2016

Intertemporal similarity: Discounting as a last resort

Jeffrey R. Stevens
University of Nebraska-Lincoln, jstevens5@unl.edu

Intertemporal similarity: Discounting as a last resort

Jeffrey R. Stevens
Max Planck Institute for Human Development
and
University of Nebraska-Lincoln

Abstract

Standard models of intertemporal choice assume that individuals discount future payoffs by integrating reward amounts and time delays to generate a discounted value. Alternative models propose that, rather than integrate across them, individuals compare within attributes (amounts and delays) to determine if differences in one attribute outweigh differences in another attribute. For instance, Leland (2002) and Rubinstein (2003) propose models that 1) compare the two reward amounts to determine whether they are similar, 2) compare the similarity of the two time delays, and then 3) make a decision based on these similarity judgments. Here, I tested discounting models against attribute-based models that use similarity judgments to make choices. I collected intertemporal choices and similarity judgments for the reward amounts and time delays from participants in three experiments. All experiments tested the ability of discounting and similarity models to predict intertemporal choices. Model generalization analyses showed that the best predicting models started with similarity judgments and then, if similarity failed to make a prediction, resorted to discounting models. Similarity judgments also matched intertemporal choice data demonstrating both the magnitude and sign effects, thereby accounting for behavioral data that contradict many discounting models. These results highlight the possibility that attribute-based models such as the similarity models provide alternatives to discounting that may offer insights into the process of making intertemporal choices.

Word count: 8531

Keywords: attribute-based decisions, intertemporal choice, process model, similarity, temporal discounting

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

Introduction

Which would you prefer, a piece of cake now or a slimmer waist next week? How about $100 today or $105 in one year? Intertemporal choices (Frederick, Loewenstein, & O'Donoghue, 2002; Read, 2004; Stevens, 2010) such as these underlie the most pressing decisions we have to make, from addressing global climate change (Stern, 2008) and the war on obesity (Komlos, Smith, & Bogin, 2004) to consuming alcohol (Rachlin, 2000) and investing in retirement plans (Laibson, Repetto, & Tobacman, 1998). In all of these cases, we must make decisions about future outcomes. Despite extensive interest in this topic, a critical gap remains in our knowledge of *how* we make intertemporal choices.

For the last 75 years, the standard models of intertemporal choice assume that we temporally discount (i.e., subjectively devalue) the future when given the choice between a smaller reward available sooner and a larger reward available later. An alternative approach, however, suggests other means by which we can make these decisions. Rather than integrate attributes to generate a discounted value for each option, these models compare attributes (reward amounts and time delays) to determine if differences in one attribute outweigh differences in another attribute (Leland, 2002; Rubinstein, 2003; Scholten & Read, 2010; Vlaev, Chater, Stewart, & Brown, 2011). Here, I explore whether attribute-wise decision making can provide a viable alternative or supplement to discounting.

Temporal Discounting

The temporal discounting approach typically offers an 'as-if' model of decision making (Berg & Gigerenzer, 2010; Kacelnik, 1997) rather than an explicit model of the process of decision making (but see Kable & Glimcher, 2007). Discounting models usually assume that individuals generate a subjective value for rewards discounted by the time delay to receiving the rewards and choose the option with the highest discounted value. For instance, in the previous monetary example, people often treat the $105 in one year as worth less than $105 today because they must wait for it.

This research was funded by the Max Planck Society and supported in part by an Alexander von Humboldt Foundation TransCoop Grant and by a National Science Foundation grant (NSF-1062045). I would like to thank Verena Köstner for translating the stimuli and data, Ralf Künster, Stefan Lindner, Sebastian Grueneisen for help with programming and data analysis, Gregor Caregnato and Emily Johnson for testing participants, and Edward Cokely, Gerd Gigerenzer, Kevin Gluck, Jon Leland, Henrik Olsson, Mike Regenwetter, Jörg Rieskamp, Marc Scholten, and Lael Schooler for valuable discussions regarding this work and/or comments on the manuscript. This research was approved by the Max Planck Institute for Human Development's Ethics Committee and the University of Nebraska-Lincoln Internal Review Board (IRB Approval # 20130313118EP).

Correspondence concerning this article should be addressed to Jeffrey R. Stevens, Department of Psychology, University of Nebraska-Lincoln, B83 East Stadium, Lincoln, NE 68588-0156, USA, E-mail: jeffrey.r.stevens@gmail.com

So, while the present value of the immediate option remains \$100, the present value of the delayed option decreases. Discounting models can make functional sense if a future benefit is uncertain. Typically, the farther in the future a benefit occurs, the lower the probability of it actually being realized. Therefore, future rewards should have a lower expected value. The form of these "hazard functions" of environmental uncertainty should map onto the discounted value functions (Kacelnik, 1997; Sozou, 1998; Stephens, 2002). Though dozens of discounting models exist (Doyle, 2013), I focus on a handful of the most commonly discussed models (Table 1).

Table 1
Intertemporal Choice Models

Models	Choose Larger, Later if...
Exponential	$A_l e^{-\delta t_l} > A_s e^{-\delta t_s}$
Hyperbolic (Mazur)	$\frac{A_l}{1+kt_l} > \frac{A_s}{1+kt_s}$
Hyperbolic (Rachlin)	$\frac{A_l}{1+kt_l^\sigma} > \frac{A_s}{1+kt_s^\sigma}$
Hyperbolic (Kirby)	$\frac{A_l}{1+kA_l^\mu t_l} > \frac{A_s}{1+kA_s^\mu t_s}$
Hyperbolic (Loewenstein & Prelec)	$\frac{A_l}{(1+\alpha t_l)^{\beta/\alpha}} > \frac{A_s}{(1+\alpha t_s)^{\beta/\alpha}}$
Arithmetic	$A_l - \lambda t_l > A_s - \lambda t_s$
Similarity	t_s and t_l are similar but A_s and A_l are dissimilar

Note. A represents reward amount; t represents time delay; δ, k, σ, μ, α, β, and λ represent model-specific parameters; and subscripts s and l refer to the smaller, sooner and larger, later option, respectively. If the inequality is reversed for the first five models, they predict choice for the smaller, sooner option. For similarity, if A_s and A_l are similar but t_s and t_l are dissimilar, it predicts choosing the smaller, sooner option. If neither of these is satisfied, it either chooses randomly (Leland, 2002) or uses some other criterion (Rubinstein, 2003).

