outcome advantage of *LL* and the time advantage of *SS*, hereafter the *ratio rule*. To illustrate, under the baseline tradeoff model,

$$\Omega_{LS} = \left(\frac{X}{\kappa T}\right)^{1/\mu}, \tag{68}$$

where $\mu > 0$ is noise in choice behavior, hereafter set to 1. Suppose that an interval ($t_S \rightarrow t_L$) is divided into two equal-length contiguous subintervals ($t_S \rightarrow t_M \rightarrow t_L$), and the decision maker exhibits constant sensitivity to delays, that is, $w(t) = t$, so that $T = t_L - t_M = t_M - t_S$. Suppose also that the decision maker exhibits constant sensitivity to outcomes, that is, $v(x) = x$, so that $X = x_L - x_M = x_M - x_S$. Then, the combined odds of choosing the later gain over the successive subintervals are

$$\Omega_{LM} \cdot \Omega_{MS} = \frac{X}{\kappa T} \cdot \frac{X}{\kappa T} = \left(\frac{X}{\kappa T}\right)^2, \tag{69}$$

whereas the odds of choosing the later gain over the undivided interval are

$$\Omega_{LS} = \frac{X + X}{\kappa(T + T)} = \frac{X}{\kappa T}. \tag{70}$$

Thus, people can exhibit subadditivity in intervals ($\Omega_{LS} > \Omega_{LM} \cdot \Omega_{MS}$) when effective outcome differences are small *relative to* effective time differences ($X/\kappa T < 1$), but superadditivity in intervals ($\Omega_{LS} < \Omega_{LM} \cdot \Omega_{MS}$) when effective outcome differences are large *relative to* effective time differences ($X/\kappa T > 1$). In support of this interpretation, Scholten et al. (2014, Figure 4) found that a ratio-rule specification of the tradeoff model was able to almost perfectly reproduce an intricate pattern of relative nonadditivity in group-level choice odds. The ratio-rule account of relative nonadditivity is highly parsimonious; however, it only covers preference data from choice: Left out are preference data from other sources, for example, preference ratings, or, indeed, indifference data.

Originally, and more fundamentally, Scholten and Read (2010) took relative nonadditivity as an indication that intertemporal choice *is* attribute-based. In their formulation of the tradeoff model, subthreshold hyposensitivity exists along *both* attributes. Moreover, the thresholds are not constant, but vary as a function of the relative similarity of the options along the attributes, referred to as *interattribute sensitivity*. In our notation, the tradeoff functions proposed by Scholten and Read (2010, Equations 23–25) are

$$Q_{T|X}(\sigma T) = \begin{cases} \kappa\sigma T, & T \geq \varepsilon_{T|X}, \\ \kappa\sigma T/\vartheta_{T|X}, & T < \varepsilon_{T|X}, \end{cases} \quad (71)$$

and

$$Q_{X|T}(X) = \begin{cases} X, & X \geq \varepsilon_{X|T}, \\ X/\vartheta_{X|T}, & X < \varepsilon_{X|T}. \end{cases} \quad (72)$$

When outcomes are more dissimilar from one another than delays, the threshold along the time attribute ($\varepsilon_{T|X}$) increases, and effective time differences, particularly the small ones associated with subintervals, are more likely to be subthreshold, increasing the likelihood of a superadditive interval effect. Conversely, when delays are more dissimilar from one another than outcomes, the threshold along the outcome attribute ($\varepsilon_{X|T}$) increases, and effective outcome differences, particularly the small ones associated with subintervals, are more likely to be subthreshold, increasing the likelihood of a subadditive interval effect. Thus, delay preference for gains may be superadditive or subadditive in intervals, depending on whether the gains are large or small in absolute magnitude, and, therefore, more or less dissimilar from one another than the delays. Scholten and Read's (2010) account does not rely on a ratio-rule specification of the tradeoff model, and, therefore, covers indifference data as well as preference data. However, the proposal of interattribute sensitivity is quite unparsimonious, invoking two thresholds ($\varepsilon_{T|X}$ and $\varepsilon_{X|T}$) as latent *variables*.

The unified tradeoff model is more parsimonious, invoking one threshold (ε) as a latent *constant*, and yet it generates relative nonadditivity. A derivation is provided in Section 3 of the Appendix, but the intuition is as follows. When the subintervals are subthreshold, but the undivided interval is not, subthreshold hyposensitivity contributes to superadditivity of intervals, thereby countervailing time bias, which contributes to subadditivity of intervals. However, time bias is treated in proportion to outcomes, a joint implication of bias integration and diminishing absolute sensitivity, so its impact decreases as outcome magnitude increases. It may, therefore, outweigh subthreshold hyposensitivity for small gains, as manifested by subadditivity of intervals, but be outweighed by subthreshold hyposensitivity for large gains, as manifested by superadditivity of intervals (see Table 2, rows [{11, 12}, {14, 15}]).

In sum, the unified tradeoff model has relative nonadditivity as its natural implication, drawing on the same assumptions as it does to generate other preference patterns as well. This does not mean that interattribute sensitivity is an invalid proposition: It is quite intuitive that, as Mellers and Biagini (1994, p. 507) phrase it, "similarity on one dimension enhances differences on another dimension" (see also Tversky & Russo, 1969). However, the unified tradeoff model does not need similarity to accommodate relative nonadditivity, and so, by Occam's razor, it should not invoke similarity, because it can achieve the same scope without admitting greater complexity. Moreover, the canonical effect of similarity in decision making, the *comparability effect*, also flows naturally from a ratio-rule specification of the tradeoff model (see Scholten & Read, 2010, Figure 5).

The Robustness of Subadditivity in Gains

Evidence on nonadditivity of intervals in gains also shows an asymmetry in robustness that does not fit well with the symmetric notion of interattribute sensitivity. On the one hand, superadditive interval effects have been observed only under carefully crafted circumstances, which, according to the unified tradeoff model, is no wonder: Effective outcome differences must be large relative to effective time differences, *and* the effective time differences over the subintervals must be subthreshold, *and* subthreshold hyposensitivity must outweigh positive time bias over these intervals, *and* the effective time difference over the undivided interval must be suprathreshold, *or*, if it is not, subthreshold hyposensitivity must be outweighed by positive time bias over this interval. On the other hand, subadditive interval effects have been observed under most other circumstances, which, according to the unified tradeoff model, is also no wonder: Time bias will propagate across subintervals, and, thus, generate subadditivity of intervals if the aforementioned conditions for generating superadditivity are not satisfied. The proposition of interattribute sensitivity fails to address this preponderance of subadditivity over superadditivity.

In response to the robustness of subadditive interval effects, Scholten and Read (2010) proposed, as an alternative to the bipartite tradeoff functions over effective time and outcome differences, an S-shaped tradeoff function over effective time differences, convex over short, subthreshold effective time differences, and concave over longer, suprathreshold effective time differences, with the inflection point depending on effective outcome differences.[16] This solution, however, does not really have another motivation than conceding to the robustness of subadditive interval effects in gains. For the unified tradeoff model, it is just one more manifestation of time bias.

Most manifestations of time bias investigated in this article concern decisions about losses, with the diversity among decision makers in the sign of time bias for losses playing a part in many distinctive implications of the unified tradeoff model. The unanimity among decision makers in their positive time bias for gains does not nearly have the same heuristic value (Meyer, 1951), but the robustness of subadditivity is its clearest manifestation.