The standard economic model of *exponential discounting* (Samuelson, 1937) assumes that discounted values should correspond to compound interest. Individuals should choose based on which option offers the best outcome should they borrow or lend money at the market interest rate (Read, 2004). Exponential discounting predicts that the present value of an option V decays at a constant rate: $V = Ae^{-\delta t}$, where A represents reward amount, t represents time delay to receiving the reward, and δ represents a discount parameter. The discount parameter δ is a function of the discount rate ρ ($\delta = -ln(1 - \rho)$), which describes how quickly value decreases over time. We would expect exponential discounting when the probability of losing a future reward is constant per unit time.

Though mathematically elegant and economically intuitive, much of the experimental evidence in humans and other animals contradicts predictions of exponential discounting (reviewed in Frederick et al., 2002). Psychologists developed the alternative notion of *hyperbolic discounting* (Ainslie, 1975; Chung & Herrnstein, 1967; Herrnstein, 1981; Rachlin, 1970), and Mazur (1987) formalized the current standard hyperbolic model: $V = \frac{A}{1+kt}$, where k is a discounting parameter that scales the steepness of discounting or the degree of preference for immediate rewards. Whereas exponential discounting corresponds to compound interest in economic terms, hyperbolic discounting corresponds to simple interest (Read, 2004). This model successfully fits people's discounting patterns, typically better than exponential models (Frederick et al., 2002; Rachlin, Raineri, & Cross, 1991) because it includes a discount rate that decreases with delay rather than remaining constant. Studies differ in how they compare models, but typically they fit various models using non-linear least-squares regression and compare R^2 values (Kirby & Maraković, 1995; McKerchar et al., 2009). Hyperbolic discounting consistently shows higher R^2 values, usually by 1-4 percentages points. Hyperbolic discounting also allows for time inconsistency, in which individuals plan to exhibit self-control when it is in the future, but as temptation nears, they often choose impulsively. A snooze bar on alarm clocks provide an example of this. In the evening, we set the alarm to wake up early to get a fresh start on the day. But once the alarm goes off, we often hit the snooze bar, succumbing to the temptation of more sleep. Hyperbolic discounting is also related to rate-based models of choice developed in the behaviorist tradition of psychology (Chung & Herrnstein, 1967; Herrnstein, 1981) and the foraging theory tradition of evolutionary biology (Kacelnik, 1997; Stephens & Krebs, 1986). If an individual maximizes his/her intake rate (rewards per unit time), this will result in a hyperbolic form (though not necessarily Mazur's specification). Mazur's hyperbolic discounting model was originally designed to describe pigeon data with repeated intertemporal choices, an ideal situation for maximizing rate. Because hyperbolic discounting can account for these phenomena, it has historically been the standard model of intertemporal choice in psychology.

The Mazur hyperbolic discounting model, however, tends to "overpredict subjective value at shorter delays, while underpredicting it at longer delays" (McKerchar et al., 2009). Researchers have modified the Mazur model by incorporating more parameters to better fit the data. Rachlin (2006) added an exponent σ to the time delay to better capture sensitivity to delay: $V = \frac{A}{1+kt^{\sigma}}$. This additional parameter improves fit by allowing a more flexible relationship between value and delay. Kirby (1997) included a parameterized amount in the denominator to capture how the discount rate is sensitive to the reward amount: $V = \frac{A}{1+kA^{\mu}t}$, where μ represents the sensitivity of discount rate to amount. Loewenstein and Prelec (1992) provide another modification of the hyperbolic discounting model that includes Mazur's hyperbolic model and the exponential model as special cases: $V = \frac{A}{(1+\alpha t)^{\beta/\alpha}}$.

Despite its success in quantitatively fitting functional forms of data, a number of qualitative empirical findings contradict Mazur's hyperbolic discounting model (reviewed in Frederick et al., 2002; Read, 2004). Here I focus on two such "anomalies": the magnitude effect and the sign effect. The magnitude effect occurs when participants' purported rate of discounting decreases as the absolute magnitude of the rewards increases (Green, Myerson, & McFadden, 1997; Thaler, 1981). Thus, people choose the smaller, sooner option more when facing \$1 today vs. \$5 in one year compared to when facing \$1,000 today vs. \$5,000 in one year, even though the ratio of rewards is the same. This constant reward ratio is important because hyperbolic discounting (along with exponential discounting) predicts that an individual preferring \$1 today over \$5 in year will always choose the smaller, sooner reward if the delays are fixed and the reward ratio is constant. The sign effect occurs when the discounting rate changes depending on whether the intertemporal choices involve positive outcomes (gains) or negative outcomes (losses). In particular, participants tend to discount gains more than losses (Estle, Green, Myerson, & Holt, 2006; Hardisty, Appelt, & Weber, 2013; Thaler, 1981), though some individuals reverse their preferences for losses, opting to advance rather than delay them (Yates & Watts, 1975). Hyperbolic discounting models with more parameters and nonlinear utility functions (e.g., Kirby, 1997; Loewenstein & Prelec, 1992) better fit the data and can allow for behavioral anomalies such as the magnitude and sign effects. Nevertheless, Mazur's hyperbolic discounting model continues to dominate the field of intertemporal choice.

The *arithmetic discounting* model[1] provides an alternative to hyperbolic discounting that converts the time delay into "disutility" and subtracts it from the reward amount (Doyle, 2013): $V = A - \lambda t$, where λ represents the discounting parameter. Doyle and Chen (2012) suggest that arithmetic discounting can outperform hyperbolic and exponential discounting.

Attribute-based Models

An alternative to discounting exists. The attribute-based approach (Payne, Bettman, & Johnson, 1993; Scholten & Read, 2010; Vlaev et al., 2011) takes a completely different view than the discounting approach. Instead of integrating the reward amount and time delay attributes to create a discounted value for each option, attribute-based models propose that individuals compare the attributes across options. Each of the models uses a different technique, but the general idea is to compare the values within an attribute (small amount compared to large amount and short delay compared to long delay) and then evaluate whether one attribute drives

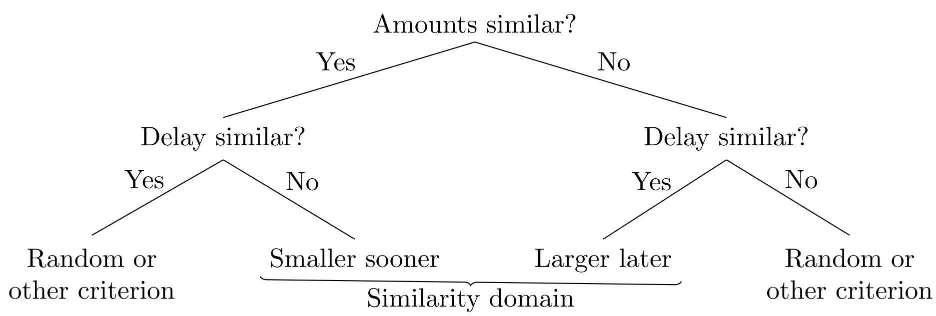
[1]Killeen (2009) has developed a more elaborate version of this model (called the additive discounting model) with nonlinear utility and time perception functions.

The tradeoff model (Scholten & Read, 2010) is an attribute-based model related to the arithmetic discounting model. In a simplified version of the model, the tradeoff between the attributes is given as $\kappa[w(t_l) - w(t_s)] = v(A_l) - v(A_s)$, where κ is a comparison parameter, w is a time-weighting function, and v is a value-weighting function. When w and v are concave (due to diminishing sensitivity), the model falls between the arithmetic discounting model and an attribute-based model. When w and v are linear, however, this model reduces to the arithmetic discounting model used here.

choice. For instance, these models would compare receiving $100 vs. $105 and waiting until today vs. one year and then assess whether the reward amount or time delay comparison (nor neither) determines choice.

Attribute-based models have been developed for two primary reasons. First, the discounting models fail to account for a number of key empirical findings in the literature (Leland, 2002; Rubinstein, 2003; Scholten & Read, 2010). Second, they do not offer accounts of the psychological process of decision making. When Rubinstein (2003) proposed an attribute-based model for intertemporal choice, he suggested that the existing discounting accounts of choice did not match the intuition one has about the psychological process experienced in making these decisions. The advantage of the attribute-based models is that they offer a window into the process of decision making by making predictions about the order of obtaining and using information about the attributes. Further, Rubinstein asserts that "the decision maker uses a procedure that aims at simplifying the choice by applying similarity relations" (p. 1210). Thus, attribute-based accounts may offer cognitively simpler processes for making intertemporal choices by avoiding integrating across attributes and focusing on potentially simpler comparison within them.

Leland (2002) and Rubinstein (2003) developed an alternative approach that examined the influence of similarity judgments on intertemporal choices. Here, similarity refers to the psychological distance between receiving the two reward amounts or between waiting the two time delays. The *similarity* models use the perceived similarity of the reward amounts and of the time delays to make a decision. The similarity model can be described by a decision tree:



If only one attribute is judged as similar, then ignore that attribute and decide based on the other. In the previous example, one might judge receiving $100 and $105 to be quite similar, whereas waiting 0 days vs. 1 year as not similar. Using the similarity model, one would ignore the amount attribute since they are similar and choose based on the time delay, therefore opting for the sooner reward of $100 today. This can generate similar behavior to the discounting models but via very different decision processes.

In situations in which either amounts or delays are judged as similar (inner two terminal branches of decision tree), I label this the *similarity domain* because the model makes a deterministic prediction in these circumstances. Two versions of this model exist that differ in their behavior outside of the similarity domain, that is, when both attributes are either similar or dissimilar (outer two terminal branches of decision tree). In the Leland (2002) version, the model predicts choosing randomly when outside the similarity domain. The Rubinstein (2003) version asserts that another criterion must be used when outside of the similarity domain. Rubinstein, however, did not specify any other possible criteria, so this form of the model makes no predictions in these circumstances. Here, I add the discounting models as the second criterion for cases outside of the similarity domain. Thus, I present seven similarity models: Leland's version with random choice outside of the similarity domain and six separate versions with the other models implemented outside of the similarity domain.

Present Study

The aim of the present study was to formally test discounting and similarity models of intertemporal choice. Thus far, the only data collected on the similarity model are Rubinstein's (2003) critical tests. These critical tests, however, did not directly measure similarity judgments.

This study offers competitive model selection tests of the similarity model using similarity judgments from participants. To test these models, I collected choice data for intertemporal choices. Unlike previous intertemporal choice studies, I provide generalization tests of predictive accuracy to offer a more robust test of models. Generalization tests fit one set of data and predict responses on a different set of data (Busemeyer & Wang, 2000; Marewski & Olsson, 2009). In addition, these experiments test whether the similarity model can account for two key anomalies associated with hyperbolic discounting: the magnitude effect and the sign effect. The combination of model generalization tests and anomaly tests provide converging methods to explore attribute-based models of intertemporal choice.

Experiment 1: Testing Similarity and the Magnitude Effect

The goals of the first experiment were to (1) compare the predictive accuracy of discounting models (exponential, hyperbolic, and arithmetic) to similarity-based models and (2) explore whether similarity-based models can account for the magnitude effect in intertemporal choice. To robustly compare the models, I first fit them to one set of data and then used generalization techniques to test the predictive accuracy of the models on a different set of data. To test the influence of similarity judgments on choice, I collected dichotomous similarity ratings from participants for pairs of reward amounts and pairs of time delays.

I also test the magnitude effect—the fact that the discount rate changes with the magnitude of the reward (Green et al., 1997; Thaler, 1981). The magnitude effect is not predicted by exponential discounting or Mazur's hyperbolic discounting. Here, I tested the magnitude effect by offering participants a series of questions in which the short delay and long delay remained constant, but the small and large amounts varied (ranging from \$2-18), though their ratio remained constant. The hyperbolic model predicts the same choice across these questions because the amount ratio is constant. The similarity model, in contrast, predicts different choices if similarity changes with the magnitude of the reward amounts.

Methods

Participants. In May 2009, I tested 64 participants (29 males and 35 females) with a mean±SD age of 25.8±3.0 (range 19-33) years, recruited from German universities via the Max Planck Institute for Human Development participant pool. They received €8 for participating in the experiment and earned an additional €7.30±2.44 (range €1-15), based on their choices in the experiment.

Materials and procedure.

Procedural overview. All materials were prepared in German. The experiment included three phases. The first two phases (binary choice phase and staircase phase) offered participants intertemporal choice questions between pairs of options. In the final phase (similarity judgment phase), participants rated the similarity of the reward amounts and time delays used in the previous intertemporal choice phases. Questions were presented using HTML forms with response buttons and are available in the Supplementary Materials.