The Delay-Speedup Asymmetry Revisited

The unified tradeoff model contemplates not only the timing of outcomes, but also changes in the timing of outcomes. The delay-speedup asymmetry is that decision makers demand more compensation for delaying a gain, or speeding up a loss, than they offer for speeding up a gain, or delaying a loss. The interpretation of Scholten and Read's (2013) time-framing model is that decision makers experience delaying a gain, or speeding up a loss, as a *loss of time*, to which they are hypersensitive, and, conversely, that decision makers experience speeding up a gain, or delaying a loss, as a *gain of time*, to which they are hyposensitive. In

[16] See Ballard et al. (2023), Cheng and González-Vallejo (2016) and Scholten et al. (2014) for applications of the S-shaped tradeoff function, but without its dependence on effective outcome differences. Indeed, Scholten et al.'s (2014) purpose was to demonstrate that relative nonadditivity in preference data could originate from a ratio-rule specification of the tradeoff model.

our notation, the tradeoff function proposed by Scholten and Read (2013, Equations 45 and 47) is

$$Q(\sigma T) = \begin{cases} \kappa\eta_l^{\sigma\delta}\sigma T, & \sigma\delta = \ \ 1, \\ \kappa\eta_g^{\sigma\delta}\sigma T, & \sigma\delta = -1, \end{cases} \tag{73}$$

where $\kappa > 0$ is positive time preference, $\eta_l > 1$ is hypersensitivity to time lost when delaying a gain ($\sigma = 1, \delta = 1$) or speeding up a loss ($\sigma = -1, \delta = -1$), and $\eta_g > 1$ is hyposensitivity to time gained when speeding up a gain ($\sigma = 1, \delta = -1$) or delaying a loss ($\sigma = -1, \delta = 1$). When neither option is the status quo ($\delta = 0$), the tradeoff function reduces to $Q(\sigma T) = \kappa\sigma T$, which is, of course, the tradeoff function of the baseline tradeoff model in Equation 4.

The time-framing model invokes two multiplicative constants (η_l, η_g) operating on *positive time preference*, whereas the unified tradeoff model invokes one additive constant (q) operating on *the later outcome*, so, clearly, the former is less parsimonious than the latter. More critically, however, the multiplicative constants operate on positive time preference, meaning that the time-framing model entails universal delay tolerance in losses. The unified tradeoff model dissociates the delay-speedup asymmetry and its origin (status-quo bias) from other preference patterns and their origins (although all forces in the model interact). It thereby generates the delay-speedup asymmetry *in addition to* all other preference patterns it accommodates, and not *at the exclusion* of many.

Impatience Reduction

When choices between single-dated outcomes are transformed into choices involving outcome sequences, the general trend is one of a preference shift from the option with the time advantage to the option with the outcome advantage: Scholten et al.'s (2016) Table 6 shows such *impatience reduction* for choices involving only gains, and evidence on the sign irrelevance of common consequences shows that it extends to choices involving sequences of gains and losses. The front-end amount effect, one motivation in our model development, goes against the tide, and is, therefore, interesting from this perspective as well.

One explanation is that sequences reduce the salience of time (Jiang et al., 2014, 2017; Sun & Jiang, 2015): When choice involves only single-dated outcomes, the decision maker only needs to attend to two different outcomes occurring at two different delays, but, when choice also involves outcome sequences, the decision maker must track the *evolution* of outcomes over time, which may draw attention away from delays, and toward outcomes (Scholten et al., 2016). This may be a plausible explanation, but it is extraneous to all existing models of intertemporal choice.

The salience explanation is also extraneous to the unified tradeoff model, but our model does offer an alternative interpretation of impatience reduction. In most studies on preferences for sequences, the outcomes of both options are distributed over a few time periods, usually just two, corresponding to the delays of the *SS–LL* pair. It is precisely the *SS–LL* pair in which the options are most remote in time, because the duration of any outcome sequence will fall between these two end points, that is, $t_S < \hat{t} < t_L$. The effective time difference (T) will, therefore, be smaller when at least one option is an outcome sequence than when both options are single-dated outcomes. This increases the probability that the option with the time disadvantage will be favored over the option with the outcome disadvantage. Moreover, when the effective time difference falls below the person's threshold ($T < \varepsilon$), subthreshold hyposensitivity sets off, further increasing the impact of the reduction in the effective time difference between the options.

Like the similarity explanation of relative nonadditivity, the salience explanation of impatience reduction is quite intuitive, and we do not dismiss it as a valid proposition. However, the unified tradeoff model does not need salience to accommodate impatience reduction, and so, by Occam's razor, it should not invoke salience, because it can achieve the same scope without admitting greater complexity.

The Psychology of Paying for Consumption

A structural condition that the unified tradeoff model introduces to the analysis of choices involving outcome sequences is cumulative time weighing by absolute accumulated outcome value, by which duration can also be assessed for sequences of positive and negative outcomes. Such sequences are prevalent in investment decisions, for example, in the calculation of return on investment and net present value, but also in consumption decisions, where the "positive outcomes" are consumption events, while the "negative outcomes" are payments for consumption. P. Samuelson's (1937) discounted-utility model famously provided the first formal description of consumption decisions. Under his model, the consumer discounts money at the market rate of interest, so the consumer should prefer to defer payments as much as possible. Prelec and Loewenstein (1998), however, found that many of their participants preferred to pay *before* consumption rather than afterward, that is, they preferred prepayment to postpayment, in obvious violation of the discounted-utility model.

Prelec and Loewenstein (1998) developed a mental-accounting model of consumption decisions, in which payments and consumptions are discounted, but later consumption events buffer the pain of sooner payments, whereas later payments attenuate the pleasure of sooner consumption events. These "hedonic interactions" between payments and consumption contribute to a preference for prepayment (with sooner payments buffered) over postpayment (with sooner consumption events attenuated). To test their model, the authors conducted a rating study on preferences among schedules of payments and consumption events, specifically, payments of $1,000 and $2,000 (denoted k and K, respectively) for vacations of 1 week and 2 weeks (denoted w and W, respectively), distributed over a period of 4 consecutive years. A selection of four out of 16 schedules is shown in Table 13.

Participants considered the full set of 16 schedules and to provide preference ratings along an 11-point scale, reserving 0 for the worst schedule, and 10 for the best one. They dispersed into three

Table 13
Selected Stimuli and Results From Prelec and Loewenstein's (1998) in Table 4

	t (years)				Average rating		
Option	1	2	3	4	*A*	*B*	*C*
#1	K	k	w	W	6.8	1.2	0.8
#8			KW	kw	6.8	8.0	3.4
#9	WK	wk			3.9	6.7	6.7
#16	W	w	k	K	0.6	2.7	8.9

preference groups; Group *A* (52.3%) preferred prepayment (#1, #8) over postpayment (#9, #16); group *B* (26.7%) preferred to concentrate payments and vacations in the same year (#8, #9), rather than segregate cost and benefits over different years (#1, #16), which the mental-accounting model interprets as an indication of loss aversion aggravating negatively valued 1-year periods; and group *C* (20.9%) preferred postpayment (#9, #16) over prepayment (#1, #8), in accord with a conventional discounting model, to which the mental-accounting model reduces when hedonic interactions are inoperative.

The unified tradeoff model can address consumption decisions as well, drawing on the structural condition of cumulative time weighing by *absolute* accumulated outcome value, that is, the absolute value of positive *or* negative accumulated outcomes. With duration thus assessed, negative accumulated outcome value will, because of loss aversion, put a greater weight on a delay than positive accumulated outcome value of the same magnitude. Table 14 provides different scenarios of outcome valuation for the prepayment and postpayment schedules from the above inset.

In the case of the postpayment schedules, the first two periods are valued positively, whereas the last two periods may or may not be. Consumers who value all periods positively get their Money's Worth, because the total benefit of the vacations exceeds their total cost. Furthermore, consumers who value the last period negatively face a Bad Deal, because the total cost of the vacations exceeds their total benefit; among these consumers, loss aversion drives up the weight assigned to the last period, so that the Bad Deal has a longer duration than Money's Worth. Finally, consumers who value both of the last two periods negatively face a Rip-Off, because even 3 weeks of vacation are not worth the money spent on 1 week; among these consumers, loss aversion drives up the weights assigned to the last two periods, so that the Rip-Off has a longer duration than the Bad Deal.