Before beginning the first phase, the computer program explained to participants that their choices determined their payoffs: The computer program would randomly select one of the intertemporal choice questions, and the participant would receive the option that they chose via bank transfer. Thus, the participants were incentivized to make choices reflecting their true preferences because they would actually receive the amount they chose after the appropriate time delay. At the end of the experiment, participants were shown the randomly selected intertemporal choice question and their choice for that question. They were given the option of accepting this outcome, or, if the outcome was delayed, they could opt for 85% of the amount in cash immediately. Participants did not know that they would receive this option while making the prior intertemporal choices or similarity judgments.

Binary choice phase. The first phase consisted of a series of 87 questions offering binary choices between options with different reward amounts and time delays, ranging from €1-20 and 0-85 days (Table S1). All participants first experienced the same two practice questions before moving to the test questions, the order of which was randomized across participants.

A subset of questions was designed to test the magnitude effect (Table S2). These questions had fixed short delays and long delays and a fixed ratio but different magnitudes of small amounts and large amounts. With these questions, a hyperbolic discounter would make the same choice across questions, assuming a consistent discount parameter k. I offered three blocks (with amount ratios of 0.50, 0.67, and 0.80) of six questions each. Within each block, three questions involved an immediate short time delay, and three questions involved a delayed short time delay. For these questions, the ratio of amounts, ratio of delays, and difference between delays remained constant, with only the difference between amounts (and therefore amount magnitude) varying across questions.

Staircase phase. In the second phase, blocks of intertemporal choice questions were presented using the staircase method. Staircase questions were presented in 20 blocks (plus 1 practice block) of 10 questions. For 17 of these blocks, the small amount varied incrementally from €1-10, while the large amount, short delay, and long delay remained constant. For example, we asked participants, "Which option would you prefer? €1 in 1 day or €10 in 6 days", then "Which option would you prefer? €2 in 1 day or €10 in 6 days". This continued until they reached "Which option would you prefer? €10 in 1 day or €10 in 6 days". For 3 of the blocks the short delay varied incrementally from 9 to 0 days, while the small amount, large amount, and long delays remained constant. Order of presentation (ascending or descending amounts or times) influences discounting parameter estimates (Hardisty et al., 2013; Robles & Vargas, 2007) suggesting that adjusting amounts and adjusting delays could yield different parameter estimates, as well. Therefore, to reduce potential variance in the parameter estimation, the adjusting-delay data were not analyzed here; I only included the adjusting-amount data. Participants began this phase of the experiment with one block of 10 practice questions. The order of trials within a block always increased from €1-10, but the order of blocks was randomized across participants. Mean choice percentages are presented in Figure S1.

Similarity judgment phase. In the final phase, participants made 60 dichotomous similar/different distinctions between reward amounts (23 questions) and between time delays (37 questions): "Indicate whether you would rate the above amounts [delays] as similar or different". All amount and delay pairs were drawn from but did not include all binary choice and staircase questions from the first two phases.

Data analysis. I processed and analyzed the data using R statistical software[2] version 3.3.1 (R Development Core Team, 2014). Data and R code[3] are available in the Supplementary Materials and will be posted on the IQSS Dataverse Network data repository (http://thedata.harvard.edu/dvn/).

I used individual participants as the unit of analysis, so all measures of choice and similarity are calculated over the mean values of each participant. When comparing measures within a participant, I used within-subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008) to remove between-participant effects.

[2]In addition to the core R program, I used the bbmle (Bolker & R Development Core Team, 2012), car (Fox & Weisberg, 2011), epicalc (Chongsuvivatwong, 2012), foreach (Revolution Analytics & Weston, 2014), Hmisc (Harrell, with contributions from Charles Dupont, & many others, 2014), lattice (Sarkar, 2008), latticeExtra (Sarkar & Andrews, 2013), plyr (Wickham, 2011), xtable (Dahl, 2013), and zoo (Zeileis & Grothendieck, 2005) packages.

[3]The original LaTeX document, with Sweave-embedded R code (Leisch, 2002) to allow reproduction of analyses (de Leeuw, 2001), is available from the author.

Model selection. I first fit the exponential discounting, hyperbolic discounting, and arithmetic discounting models to each participant's staircase data using maximum likelihood estimation with an inverse logit function and a binomial distribution (median parameter estimates available in Table S3). I removed from the analysis participants whose maximum likelihood estimates failed to converge (typically due to nearly exclusive choice of the larger, later option), yielding data from 51 participants. To report fit for these models, I include AICc values (Burnham & Anderson, 2010) computed both over all data and separately for each participant. The similarity models had no parameters to fit for this analysis.

Next, I used the fitted parameters from each model to predict responses for binary choice questions. I generated a prediction for each binary choice question, using participant-specific parameters estimated from the staircase data. For each participant and each model, I calculated predictive accuracy as the percentage of questions for which the model correctly predicted the participant's choice.

I used the dichotomous similarity ratings as the input into the similarity model. The 60 similarity judgments did not cover all attribute pairs, allowing the similarity models to make predictions for 46 of the 87 questions (53%). I restricted the model selection analysis to this subset of questions to allow a similar comparison across all models. I tested seven forms of the similarity model. Leland's (2002) version of the model chose randomly when both attributes were judged as similar or dissimilar (outside of the similarity domain). Predictive accuracy for a participant was calculated as the mean predictive accuracy of the deterministic predictions and of the random predictions, weighted by the number of questions in each of those categories[4]. The remaining six similarity models employed the discounting models when outside the similarity domain. Thus, they were two-stage models with a similarity judgment stage and, if similarity did not make a deterministic prediction, a second stage used another model. The mean percentage of questions in the similarity domain for participants was 62% (median: 64%), ranging from 4-100%.

Results and Discussion

Model selection. Table 2 shows the mean AICc values (lower is better fit) and predictive accuracy (higher is better performance) for all models tested

[4]Predictive accuracy was measured by assessing whether data matched the deterministic predictions of the models. For random predictions, the expected predicted choice was 50% since individuals were randomly choosing between two options. Therefore, for each participant, I calculated the percent choice for the larger, later option in the questions for which the similarity model predicted random choice (separately for both similar and both dissimilar). I then measured the absolute deviation of the observed choice percentage from the expected percentage (50) and divided by the expected percentage:

$$\text{predictive accuracy} = 1 - \frac{|\text{observed} - 50|}{50}.$$

in Experiment 1. Rachlin's two-parameter hyperbolic discounting model best fit the aggregated data, and arithmetic discounting best fit the individual data. Yet, when predicting new data, all discounting models performed about equally well, predicting 70.5-74.2%. The two-stage similarity models, however, outpredicted the discounting models with a predictive accuracy of 77.0-79.2%. As an exploratory analysis, I compared the single-parameter hyperbolic model (Mazur) to the matching two-stage similarity model (similarity+Mazur). I chose Mazur's model because it performed as well as all other models, offers parsimony with a single parameter, and is the standard model used in intertemporal choice. The two-stage model significantly outperformed the discounting only model by 7.0±3.1 percentage points, a medium-sized effect (Cohen's d = 0.63). With the exception of Leland's model, all of the two-stage similarity models performed at fairly comparable levels and better than the discounting models. Figure 1 shows boxplots of individual participant predictive accuracy to illustrate the variation in accuracy across models.