In case of the prepayment schedules, the first two periods are valued negatively, while the last two periods may or may not be. Consumers who value all periods negatively face a Bad Deal, and loss aversion drives up the weights of all periods. Furthermore, consumers who value the last period positively get their Money's Worth; among these consumers, loss aversion drives up the weights of all but the fourth period, so that the duration of Money's Worth is shorter than the duration of the Bad Deal. Finally, consumers who value both of the last two periods positively gets a Bargain, because even 1 week of vacation is worth the money spent on 3 weeks; among these consumers, loss aversion drives up the weights of the first two periods, so that the duration of the Bargain is shorter than the duration of Money's Worth.

Given these scenarios, the unified tradeoff model interprets group *A* as participants who get their money's worth: Overall outcome value is positive, so that, by positive time preference, a shorter duration is preferred to a longer one, generating a preference for prepayment over postpayment. Conversely, our model interprets group *C* as participants who face a bad deal: Overall outcome value is negative, so that, by positive time preference, a longer duration is preferred to a shorter one, generating a preference for postpayment over prepayment.

Groups *A* and *C* act consistently with the structural condition of outcome valuation by outcome integration. In the specific context of consumption decisions, outcome integration is consistent with standard economic theory of consumer behavior, in which the decision to buy a good is not treated as a loss of money that must be compensated by the good, but, rather, as a choice between the good and other goods that the money could buy (Tversky & Kahneman, 1991). Moreover, segregation of payments and consumption events would be hedonically inefficient, especially in routine transactions (Thaler, 1985), which typically afford consumer surplus, that is, a positive net cost–benefit.

In contrast with groups *A* and *C*, participants in group *B* appear to integrate payments and consumption events that are temporally proximate (occurring within the same year), but *segregate* payments and consumption events that are temporally remote (occurring in different years). As just mentioned, segregation is hedonically inefficient, which would explain why this group gives very low ratings to schedules of temporally remote costs and benefits. As a matter of fact, these participants act consistently with

Table 14
Outcome Valuation of Prelec and Loewenstein's (1998) Postpayment and Prepayment Schedules Under the Unified Tradeoff Model

	t (years)			
	Outcome valuation of postpayment schedule (#16)			
	1	2	3	4
Conclusion	$v(W)$	$v(W + w)$	$v(W + w - \mu k)$	$v(W + w - \mu k - \mu K)$
Money's worth	+	+	+	+
Bad deal	+	+	+	−
Rip-off	+	+	−	−
	Outcome valuation of prepayment schedule (#1)			
Conclusion	$v(-\mu K)$	$v(-\mu K - \mu k)$	$v(-\mu K - \mu k + w)$	$v(-\mu K - \mu k + w + W)$
Bad deal	−	−	−	−
Money's worth	−	−	−	+
Bargain	−	−	+	+

Note. Negative accumulated outcome value will, in assessment of duration, put a greater weight on a delay than positive accumulated outcome value of the same magnitude: If we let $\mu > 0$ be a constant that assigns consumption units and monetary units to a common scale, loss aversion is that $-v(-\mu K) > v(W)$ if $\mu K = W$.

Thaler and Johnson's (1990) temporal-spacing principle of hedonic editing, under which temporal proximity makes outcome integration more likely, and temporal remoteness makes segregation more likely.

The above analysis is silent on time bias, because it is unidentifiable given the rating task that Prelec and Loewenstein (1998) administered: When no record is being kept of which options are compared to which while the task is being performed, it is impossible to assign time bias to any particular option. Yet, even without resorting to time bias, the unified tradeoff model provides a cogent explanation of Prelec and Loewenstein's (1998) results, although the role of temporal spacing in Group *B* is an extraneous factor to our model that merits closer attention in future developments. For instance, we do not yet know whether its role is restricted to same-different contrasts between days (Thaler & Johnson, 1990) or years (Prelec & Loewenstein, 1998), or whether its operation is more continuous.

The explanation given by the mental-accounting model is no less cogent, *but* it is tailored to consumption decisions: Hedonic interactions require sequences of both positive and negative outcomes, without which the mental-accounting model reduces to a standard discounting model, thus failing to accommodate most preference patterns associated with decisions in which the outcomes are of a single sign. The unified tradeoff model, in contrast, offers a broad coverage of preference patterns in intertemporal choice, *including*, as it now appears, preferences patterns in consumption decisions. The full statement of the model, provided in Table 11, covers such decisions as well if we let $x_i = y_i$ when the outcome in period i is a consumption good, or a bundle of consumption goods, and y_i is the reservation price put on it, meaning the maximum amount of money that the decision maker would be willing to pay for the good or the bundle. In case reservation prices are unavailable, money and consumption must be assigned to a common scale by a conversion parameter μ, as in Table 14.

Caveats

It should be noted that the unified tradeoff model, like any model of human behavior, omits many elements from its representation of intertemporal choice. For instance, we have not addressed "framing effects," meaning preference shifts caused by the redescription of options. Semantics can strongly influence the coding of outcomes and delays, and, therefore, choices made. Formal models like the unified tradeoff model operate on coded outcomes and delays, which leaves coding itself as a separate topic in the modeling of choice. For instance, Scholten et al. (2019) argued that a zero outcome acquires positive valence when the wording suggests that it could have been worse ("losing $0," "paying $0"), and negative valence when the wording suggests that it could have been better ("gaining $0," "receiving $0"). They showed that prospect theory and the mental-accounting model accurately predicted "mutable-zero effects" in risky and intertemporal choices, respectively, once the valence of zero outcomes was taken into account.

We also have not addressed the mixed evidence on the common-difference effect, mostly obtained in the restricted domain of monetary gains (see Footnote 3): This "core anomaly" is seen in some studies, but vanishes or even reverses in others. Reversals over (very) short time spans have been interpreted as manifestations of an *extended present* (Cavagnaro et al., 2016; Ebert & Prelec, 2007; Sayman & Öncüler, 2009; Takeuchi, 2011), meaning a time interval contiguous to the present that is treated as if it were the present, thus curtailing or even eroding the impact of positive time preference over that interval. For instance, in a choice between $7 today and $10 in 2 days ("the day after tomorrow"), the larger gain may fall inside the extended present, and will most likely be chosen, but, in a choice between $7 in 2 days and $10 in 4 days, the larger gain may no longer fall inside the extended present, so that the impact of positive time preference increases, and with it the probability that the smaller gain will be chosen. Thus, the introduction of a front-end delay induces a shift from *LL* to *SS*, the reverse of what is "supposed" to happen, no less according to the unified tradeoff model.

Formally, researchers have incorporated the extended present into discounting models by means of an inverse S-shaped discount function that is concave over the extended present, and convex beyond (Ebert & Prelec, 2007; Sayman & Öncüler, 2009; Takeuchi, 2011). Similarly, we might incorporate it into the unified tradeoff model by means of an S-shaped time-weighing function w that is convex over the extended present, and concave beyond. However, there are documented reversals of the common-difference effect (e.g., Attema et al., 2010; He et al., 2019) that cannot reasonably be ascribed to an extended present, unless we accept that the present is as extended as the data tell us it is, which amounts to restating the data.

General Discussion

We have restricted ourselves to a qualitative analysis of intertemporal choice, using scope as the criterion of model validity: The number of preference patterns that a model can account for, both in terms of their existence (e.g., delay aversion for losses) and in terms of the boundary conditions (e.g., the limits set by outcome magnitude and delay length on delay aversion for losses). We have shown that the unified tradeoff model offers by far the broadest coverage of preference patterns in intertemporal choice: All of those previously reported in the literature, many of which informed the formulation of our model, and a series of heretofore unreported patterns that we derived from our model, and confirmed across five studies. That said, the unified tradeoff model awaits specification as a computational model, and validation by criteria that are proper to quantitative analyses. We conclude by elaborating on both of these issues.