Table 2
Model Selection Results for Experiment 1

Model	Aggregate AICc	Individual AICc	Predictive Accuracy
Exponential	6265.6	62.1	71.7±2.3
Hyperbolic (Mazur)	6096.1	62.5	72.2±2.4
Hyperbolic (Rachlin)	**5989.6**	60.3	70.5±2.8
Hyperbolic (Kirby)	6078.7	62.4	73.1±2.1
Hyperbolic (Loewenstein & Prelec)	5992.9	61.3	72.1±2.8
Arithmetic	6451.4	**60.2**	74.2±1.8
Similarity (Leland)	NA	NA	69.4±7.3
Similarity+exponential	NA	NA	79.0±1.5
Similarity+Mazur	NA	NA	**79.2±1.6**
Similarity+Rachlin	NA	NA	77.7±1.7
Similarity+Kirby	NA	NA	78.2±1.6
Similarity+L&P	NA	NA	78.7±1.5
Similarity+arithmetic	NA	NA	77.0±1.7

Note. Aggregate AICc values are calculated using all staircase data. Individual AICc values are the median AICc values calculated separately for each participant. Predictive Accuracy is the mean percentage (± within-subjects 95% confidence intervals) of correctly predicted binary choice data calculated over all participant means. Best fitted or predicted models for each measure are in boldface. NA refers to the fact that the similarity models are not fitted to staircase data. Data are based on 51 participants.

Leland's (2002) similarity model had the lowest mean predictive accuracy of all models at 69.4%, though this was comparable to the discounting models. As illustrated in Figure 1, Leland's similarity model included a large number of participants for whom it had very low predictive accuracy. Many participants were clearly not choosing randomly outside of the similarity domain, and the model was severely penalized by them in terms of overall predictive accuracy. This similarity+random choice model, however, performed as well as the discounting models.

When restricting the model selection analysis only to questions within the similarity domain, the models resulted in the following predictive accuracies: exponential discounting 64.2%, Mazur hyperbolic discounting 64.9%, Rachlin hyperbolic discounting 63.3%, Kirby hyperbolic discounting 69.1%, Loewenstein and Prelec hyperbolic discounting 65.0%, arithmetic discounting 76.8%, similarity 85.7%. Thus, when it could make a deterministic prediction, the similarity model outperformed all other models.

Magnitude effect. To test the magnitude effect, I varied the amount magnitude, while holding the amount ratio, short delay, and long delay within a block constant for both similarity judgments and choice. To test whether the similarity model predicts different choices within a block, I examined how the similarity ratings of reward amounts varied at different reward magnitudes. Increasing amount magnitudes reduced similarity judgments (Figure 2a), predicting an increase in choosing the larger, later options in intertemporal choice. As predicted by the similarity judgments, actual choices for the larger, later option increased as the amount magnitude increased (Figure 2b). Mazur's hyperbolic discounting predicts similar choices (i.e., a flat line) across these magnitudes. Therefore, these findings contradict Mazur's hyperbolic discounting but are consistent with predictions of the similarity model, suggesting that similarity could underly the magnitude effect observed here.

Experiment 2: Testing Similarity without the Magnitude Effect

The goal of the second experiment was to test whether the superior predictive accuracy observed in the similarity model in Experiment 1 was only due to its ability to account for the magnitude effect. To test this, I controlled for the magnitude effect by holding both the amounts and the k parameter values at indifference constant. I then varied only the delay magnitudes to determine whether similarity judgments tracked delays and continued to outperform the discounting models.

Methods

Participants. In December 2014, I tested 62 participants (23 males and 39 females) with a mean±SD age of 20.1±3.5 (range 18-45) years, recruited from the University of Nebraska-Lincoln Department of Psychology undergraduate participant pool. Participants received one course credit rather than money for their participation.

Materials and procedure. This experiment was conducted using the web-based Qualtrics Survey Software and included five phases. The first phase presented a set of 31 binary choice questions (plus two practice questions). I restricted the analysis here to questions with a small amount of \$7, which resulted in 25 questions (Figure S5). Results were the same when including the six questions with small amount of \$8. All questions had small amounts of \$7 and large amounts of \$10. Questions had k parameters at indifference of 0.333, 0.5, 0.6, 0.75, and 1.0. However, I varied the magnitude of the time delays from 18-309 days. I chose delays that would act as critical tests that result in different predictions for the hyperbolic and similarity models. In particular, given the k parameters, most participants should choose the larger, later option for all questions if they are hyperbolically discounting. However, the similarity model predicts choosing the smaller, sooner option in most of these questions because the amounts would likely be rated as similar but the delays rated as dissimilar.

The second phase included a set of staircase choice questions consisting of eight blocks of 10 questions in which the small amount varied from \$1-10, while the large amount (\$10), short delay (0 days), and long delay remained constant within a block. Across blocks, the long delay varied between 2, 7, 14, 30, 60, 90, 180, and 365 days, with the order of presentation randomized across participants. Mean choice percentage for binary choice data are presented in Table S5 and staircase data are presented in Figure S2.

The next two phases measured similarity judgments. Participants judged the similarity of receiving monetary rewards (e.g., "Would you rate receiving \$1 or \$10 as similar or different?") and then the similarity of waiting (e.g., "Would you rate waiting 0 days or 2 days as similar or different?"). The amount and time delay values used in the similarity judgments included all values used in the intertemporal choices. The final phase collected demographic information, including age, gender, university major, ethnicity, employment status, number of children, and parental income.

Data analysis. Data are available as supplementary materials. As in Experiment 1, for the model selection analysis, I removed participants whose maximum likelihood estimates did not converge. This yielded data from 54 participants. From these participants, I calculated predictive accuracy for all models.

Results and Discussion

Participants chose the larger, later option less as the overall delay magnitude increased, even when amount magnitude and k values were held constant (Figure S3). This is not predicted by Mazur's hyperbolic discounting model. As demonstrated with the amount magnitude effect in Experiment 1, the similarity judgments for these same delay pairs matched the choice proportions in the intertemporal choice questions, again suggesting that choices mirror similarity judgments (Figure S3).