Probabilistic and Dynamic Choice

Busemeyer and Townsend (1993, p. 432) classified models of choice along two axes: deterministic versus probabilistic, and static versus dynamic. The unified tradeoff model is a *deterministic theory* in postulating "a binary preference relation that is either true or false for any pair of actions." In contrast, a *probabilistic theory* postulates "a probability function that maps each pair of actions into the closed interval [0, 1]." The unified tradeoff model is also a *static theory* in positing that "the preference relation (for deterministic models) or the probability function (for probabilistic models) is independent of the length of deliberation time." In contrast, a *dynamic theory* describes "how the preference relation or probability function changes as a function of deliberation time." The unified tradeoff model can be given a probabilistic and/or dynamic specification for quantitative analyses.

Dai and Busemeyer (2014) conducted model contests on *SS–LL* choices in the domain of gains, considering a range of candidate models, both alternative- and attribute-based, a range of probabilistic-dynamic specifications, and one probabilistic-dynamic specification. They found that the probabilistic-dynamic specification of the baseline tradeoff model performed best, not only in terms of overall fit to the data, but also in terms of detailed fit, that is, its ability to reproduce (a) the effects of three experimental manipulations (absolute outcome magnitude, relative delay length, and absolute delay length; see Dai & Busemeyer, 2014, Figures 5 and 7), and (b) an inverse U-shaped relation between the probability of *LL* choices and decision times (see Dai & Busemeyer, 2014, Figure 6 and 8).[17] Inspection of computational models on their detailed fit is a practice that merits broader adoption (see also Scholten & Read, 2013, Figures 2–7, and Scholten et al., 2014, Figures 3–5), and there is wealth of preference patterns on which these models are yet to be validated, first and foremost, those involving obligations to pay (uncompensated losses), and those involving payments for consumption (losses compensated by gains).

In the domain of risky choice, it has long been recognized that probabilistic specifications of deterministic models are not neutral regarding the patterns they produce, and are, therefore, an integral part of the theory that is being put to the test (Loomes & Sugden, 1995). The same is true in the domain of intertemporal choice, as we illustrate next.

Consider two *SS–LL* choices involving monetary gains: A choice from a Base pair, and a choice from an Alternative pair, in which the magnitude of both gains is increased by a common multiplicative constant. Further consider two probabilistic specifications: One in which choice probabilities derive from the difference between the "response strengths" (Luce, 1977, p. 224) of the options (the *difference rule*), and another in which choice probabilities derive from the ratio between the response strengths (the *ratio rule*). The tradeoff model will generate the absolute magnitude effect under either choice rule. For instance, if the baseline tradeoff model is the deterministic core, then $X_A/\kappa T > X_B/\kappa T \therefore X_A > X_B \therefore X_A - \kappa T > X_B - \kappa T$. Alternative-based models, however, do not ensure such ordinal consistency between the choice rules.

To illustrate, consider Loewenstein and Prelec's (1992) value-discounting model, which is the only alternative-based model that offers a principled account of the absolute magnitude effect. In this model, increasing the absolute magnitude of gains not only increases the value *difference* between a larger and a smaller gain, that is, $v(x_{L_A}) - v(x_{S_A}) > v(x_{L_B}) - v(x_{S_B})$, but also increases the value *ratio* between the gains, that is, $v(x_{L_A})/v(x_{S_A}) > v(x_{L_B})/v(x_{S_B})$. A ratio-rule specification of the model will generate the absolute magnitude effect, because increasing the value ratio amounts to increasing the *discounted*-value ratio: Letting $V = d(t)v(x)$ be the discounted value of (x, t), the model ensures that $V_{L_A}/V_{S_A} > V_{L_B}/V_{S_B}$. A difference-rule specification, however, will generate an absolute *or a reverse* magnitude effect, depending on whether the discounted-value difference in the Base pair favors *LL*, that is, $V_{L_A} - V_{S_A} > V_{L_B} - V_{S_B} > 0$, or disfavors *LL*, that is, $V_{L_A} - V_{S_A} < V_{L_B} - V_{S_B} < 0$, respectively.[18] This is just the tip of the iceberg, so we share Regenwetter et al.'s (2018) call for a comprehensive look at the role of probabilistic specifications in the modeling of intertemporal choice.

Parsimony, Scope, and Generalizability

Computational models can be evaluated on their validity by several criteria, each of which conceives of model fit (overall or detailed) in a certain way. One criterion is *descriptive adequacy* (Myung & Pitt, 2002), or the ability of a model to describe the data on which it was estimated, that is, to reproduce the data. Insofar as the data exhibit preference patterns within the scope of the deterministic core, and the probabilistic specification preserves the direction of the preference shifts generated by the deterministic core (see the above discussion), the criterion of descriptive adequacy is particularly useful to see whether a computational model reproduces well the size of the shifts. This means that the emphasis must be on the detailed fit, and not on the overall fit, of the computational model.

Evaluating model validity by descriptive adequacy has been widely criticized, because a model can be made to fare well just by increasing its complexity (the number of free parameters in the model, and the functional form combining the parameters). However, the more we increase model complexity to adequately describe a given data set with all of its idiosyncrasies, the less we are able to accurately describe other data sets with their own idiosyncrasies. The precision with which data are described must, therefore, be traded off against the parsimony of the model that describes the data, because a moderate complexity of the model best serves the goal of achieving *generalizability* (Myung & Pitt, 2002, Figure 11.4).

Drawing on Mosier (1951), a distinction can be made between generalizability to data from *similar* experimental designs, which is the cross-validation criterion, and generalizability to data from *dissimilar* experimental designs, which is the generalization criterion (Busemeyer & Wang, 2000). A central building block of the experimental design is the stimulus design, which, in judgment and decision-making research, is typically the design of option pairs. Whether or not two stimulus designs are "similar" is a matter of whether or not the option pairs in the validation design are sampled from the same population as those in the estimation design. For instance, two populations in which outcomes range from \$100 to \$400, delays range from 0 to 4 years, and interest rates range from −40% to +40%, might qualify as "similar" when both populations include single-dated outcomes of either sign. However, they qualify as "dissimilar" when one population includes only single-dated gains, while the other includes only single-dated losses. Busemeyer and Wang (2000) formally demonstrated that the generalization criterion is better than the cross-validation criterion in weeding out overly simple models, which are relatively bad at

[17] The effects of these manipulations are the absolute magnitude effect, the common-difference effect, and the *common-ratio effect* (Scholten & Read, 2010), which Dai and Busemeyer (2014) refer to as the "delay-duration effect:" Increasing both delays by the same multiplicative constant changes indifference into a preference for *SS* in gains, consistent with normative theory, because this manipulation decreases the interest rate implied by the options.

[18] Wulff and van den Bos (2018) reanalyzed Ericson et al.'s (2015) data, and found alternative-based models to perform better with a ratio rule than with a difference rule. Oddly, they specified attribute-based models only with a difference rule, even though Scholten et al.'s (2014) evidence on relative nonadditivity had shown that the ratio rule is vital to the performance of these models. Wulff and van den Bos (2018) found ratio-rule specifications of alternative-based models to perform about as well as difference-rule specifications of attribute-based models. However, in view of the failure to specify attribute-based models with a ratio rule, their conclusion that "the jury is still out on which model—or which type of model—is best" (p. 1890) must be regarded premature.

capturing regularities, and overly complex models, which are relatively bad at filtering out idiosyncrasies.

Application of the generalization criterion is nontrivial. Consider two populations that differ only in outcome sign. Estimating a probabilistic specification of the unified tradeoff model on choices between single-dated gains, and validating the estimated model on choices between single-dated losses, is inappropriate for at least two reasons. First, people are unanimous in their positive time bias for gains, but diverse in the sign of time bias for losses, so, for many, the sign of time bias reverses in the transition from gains to losses; neglecting this complication inevitably compromises the performance of the model on the generalization test. Second, many people are loss averse, which tempers the contribution of positive time preference to delay tolerance for losses (see Equation 47), so neglecting loss aversion further compromises the performance of the model on the generalization test.

We can conceive of two strategies to deal with this. The *downward strategy* is to use the stimulus design that requires a more complete specification of the model (the loss design) in the estimation phase, and to use the stimulus design that requires a less complete specification of the model (the gain design) in the validation phase. The *upward strategy* is to run the generalization test in the reverse direction, but on the provision that the parameters omitted from the estimation phase (the sign of time bias for losses, the strength of loss aversion) be estimated in the validation phase instead. It would appear prudent to run generalization tests in both directions, and check on the implications for test results.

In addition to the ability of a model to accurately predict choice behavior across similar or dissimilar designs, there is the issue of whether the model derives its predictions from stable parameter estimates, that is, whether parameter estimates are consistent across experimental designs (Ballard et al., 2023; Yechiam & Busemeyer, 2008). Parameters can represent many things, and parameter estimates may be meaningful either because they are consistent across, or because they are contingent on, the circumstances under which they are obtained.

A core premise in our qualitative analysis has been that time preference is of positive sign across decision makers, and of constant strength for each individual decision maker. Therefore, estimates of time preference would be meaningful if they were consistent across experimental designs, and constant estimates of this parameter would bolster our belief that the model under investigation is a valid representation of choice behavior. Indeed, constant estimates of *any* parameter would be reassuring, because nonconstant estimates mean that the model itself changes as a function of the reality it represents, and mimicry is not quite the hallmark of generality. Meanwhile, people themselves may be accused of engaging in mimicry. For instance, the interest rates people demand or offer vary tremendously as a function of whether delays are stated as days, months, or years (see Scholten et al., 2014, Footnote 10).

This said, fully consistent estimates of computational models may be hard to achieve: A model is by definition a partial representation of reality, and is, therefore, unequipped to deal with all possible reactions to experimental designs. In Study 5, we carefully crafted different loss tasks across which we expected time bias to be consistent in sign, and the results suggested that it was. In Studies 3 and 4, however, we always administered loss tasks prior to gain tasks, informed (or alarmed) as we were by Hardisty et al.'s (2013, Study 3) finding that average delay aversion for losses was weaker when loss tasks were administered last, but average delay aversion for gains did not vary with task order. As interpreted by the unified tradeoff model, this *asymmetric task-order effect* may be an indication that, (a) among individuals with bipolar time bias, the sign of time bias for losses reverses from negative to positive once positive time bias for gains has been activated, or that (b) the strength of negative time bias among these individuals decreases upon activation of positive time bias for gains. In any event, counterbalancing loss and gain tasks would deteriorate the performance of the unified tradeoff model in a generalization test when specified with time bias that is constant in sign and strength for each individual participant. The application of the unified tradeoff model in quantitative analyses of intertemporal choice may tell us a lot more about time bias and its vicissitudes.

References

Abdellaoui, M., Attema, A. E., & Bleichrodt, H. (2010). Intertemporal tradeoffs for gains and losses: An experimental measurement of discounted utility. *Economic Journal*, *120*(545), 845–866. https://doi.org/10.1111/j.1468-0297.2009.02308.x

Abdellaoui, M., Bleichrodt, H., & l'Haridon, O. (2013). Sign-dependence in intertemporal choice. *Journal of Risk and Uncertainty*, *47*(3), 225–253. https://doi.org/10.1007/s11166-013-9181-9

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*(10), 1659–1674. https://doi.org/10.1287/mnsc.1070.0711

Amasino, D. R., Sullivan, N. J., Kranton, R. E., & Huettel, S. A. (2019). Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour*, *3*(4), 383–392. https://doi.org/10.1038/s41562-019-0537-2

Appelt, K. C., Hardisty, D. J., & Weber, E. U. (2011). Asymmetric discounting of gains and losses: A query theory account. *Journal of Risk and Uncertainty*, *43*(2), 107–126. https://doi.org/10.1007/s11166-011-9125-1

Arfer, K. B., & Luhmann, C. C. (2015). The predictive accuracy of intertemporal-choice models. *British Journal of Mathematical & Statistical Psychology*, *68*(2), 326–341. https://doi.org/10.1111/bmsp.12049

Arieli, A., Ben-Ami, Y., & Rubinstein, A. (2011). Tracking decision makers under uncertainty. *American Economic Journal. Microeconomics*, *3*(4), 68–76. https://doi.org/10.1257/mic.3.4.68

Attema, A. E., Bleichrodt, H., Rohde, K. I. M., & Wakker, P. P. (2010). Time-tradeoff sequences for analyzing discounting and time inconsistency. *Management Science*, *56*(11), 2015–2030. https://doi.org/10.1287/mnsc.1100.1219

Baker, F., Johnson, M. W., & Bickel, W. K. (2003). Delay discounting in current and never-before cigarette smokers: Similarities and differences across commodity, sign, and magnitude. *Journal of Abnormal Psychology*, *112*(3), 382–392. https://doi.org/10.1037/0021-843X.112.3.382

Ballard, T., Luckman, A., & Konstantinidis, E. (2023). A systematic investigation into the reliability of inter-temporal choice model parameters. *Psychonomic Bulletin and Review*, *30*, 1294–1322. https://doi.org/10.3758/s13423-022-02241-7

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961–981. https://doi.org/10.1287/opre.30.5.961

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, *69*(2), 205–223. https://doi.org/10.1016/j.geb.2009.11.003

Benzion, U., Rapoport, A., & Yagil, J. (1989). Discount rates inferred from decisions: An experimental study. *Management Science*, *35*(3), 270–284. https://doi.org/10.1287/mnsc.35.3.270

Bernoulli, D. (1954). Specimen theoriae novae de mensura sortis. Commentarii academiae scientiarum imperialis petropolitanae [Exposition of a new theory on the measurement of risk] (D. Bernoulli, Trans.). *Econometrica*, *22*(1), 23–36. https://doi.org/10.2307/1909829 (Original work published 1738)

Brañas-Garza, P., Jorrat, D., Espín, A. M., & Sánchez, A. (2023). Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics*, *26*(2), 412–434. https://doi.org/10.1007/s10683-022-09776-5

Breuer, W., Soypak, C. K., & Steininger, B. I. (2023). Conventional or reverse magnitude effect for negative outcomes: A matter of framing. *Review of Financial Economics*. Advance online publication. https://doi.org/10.1002/rfe.1190

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. https://doi.org/10.1037/0033-295X.100.3.432

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. https://doi.org/10.1006/jmps.1999.1282

Camerer, C. F., & Ho, T. H. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, *8*(2), 167–196. https://doi.org/10.1007/BF01065371

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. https://doi.org/10.1007/s11166-016-9242-y

Cheng, J., & González-Vallejo, C. (2016). Attribute-wise vs. alternative-wise mechanism in intertemporal choice: Testing the proportional difference, trade-off, and hyperbolic models. *Decision*, *3*(3), 190–215. https://doi.org/10.1037/dec0000046

Cohen, J., Ericson, K. M., Laibson, D., & White, J. M. (2020). Measuring time preferences. *Journal of Economic Literature*, *58*(2), 299–347. https://doi.org/10.1257/jel.20191074

Dai, J. (2017). Are intertemporal preferences transitive? A Bayesian analysis of repeated individual intertemporal choices. *Decision*, *4*, 1–24. https://doi.org/10.1037/dec0000054

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. https://doi.org/10.1037/a0035976

Ebert, J. E. J., & Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of the near and far future. *Management Science*, *53*(9), 1423–1438. https://doi.org/10.1287/mnsc.1060.0671

Ericson, K. M. M., White, J. M., Laibson, D., & Cohen, J. D. (2015). Money earlier or later? Simple heuristics explain intertemporal choices better than delay discounting does. *Psychological Science*, *26*(6), 826–833. https://doi.org/10.1177/0956797615572232

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

Fishburn, P. C. (1974). Lexicographic orders, utilities and decision rules: A survey. *Management Science*, *20*(11), 1442–1471. https://doi.org/10.1287/mnsc.20.11.1442

Fisher, I. (1930). *The theory of interest*. Macmillan.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal Economic Literature*, *40*(2), 351–401. https://doi.org/10.1257/002205102320161311

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review*, *109*(1), 137–155. https://doi.org/10.1037/0033-295X.109.1.137