To test whether similarity judgments are not only consistent with choice but consistent with the use of the similarity model, I calculated predictive accuracy for all models using this data. Table 3 and Figure 3 show that similarity models greatly outpredict discounting models alone. Similarity+Mazur discounting outpredicts Mazur discounting alone by 23.0±11.0 percentage points, a medium-sized effect

(Cohen's $d = 0.57$). Therefore, similarity outpredicts discounting alone because it accounts for magnitude effects in both amounts and delays.

Table 3
Model Selection Results for Experiment 2

Model	Aggregate AICc	Individual AICc	Predictive Accuracy
Exponential	4886.8	32.5	52.9±7.8
Hyperbolic (Mazur)	4323.8	27.4	42.4±7.4
Hyperbolic (Rachlin)	3894.6	26.1	43.9±6.6
Hyperbolic (Kirby)	4323.8	30.8	42.1±7.2
Hyperbolic (Loewenstein & Prelec)	**3890.2**	**24.7**	42.4±7.1
Arithmetic	6361.9	45.1	48.3±6.2
Similarity (Leland)	NA	NA	55.6±14.1
Similarity+exponential	NA	NA	**68.4±7.1**
Similarity+Mazur	NA	NA	65.3±5.7
Similarity+Rachlin	NA	NA	64.9±5.6
Similarity+Kirby	NA	NA	64.5±5.7
Similarity+L&P	NA	NA	64.1±5.7
Similarity+arithmetic	NA	NA	67.2±5.2

Note. Aggregate AICc values are calculated using all staircase data. Individual AICc values are the median AICc values calculated separately for each participant. Predictive Accuracy is the mean percentage (± within-subjects 95% confidence intervals) of correctly predicted binary choice data calculated over all participant means. Best fitted or predicted models for each measure are in boldface. NA refers to the fact that the similarity models are not fitted to staircase data. Data are based on 54 participants.

Experiment 3: Testing Similarity and the Sign Effect

The goals of the third experiment were to (1) replicate key model selection results from Experiment 1 and (2) explore whether similarity-based models can account for the sign effect in intertemporal choice. This experiment allowed confirmatory tests of the exploratory analyses comparing Mazur's hyperbolic model and the two-stage similarity model with Mazur's hyperbolic discounting. This tested whether adding similarity as the first step robustly improves the predictive accuracy of the Mazur hyperbolic model. As in Experiment 1, I first fit the hyperbolic models to one set of data and then tested the predictive accuracy of the models on a different set of data.

To test the sign effect, I offered participants a series of intertemporal choices in which they would *receive* money after a delay (gain condition) or *pay* money after a delay (loss condition). I then asked them to judge the similarity of receiving monetary amounts, paying monetary amounts, and waiting for time delays. This allowed me to map similarity judgments for gains and losses on to the intertemporal choices for gains and losses, thereby testing whether the similarity model can account for the sign effect.

Methods

Participants. From September to October 2013, I tested 68 participants (14 males and 54 females) with a mean±SD age of 19.8±2.8 (range 17-39) years, recruited from the University of Nebraska-Lincoln Department of Psychology undergraduate participant pool. Participants received one course credit rather than money for their participation.

Materials and procedure. This experiment was conducted using Qualtrics Survey Software and included eight phases. The first two phases presented a set of 40 binary choice questions from Luhmann (2013) (plus two practice questions). The second phase included a set of staircase choice questions consisting of six blocks (plus 1 practice block) of 10 questions in which the small amount varied from $1-10, while the large amount ($10), short delay (0 days), and long delay remained constant within a block. Across blocks, the long delay varied between 2, 7, 14, 30, 60, and 90 days, with the order of presentation randomized across participants. For both phases, the questions were phrased as hypothetical gains (e.g., Would you prefer to RECEIVE $47 in 30 days or $58 in 80 days?). The third and fourth phases consisted of sets of the same binary and staircase questions in which the amounts were hypothetical losses (e.g., Would you prefer to PAY $47 in 30 days or $58 in 80 days?). Mean choice percentage for binary choice data are presented in Table S4 and staircase data are presented in Figures S4 and S5.

The next three phases measured similarity judgments. Participants judged the similarity of receiving monetary gains (e.g., "Would you rate RECEIVING $1 or $10 as similar or different?"), the similarity of paying monetary losses (e.g., "Would you rate PAYING $1 or $10 as similar or different?"), and then the similarity of waiting (e.g., "Would you rate WAITING 0 days or 2 days as similar or different?"). The amount and time delay values used in the similarity judgments included all values used in the intertemporal choices. The final phase collected demographic information, including age, gender, university major, ethnicity, employment status, number of children, and parental income.

Data analysis. Data are available as supplementary materials. For the model selection analysis, I removed participants whose maximum likelihood estimates did not converge. This yielded data from 57 participants for the gain condition and 28 participants for the loss condition.

Thirty-nine of the forty participants that were dropped in the loss condition almost always chose the smaller, sooner option, and one participant almost always chose the larger, later option (Figure S5). This likely occurred because some participants prefer losses to be advanced while other prefer them to be delayed (Yates & Watts, 1975). I tested this by measuring choice in the staircase questions in which both options had the same amount ($10) but at different delays. Each participant experienced six of these questions (one for each staircase block), and I categorized each participant as preferring losses (1) advanced if they chose the sooner option four or more times, (2) delayed if they chose the later option four or more times, and (3) neutral if they chose both options equally often. Whereas in the gain condition, 66 of 68 participants advanced gains (with the other two being neutral), in the loss condition, 35 advanced losses and 31 delayed losses, roughly matching the even split shown by Yates and Watts (1975). Moreover, in the loss condition, 28 of the 40 dropped participants (70%) were categorized as preferring advanced losses compared to 7 of the 28 retained participants (25%). Advancing losses implies a negative discount rate. Therefore, the drop in participants in the loss condition seems to result from a high number of participants with negative discount rates, which the stimuli were not designed to detect.