Guyse, J. L., & Simon, J. (2011). Consistency among elicitation techniques for intertemporal choice: A within-subjects investigation of the anomalies. *Decision Analysis*, *8*(3), 233–246. https://doi.org/10.1287/deca.1110.0212

Hardisty, D. J., Appelt, K. C., & Weber, E. U. (2013). Good or bad, we want it now: Fixed-cost present for gains and losses explains magnitude asymmetries in intertemporal choice. *Journal of Behavioral Decision Making*, *26*(4), 348–361. https://doi.org/10.1002/bdm.1771

Hardisty, D. J., & Weber, E. U. (2020). Impatience and savoring vs. dread: Asymmetries in anticipation explain consumer time preferences for positive vs. negative events. *Journal of Consumer Psychology*, *30*(4), 598–613. https://doi.org/10.1002/jcpy.1169

He, L., Golman, R., & Bhatia, S. (2019). Variable time preference. *Cognitive Psychology*, *111*(12), 53–79. https://doi.org/10.1016/j.cogpsych.2019.03.003

He, L., Zhao, W. J., & Bhatia, S. (2022). An ontology of decision models. *Psychological Review*, *129*(1), 49–72. https://doi.org/10.1037/rev0000231

Hirshleifer, J. (1958). On the theory of optimal investment decision. *Journal of Political Economy*, *66*(4), 329–352. https://doi.org/10.1086/258057

Holt, D. D., Green, L., Myerson, J., & Estle, S. J. (2008). Preference reversal with losses. *Psychonomic Bulletin and Review*, *15*(1), 89–95. https://doi.org/10.3758/PBR.15.1.89

Jiang, C.-M., Hu, F.-P., & Zhu, L.-F. (2014). Introducing upfront losses as well as gains decreases impatience in intertemporal choices with rewards. *Judgment and Decision Making*, *9*(4), 297–302. https://doi.org/10.1017/S1930297500006173

Jiang, C.-M., Sun, H.-M., Zhu, L.-F., Zhao, L., Liu, H.-Z., & Sun, H.-Y. (2017). Better is worse, worse is better: Reexamination of violations of dominance in intertemporal choice. *Judgment and Decision Making*, *12*(3), 253–259. https://doi.org/10.1017/S1930297500005866

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Keren, G., & Teigen, K. H. (2004). Yet another look at the heuristics and biases approach. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 89–109). Blackwell Publishing. https://doi.org/10.1002/9780470752937.ch5

Kinari, Y., Ohtake, F., & Tsutsui, Y. (2009). Time discounting: Declining impatience and interval effect. *Journal of Risk and Uncertainty*, *39*, 87–112. https://doi.org/10.1007/s11166-009-9073-1

Leland, J. W. (2002). Similarity judgments and anomalies in intertemporal choice. *Economic Inquiry*, *40*(4), 574–581. https://doi.org/10.1093/ei/40.4.574

Linville, P. W., & Fischer, G. W. (1991). Preferences for separating or combining events. *Journal of Personality and Social Psychology*, *60*(1), 5–23. https://doi.org/10.1037/0022-3514.60.1.5

Liu, H. Z., Lyu, X. K., Wei, Z. H., Mo, W. L., Luo, J. R., & Su, X. Y. (2021). Exploiting the dynamics of eye gaze to bias intertemporal choice. *Journal of Behavioral Decision Making*, *34*(3), 419–431. https://doi.org/10.1002/bdm.2219

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. *Economic Journal*, *97*(387), 666–684. https://doi.org/10.2307/2232929

Loewenstein, G. (1988). Frames of mind in intertemporal choice. *Management Science*, *34*(2), 200–214. https://doi.org/10.1287/mnsc.34.2.200

Loewenstein, G. (1992). The fall and rise of psychological explanations in the economics of intertemporal choice. In G. Loewenstein & J. Elster (Eds.), *Choice over time* (pp. 5–34). SAGE Publications.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics*, *107*(2), 573–597. https://doi.org/10.2307/2118482

Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, *100*(1), 91–108. https://doi.org/10.1037/0033-295X.100.1.91

Loomes, G. (2010). Modeling choice and valuation in decision experiments. *Psychological Review*, *117*(3), 902–924. https://doi.org/10.1037/a0019807

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*(368), 805–824. https://doi.org/10.2307/2232669

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review*, *39*(3–4), 641–648. https://doi.org/10.1016/0014-2921(94)00071-7

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. Wiley.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of Mathematical Psychology*, *15*(3), 215–233. https://doi.org/10.1016/0022-2496(77)90032-3

Luckman, A., Donkin, C., & Newell, B. R. (2017). People wait longer when the alternative is risky: The relation between preferences in risky and intertemporal choice. *Journal of Behavioral Decision Making*, *30*(5), 1078–1092. https://doi.org/10.1002/bdm.2025

Luckman, A., Donkin, C., & Newell, B. R. (2020). An evaluation and comparison of models of risky intertemporal choice. *Psychological Review*, *127*(6), 1097–1138. https://doi.org/10.1037/rev0000223

McAlvanah, P. (2010). Subadditivity, patience, and utility: The effects of dividing time intervals. *Journal of Economic Behavior & Organization*, *76*(2), 325–337. https://doi.org/10.1016/j.jebo.2010.06.009

Mellers, B. A., & Biagini, K. (1994). Similarity and choice. *Psychological Review*, *101*(3), 505–518. https://doi.org/10.1037/0033-295X.101.3.505

Meyer, H. (1951). On the heuristic value of scientific models. *Philosophy of Science*, *18*(2), 111–123. https://doi.org/10.1086/287137

Mischel, W., & Metzner, R. (1962). Preference for delayed reward as a function of age, intelligence, and length of delay interval. *Journal of Abnormal and Social Psychology*, *64*(6), 425–431. https://doi.org/10.1037/h0045046

Mosier, C. I. (1951). Problems and designs of cross-validation. *Educational and Psychological Measurement*, *11*(1), 5–11. https://doi.org/10.1177/001316445101100101

Myerson, J., Baumann, A. A., & Green, L. (2017). Individual differences in delay discounting: Differences are quantitative with gains, but qualitative with losses. *Journal of Behavioral Decision Making*, *30*(2), 359–372. https://doi.org/10.1002/bdm.1947

Myung, J., & Pitt, M. A. (2002). Mathematical modeling. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology. Methodology in experimental psychology* (pp. 429–460). Wiley.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552. https://doi.org/10.1037/0278-7393.14.3.534

Prelec, D., & Loewenstein, G. (1991). Decision making over time and under uncertainty: A common approach. *Management Science*, *37*(7), 770–786. https://doi.org/10.1287/mnsc.37.7.770

Prelec, D., & Loewenstein, G. (1998). The red and the black: Mental accounting of savings and debt. *Marketing Science*, *17*(1), 4–28. https://doi.org/10.1287/mksc.17.1.4

Rao, L. L., & Li, S. (2011). New paradoxes in intertemporal choice. *Judgment and Decision Making*, *6*(2), 122–129. https://doi.org/10.1017/S193029750000406X

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty*, *23*(1), 5–32. https://doi.org/10.1023/A:1011198414683

Read, D., Frederick, S., & Scholten, M. (2013). DRIFT: An analysis of outcome framing in intertemporal choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 573–588. https://doi.org/10.1037/a0029177

Read, D., Olivola, C., & Hardisty, D. J. (2017). The value of nothing: Asymmetric attention to opportunity costs drives intertemporal decision making. *Management Science*, *63*(12), 4277–4297. https://doi.org/10.1287/mnsc.2016.2547

Read, D., & Roelofsma, P. H. M. P. (2003). Subadditive versus hyperbolic discounting: A comparison of choice and matching. *Organizational Behavior and Human Decision Processes*, *91*(2), 140–153. https://doi.org/10.1016/S0749-5978(03)00060-8

Read, D., & Scholten, M. (2012). Tradeoffs between sequences: Weighing accumulated outcomes against outcome-adjusted delays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1675–1688. https://doi.org/10.1037/a0028216