Results and Discussion

Replication. For the model selection replication, I used only the gain condition data to provide the clearest comparision to Experiment 1. As in Experiment 1, the two-stage similarity models yielded higher predictive accuracy than the discounting models alone (Table 3). Mazur's hyperbolic model correctly predicted 65.6±2.0% of the gain binary choice data, and the two-stage similarity+Mazur model correctly predicted 68.7±2.1% of the data. Therefore, confirmatory analysis indicates that adding the similarity assessment before discounting significantly improved predictive accuracy by 3.1±1.8 percentage points, a small effect size (Cohen's d = 0.45). This benefit likely results from the high predictive accuracy of 86.8% for the similarity model in the similarity domain. This result replicates the findings of Experiment 1 despite testing in different countries (Germany vs. U.S.), different payment schemes (performance-based pay vs. hypothetical rewards), and different sex ratios (even vs. skewed toward females). Thus, the similarity model provides robust predictive accuracy over discounting models alone.

Sign effect. To investigate whether the similarity model can account for the sign effect, I conducted the previously described model selection analysis on the loss data. Table 3 shows that all models, except Leland's similarity model performed at comparable levels. Notably, the similarity models provided the same predictive accuracy as the discounting models. Mazur's hyperbolic model correctly predicted 68.3±4.2% of the loss binary choices, and the two-stage similarity model with Mazur hyperbolic discounting predicted a comparable 67.5±4.5%. Therefore, though similarity models do not outperform discounting models in the loss domain, they perform equally well, thereby accounting for the sign effect as well as discounting models.

To more thoroughly explore the sign effect, I calculated discount rates for both the gain and loss data. Because the previously described analyses on gain and loss data are based on different sets of participants (57 participants for the gain condition and 28 participants for the loss condition), I restricted this analysis to only

Table 4
Model Selection Results for Experiment 3

Model	Gain			Loss		
	Aggregate AICc	Individual AICc	Predictive Accuracy	Aggregate AICc	Individual AICc	Predictive Accuracy
Exponential	2879.0	22.1	66.6±2.8	2881.8	27.5	68.4±4.0
Hyperbolic (Mazur)	2701.7	21.2	65.6±2.0	2798.5	25.3	68.3±4.2
Hyperbolic (Rachlin)	2546.8	**18.1**	61.4±2.6	**1903.0**	19.7	68.1±3.7
Hyperbolic (Kirby)	2701.7	22.3	55.5±2.8	2798.5	25.9	67.1±3.5
Hyperbolic (L&P)	**2533.0**	23.7	52.4±3.6	2420.7	**19.1**	65.6±3.6
Arithmetic	3120.4	21.4	43.9±3.7	2965.9	30.5	64.4±7.1
Similarity (Leland)	NA	NA	64.0±7.4	NA	NA	42.1±17.7
Similarity+exponential	NA	NA	**70.2±2.4**	NA	NA	68.3±3.5
Similarity+Mazur	NA	NA	68.7±2.1	NA	NA	67.5±4.5
Similarity+Rachlin	NA	NA	65.9±2.1	NA	NA	**68.6±2.6**
Similarity+Kirby	NA	NA	61.7±2.2	NA	NA	67.9±2.9
Similarity+L&P	NA	NA	61.4±2.4	NA	NA	66.5±2.8
Similarity+arithmetic	NA	NA	55.4±3.7	NA	NA	66.5±5.7

Note. Aggregate AICc values are calculated using all staircase data. Individual AICc values are the median AICc values calculated separately for each participant. Predictive Accuracy is the mean percentage (± within-subjects 95% confidence intervals) of correctly predicted binary choice data calculated over all participant means. Best fitted or predicted models for each measure are in boldface. NA refers to the fact that the similarity models are not fitted to staircase data. Data are based on 57 participants for the gain condition and 28 participants for the loss condition.

participants for whom I could calculate maximum likelihood estimates for both gain and loss data (i.e., the 28 participants from the loss condition). The discount rate for gains (ρ =0.016±0.001) significantly differs from that for losses (ρ =0.009±0.001), with steeper discounting for gains. This finding replicates previous work in the field demonstrating steeper discounting for gains compared to losses (Estle et al., 2006; Hardisty et al., 2013; Thaler, 1981). I also calculated the similarity ratings of the reward amounts for gains and losses in both binary and staircase intertemporal choice data. Participants judged the amounts as similar in 30% of gain amount pairs and 28% of loss amount pairs, a significant difference of 2.2±2.0% with a small effect size (Cohen's d = 0.26). Since amounts are judged as more similar for gains than losses, this suggests that participants will ignore amounts and focus on delays more for gains than losses. This emphasis on delays will favor choosing the smaller, sooner option more, which results in higher discount rates for gains. Thus, differences in similarity judgments match those observed in intertemporal choices, though replications with larger samples are needed to confirm reliability.

A key limitation of interpreting the sign effect data is the fact that so many participants were dropped due to what appears to be negative discount rates for losses. Therefore, the analysis provided here applies to only a subset of decision makers, most of which have positive discount rates. Though Yates and Watts (1975) showed clear individual differences in positive or negative discount rates for losses, little research has expanded on or even recognized the possibility of negative discount rates when fitting models to loss data. Future work must acknowledge this variation to fully capture intertemporal choice data.

General Discussion

In Experiment 1, the discounting models all predicted new data with roughly equal success. Yet, the two-stage similarity-based models provided the highest mean predictive accuracy rates, with comparable levels of performance across the different discounting models. Moreover, similarity judgments tracked differences in amount magnitude, consistent with the magnitude effect observed in intertemporal choices. In Experiment 2, similarity judgments tracked choices and the similarity model outpredicted discounting models even when the magnitude effect was removed. While holding reward amounts constant (thereby removing the magnitude effect), varying the delay magnitudes influenced choices consistent with predictions from the similarity model. In Experiment 3, a replication of Experiment 1 again showed that adding similarity improved predictive accuracy, as the two-stage similarity-hyperbolic (Mazur) model outpredicted the hyperbolic (Mazur) model alone for the gain data. The similarity model also accounted for the sign effect both by predicting choices framed as losses as well as the hyperbolic discounting model and by demonstrating that similarity judgments tracked the gain/loss difference observed in discount rates. Thus, model generalization tests and tests of anomalies provide converging evidence supporting attribute-based models of intertemporal choice, such as the similarity model, as viable alternatives or precursors to discounting models.

Leland (2002) provided a similarity-based model of intertemporal choice that randomly chooses when similarity does not discriminate between attributes. This model is probably not an accurate model of choice given the random component of choice. In fact, this model cannot account for preference reversals[5] observed in participant data (Green, Fristoe, & Myerson, 1994; Kirby & Herrnstein, 1995). Yet, this simple model performed as well as discounting models for gain data. Viewing the distribution of participant accuracies suggests that this model yielded the largest range in predictive accuracies (Figure 1).