Read, D., & Scholten, M. (2017). Future-oriented decisions: Intertemporal choice. In R. Ranyard (Ed.), *Economic psychology* (pp. 35–50). Wiley. https://doi.org/10.1002/9781118926352.ch3

Reeck, C., Wall, D., & Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(45), 11890–11895. https://doi.org/10.1073/pnas.1707040114

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. https://doi.org/10.1037/dec0000069

Roelofsma, P. H. M. P., & Read, D. (2000). Intransitive intertemporal choice. *Journal of Behavioral Decision Making*, *13*(2), 161–177. https://doi.org/10.1002/(SICI)1099-0771(200004/06)13:2<161::AID-BDM348>3.0.CO;2-P

Rosenbaum, D. A., Fournier, L. R., Levy-Tzedek, S., McBride, D. M., Rosenthal, R., Sauerberger, K., VonderHaar, R. L., Wasserman, E. A., & Zentall, T. R. (2019). Sooner rather than later: Precrastination rather than procrastination. *Current Directions in Psychological Science*, *28*(3), 229–233. https://doi.org/10.1177/0963721419833652

Rubinstein, A. (2003). "Economics and psychology"? The case of hyperbolic discounting. *International Economic Review*, *44*(4), 1207–1216. https://doi.org/10.1111/1468-2354.t01-1-00106

Rubinstein, A. (2006). Dilemmas of an economic theorist. *Econometrica*, *74*(4), 865–883. https://doi.org/10.1111/j.1468-0262.2006.00689.x

Samuelson, P. (1937). A note on measurement of utility. *The Review of Economic Studies*, *4*(2), 155–161. https://doi.org/10.2307/2967612

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical Association*, *46*(253), 55–67. https://doi.org/10.1080/01621459.1951.10500768

Sayman, S., & Öncüler, A. (2009). An investigation of time inconsistency. *Management Science*, *55*(3), 470–482. https://doi.org/10.1287/mnsc.1080.0942

Scholten, M., & Read, D. (2006). Discounting by intervals: A generalized model of intertemporal choice. *Management Science*, *52*(9), 1426–1438. https://doi.org/10.1287/mnsc.1060.0534

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review*, *117*(3), 925–944. https://doi.org/10.1037/a0019619

Scholten, M., & Read, D. (2013). Time and outcome framing in intertemporal tradeoffs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1192–1212. https://doi.org/10.1037/a0031171

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. https://doi.org/10.1111/cogs.12104

Scholten, M., Read, D., & Sanborn, A. (2016). Cumulative weighing of time in intertemporal tradeoffs. *Journal of Experimental Psychology: General*, *145*(3), 1177–1205. https://doi.org/10.1037/xge0000198

Scholten, M., Read, D., & Stewart, N. (2019). The framing of nothing and the psychology of choice. *Journal of Risk and Uncertainty*, *59*(2), 125–149. https://doi.org/10.1007/s11166-019-09313-5

Stevens, J. R. (2016). Intertemporal similarity: Discounting as a last resort. *Journal of Behavioral Decision Making*, *29*(1), 12–24. https://doi.org/10.1002/bdm.1870

Sun, H.-Y., & Jiang, C.-M. (2015). Introducing money at any time can reduce discounting in intertemporal choices with rewards: An extension of the upfront money effect. *Judgment and Decision Making*, *10*(6), 564–570. https://doi.org/10.1017/S1930297500007002

Takeuchi, K. (2011). Non-parametric test of time consistency: Present bias and future bias. *Games and Economic Behavior*, *71*(2), 456–478. https://doi.org/10.1016/j.geb.2010.05.005

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, *8*(3), 201–207. https://doi.org/10.1016/0165-1765(81)90067-7

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, *4*(3), 199–214. https://doi.org/10.1287/mksc.4.3.199

Thaler, R., & Johnson, E. J. (1990). Gambling with the house money effect and trying to break even: The effects of prior outcomes on risky choice. *Management Science*, *36*(6), 643–660. https://doi.org/10.1287/mnsc.36.6.643

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *107*(4), 1039–1061. https://doi.org/10.2307/2937956

Tversky, A., & Russo, J. (1969). Substitutability and similarity in binary choice. *Journal of Mathematical Psychology*, *6*(1), 1–12. https://doi.org/10.1016/0022-2496(69)90027-3

Wulff, D. U., & van den Bos, W. (2018). Modeling choices in delay discounting. *Psychological Science*, *29*(11), 1890–1894. https://doi.org/10.1177/0956797616664342

Yates, J. F., & Watts, R. A. (1975). Preference for deferred losses. *Organizational Behavior and Human Performance*, *13*(2), 294–306. https://doi.org/10.1016/0030-5073(75)90051-3

Yechiam, E., & Busemeyer, J. R. (2008). Evaluating generalizability and parameter consistency in learning models. *Games and Economic Behavior*, *63*(1), 370–394. https://doi.org/10.1016/j.geb.2007.08.011

Zauberman, G., Kim, B. K., Malkoc, S. A., & Bettman, J. R. (2009). Discounting time and time discounting: Subjective time perception and intertemporal preferences. *JMR: Journal of Marketing Research*, *46*(4), 543–556. https://doi.org/10.1509/jmkr.46.4.543

Appendix

Additivity, Nonadditivity, and Relative Nonadditivity of Intervals

Suppose that an undivided interval, $0 \rightarrow 2t$, is divided into two subintervals, $0 \rightarrow t$ and $t \rightarrow 2t$, and that three indifference points are obtained: One for the undivided interval, $(x, 0) \sim (\tilde{x}_{2t}, 2t)$, and two for the subintervals, $(x, 0) \sim (\tilde{x}_t, t)$ and $(\tilde{x}_t, t) \sim (\tilde{\tilde{x}}_{2t}, 2t)$. The two indifference points for the subintervals are obtained by means of a *chained matching procedure*, in which the later outcome that has been matched to the sooner outcome over the first subinterval $(\tilde{x}_t)$ becomes the sooner outcome to which the later outcome must be matched over the second subinterval $(\tilde{\tilde{x}}_{2t})$. In this Appendix, we derive three results regarding delay preference under the chained matching procedure.

In Section 1, we derive additivity of intervals from alternative-based intertemporal choice. In Section 2, we derive the implications of the unified tradeoff model for nonadditivity of intervals in monetary losses: Under unipolar time bias, the model generates either subadditivity or superadditivity of intervals, depending on whether time bias outweighs, or is outweighed by subthreshold hyposensitivity; under bipolar time bias, the model unconditionally generates superadditivity of intervals. In Section 3, we derive relative nonadditivity from the unified tradeoff model.

Section 1: Additivity of Intervals

Delay preference is additive in intervals when its strength does not depend on subdivision of an interval, that is, $\tilde{x}_{2t} = \tilde{\tilde{x}}_{2t}$. Alternative-based models entail additivity of intervals. To show this under the chained matching procedure, consider Loewenstein and Prelec's (1992) model of value discounting by delays, which describes the indifference points as

$$d(2t)v(\tilde{x}_{2t}) = v(x) \text{ for } 0 \rightarrow 2t, \tag{A1}$$

$$d(t)v(\tilde{x}_t) = v(x) \text{ for } 0 \rightarrow t, \text{ and} \tag{A2}$$

$$d(2t)v(\tilde{\tilde{x}}_{2t}) = d(t)v(\tilde{x}_t) \text{ for } t \rightarrow 2t. \tag{A3}$$

Solving each indifference point for the later outcome,

$$\tilde{x}_{2t} = v^{-1}[v(x)/d(2t)], \tag{A4}$$

$$\tilde{x}_t = v^{-1}[v(x)/d(t)], \text{ and} \tag{A5}$$

$$\tilde{\tilde{x}}_{2t} = v^{-1}[d(t)v(\tilde{x}_t)/d(2t)]. \tag{A6}$$

Substituting Equation A5 into Equation A4,

$$\tilde{\tilde{x}}_{2t} = v^{-1}[d(t)v(v^{-1}[v(x)/d(t)])/d(2t)], \text{ or} \tag{A7}$$

$$\tilde{\tilde{x}}_{2t} = v^{-1}[v(x)/d(2t)], \tag{A8}$$

meaning that $\tilde{\tilde{x}}_{2t} = \tilde{x}_t$.

The fixed-cost discounting model entails additivity of intervals as well, which describes the indifference points as

$$d(2t)\tilde{x}_{2t} - c = x \text{ for } 0 \rightarrow 2t, \tag{A9}$$

$$d(t)\tilde{x}_t - c = x \text{ for } 0 \rightarrow t, \text{ and} \tag{A10}$$

$$d(2t)\tilde{\tilde{x}}_{2t} = d(t)\tilde{x}_t \text{ for } t \rightarrow 2t. \tag{A11}$$

Solving Equation A10 for $\tilde{x}_t$, and substituting it into Equation A11, yields Equation A9 with $\tilde{x}_{2t} = \tilde{\tilde{x}}_{2t}$, which is additivity of intervals.

Section 2: Nonadditivity of Intervals in Losses

Delay preference is superadditive/subadditive in intervals when $\tilde{x}_t$ deviates more/less from the neutral outcome than $\tilde{\tilde{x}}_{2t}$, that is, when $\tilde{x}_{2t} \gtrless \tilde{\tilde{x}}_{2t}$ for $x > 0$, and $\tilde{x}_{2t} \lessgtr \tilde{\tilde{x}}_{2t}$ for $x < 0$. We assume, for simplicity, but also motivated by evidence (see McAlvanah, 2010; Read, 2001; Scholten & Read, 2006), that the decision maker exhibits constant sensitivity to delays, that is, $w(t) = t$, as a result of which the effective subintervals are of the same quantity, that is, $T = w(t) - w(0) = w(2t) - w(t) = t$.

(*Appendix continues*)

Unipolar Time Bias

On the assumption that the undivided interval is suprathreshold, a decision maker who operates under unipolar time bias will be indifferent between a sooner and a later loss when

$$-\kappa 2T = v(\tilde{x}_{2t} + c) - v(x) \text{ for } 0 \rightarrow 2t, \tag{A12}$$

$$-\kappa T/\vartheta = v(\tilde{x}_t + c) - v(x) \text{ for } 0 \rightarrow t, \text{ and} \tag{A13}$$

$$-\kappa T/\vartheta = v(\tilde{\tilde{x}}_{2t} + c) - v(\tilde{x}_t) \text{ for } t \rightarrow 2t. \tag{A14}$$

Where $x < 0$. Solving each indifference point for the later outcome,

$$\tilde{x}_{2t} = v^{-1}[-\kappa 2T + v(x)] - c, \tag{A15}$$

$$\tilde{x}_t = v^{-1}[-\kappa T/\vartheta + v(x)] - c, \text{ and} \tag{A16}$$

$$\tilde{\tilde{x}}_{2t} = v^{-1}[-\kappa T/\vartheta + v(\tilde{x}_t)] - c. \tag{A17}$$

Substituting Equation A16 into Equation A17,

$$\tilde{\tilde{x}}_{2t} = v^{-1}[-\kappa T/\vartheta + v(v^{-1}[-\kappa T/\vartheta + v(x)] - c)] - c. \tag{A18}$$

Delay preference for losses is superadditive/subadditive in intervals when $\tilde{x}_t$ is more/less negative than $\tilde{\tilde{x}}_{2t}$, that is, $\tilde{x}_{2t} \lesseqgtr \tilde{\tilde{x}}_{2t}$ for $x < 0$, which, under unipolar time bias, means that delay tolerance is stronger/weaker over the undivided interval than over the divided one:

$$v^{-1}[-\kappa 2T + v(x)] - c \lesseqgtr$$
$$v^{-1}[-\kappa T/\vartheta + v(v^{-1}[-\kappa T/\vartheta + v(x)] - c)] - c, \text{ or} \tag{A19}$$

$$-\kappa 2T + v(x) \lesseqgtr \kappa T/\vartheta + v(v^{-1}[-\kappa T/\vartheta + v(x)] - c). \tag{A20}$$

In the absence of time bias ($c = 0$), but presence of subthreshold hyposensitivity ($\vartheta > 1$), Inequality Equation A20 reduces to

$$1/\vartheta < 1, \tag{A21}$$

and delay preference is superadditive in intervals. Conversely, in the absence of subthreshold hyposensitivity ($\vartheta = 1$), but presence of time bias ($c > 0$), Inequality Equation 20 reduces to

$$-\kappa T + v(x) > v(v^{-1}[-\kappa T + v(x)] - c), \tag{A22}$$

and delay preference is subadditive in intervals. Thus, under unipolar time bias, the unified tradeoff model generates either subadditivity or superadditivity of intervals, depending on whether time bias outweighs, or is outweighed by, subthreshold hyposensitivity.

Bipolar Time Bias

Whereas unipolar time bias operates in favor of the later loss, that is, $v(x + c)$, bipolar time bias operates against it, that is, $v(x - c)$. Because the impact of time bias reverses, Inequality Equation A22 reverses into

$$-\kappa T + v(x) < v(v^{-1}[-\kappa T + v(x)] + c), \tag{A23}$$

and delay preference is superadditive. Thus, bipolar time bias operates in the same direction as subthreshold hyposensitivity, toward superadditivity of intervals, so that, under this variant of time bias, the unified tradeoff model unconditionally generates superadditivity of intervals. However, bipolar time bias operates in the opposite direction to positive time preference: When positive time preferences prevails, as manifested by delay tolerance, superadditivity of intervals refers to a stronger delay tolerance over the undivided interval than over the divided one ($\tilde{x}_{2t} < \tilde{\tilde{x}}_{2t} < x < 0$), but, when bipolar time bias prevails, as manifested by delay aversion, superadditivity of intervals refers to a *weaker* delay aversion over the undivided interval than over the divided one ($x < \tilde{x}_{2t} < \tilde{\tilde{x}}_{2t} < 0$).

Section 3: Relative Nonadditivity of Intervals

By diminishing absolute sensitivity to outcomes, that is, the concavity of the value function over gains, and its convexity over losses, and bias integration, time bias is treated in proportion to outcomes, implying that, as the magnitude of the outcomes increases, the impact of time bias decreases, and the impact of subthreshold hyposensitivity correspondingly increases. Thus, when time bias contributes to subadditivity of intervals, as it does for monetary gains under either variant of time bias, and for monetary losses under unipolar time bias, delay preference may be subadditive in intervals when the outcomes are small, but superadditive in intervals when the outcomes are large, which is relative nonadditivity.

Inequality Equation A20 describes the situation for monetary losses under unipolar time bias. For monetary gains, time bias universally operates against the later gain, that is, $v(x - c)$, and so delay preference will be superadditive or subadditive in intervals when

$$\kappa 2T + v(x) \gtreqless \kappa T/\vartheta + v(v^{-1}[\kappa T/\vartheta + v(x)] + c). \tag{A24}$$

In the absence of time bias ($c = 0$), but presence of subthreshold hyposensitivity ($\vartheta > 1$), Inequality Equation A24 reduces to

$$1 > 1/\vartheta, \tag{A25}$$

and delay preference is superadditive in intervals. Conversely, in the absence of subthreshold hyposensitivity ($\vartheta = 1$), but presence of time bias ($c > 0$), Inequality Equation A24 reduces to

$$\kappa T + v(x) < v(v^{-1}[\kappa T + v(x)] + c), \tag{A26}$$

and delay preference is subadditive in intervals. However, as x goes to infinity, c vanishes, and so does subadditivity. With c vanishing from Inequality Equation A24, nonadditivity of intervals derives only from subthreshold hyposensitivity, and delay preference is, therefore, superadditive in intervals. This reversal is relative nonadditivity.

Received November 11, 2022
Revision received September 29, 2023
Accepted October 7, 2023 ■