In Rubinstein's (2003) version of the similarity model, individuals are expected to use similarity to make a choice, and, if similarity does not distinguish, then use another criterion. Two-stage models of similarity were, in fact, quite successful in predicting participant choices. Models that start out using similarity models and

[5] For example, the large amount is typically chosen over the small amount when both delays are large. Preference reversals occur when choice switches from larger, later to the smaller, sooner option as the delay decreases (holding amounts constant). Leland's model would predict that choice should switch from larger, later (because delays are similar) to random as delays decrease (because they become more dissimilar).

then use discounting models if similarity does not make a deterministic prediction outperformed all other models for gain data. This raises the intriguing possibility that people start out with an attribute-based strategy for intertemporal choice and then may switch to discounting or other strategies as a last resort.

Though discounting models performed well as the second stage outside of the similarity domain, this does not imply that only discounting models are needed. In point of fact, if analysis is restricted to only questions found within the similarity domain for gains, the similarity model outperformed the next best models by 9-40 percentage points. Therefore, when the similarity model can make a deterministic prediction, it predicts choice at a much greater level than any of the discounting models. This indicates that similarity adds a unique contribution to intertemporal choice beyond discounting for gains.

For losses, similarity performed as well as but not better than discounting models. This may result from assessing delay similarity with a single set of judgments that did not discriminate between gains and losses. Including the gain and loss dimension for delay similarity judgments may further improve the accuracy of the similarity model in the loss domain.

Most studies of intertemporal choice typically rely on nonlinear regression of choice data to discriminate between models (e.g., Green, Myerson, & Macaux, 2005; McKerchar et al., 2009). In these analyses, hyperbolic discounting usually does a good job of *fitting* data, as it did in these two experiments. To improve fit, modelers often add more parameters to the hyperbolic model (Loewenstein & Prelec, 1992; Myerson & Green, 1995; Rachlin, 2006). Simply fitting models is problematic, however, because of the possibility of overfitting data (Pitt & Myung, 2002). Having more parameters allows a model to fit the noise in the data at the expense of capturing the overall relationship. One way to properly test the models and avoid overfitting is to *predict* new data (Marewski & Olsson, 2009). Though a common practice in machine learning and some areas of psychology, few if any studies of intertemporal choice use either cross validation (fitting a proportion of a single data set and predicting the rest; reviewed in Shiffrin, Lee, Kim, & Wagenmakers, 2009) or generalization (fitting one data set and predicting a different set; Busemeyer & Wang, 2000). This study used a generalization technique in intertemporal choice by fitting model parameters on the staircase data and measuring predictive accuracy on a different set of binary choice data.

In both experiments, adding more parameters to the Mazur hyperbolic model (e.g., using the Rachlin, Kirby, and Loewenstein & Prelec models) typically improved fit of the gain data. In predictive accuracy, however, at best the multi-parameter hyperbolic models performed only as well as the single-parameter hyperbolic model (Tables 2 & 3), and, in some cases, the single-parameter model predicted better. In addition, when combined with the similarity models, the two-parameter discounting models did not increase predictive accuracy over the one-parameter version. These two findings supports the notion that high-parameter models can overfit the data, especially when they are not constructed to accommodate psychological processes. Therefore, the current practice of comparing intertemporal choice models based on model fitting does not translate well to predicting new data.

Limitations and Future Directions

One limitation of interpreting the results of these studies is that the predictive accuracies of many of the models was fairly similar (Tables 2 & 3). In Experiment 1, the discounting models performed quite similarly. For gains, similarity models yield accuracies 3-11 percentage points higher than discounting models alone, matching the differences typically used to distinguish between fits of exponential and hyperbolic models (Kirby & Maraković, 1995; McKerchar et al., 2009). Thus, including similarity increases predictive accuracy. However, within these two tiers of models (discounting alone and similarity+discounting), the models perform similarly. We need to design future experimental stimuli specifically for discriminating among these models to better understand the relative success of discounting and similarity models. Scholten, Read, and Sanborn (2014) designed their studies to discriminate among several discounting models and their tradeoff model, with time-weighting and value functions included for both model types. The attribute-based tradeoff model outperformed the Loewenstein and Prelec (1992) hyperbolic model. Further, Dai and Busemeyer (2014) demonstrated that an attribute-based diffusion model can outpredict discounting models when using probabilistic and dynamic specifications. Thus, we have evidence from multiple studies that attribute-based models can better account for intertemporal choices than discounting models. An obvious next step is to begin testing attribute-based models against each other.

A limitation of the similarity model is that it lacks an explanation of the similarity judgment itself. It effectively pushes the explanatory question from the intertemporal choice to the similarity judgment. Thus, further refinements of the similarity model are needed to explore how individuals make similarity judgments for reward amounts and time delays. Rubinstein (1988), for instance, proposed that the ratio between rewards could drive similarity judgments. Though a nice start, this does not completely capture the nature of similarity judgments, because both ratios and differences influence similarity judgments for amounts and delays. Similarity judgments in models of choice clearly require more in-depth investigation.

Both cognitive psychology and machine learning have a long history of exploring similarity concepts (Aha, Kibler, & Albert, 1991; Goldstone & Son, 2005; Hahn & Chater, 1998; Shepard, 1987; Tversky, 1977). At the moment, there does not appear to be much work on similarity in monetary rewards or time delays, though researchers have investigated the role of time estimation on intertemporal choice (Wittmann & Paulus, 2008; Zauberman, Kim, Malkoc, & Bettman, 2009).

One key finding in the similarity literature is that context matters greatly. We would not expect people to rate \$1 vs. \$3 in the same way as they rate 1 cent vs. 3 cents or 1 day vs. 3 days or 1 year vs. 3 years. In fact, each of these four pairs could very well elicit different similarity ratings, despite sharing 1 vs. 3 in common. Moreover, even within identical magnitudes and currencies, data presented here show that gaining rewards vs. losing rewards are different contexts that influence similarity judgments. States such as an individual's socio-economic status also likely shape similarity judgments: an undergraduate will judge the similarity of $100 and $200 differently than a billionaire. Thus, contextual factors play a key role in similarity judgments, highlighting important open areas of research.

As a further example of context effects, the pairing of the amounts and delays together in an intertemporal choice question may influence their similarity judgments. For example, $1 vs. $3 may be rated as more similar when paired with long delays than when paired with short delays, a phenomenon termed inseparability (Scholten & Read, 2010). This interdependency suggests that the current estimates of accuracy for the similarity models are a lower bound because similarity was measured separately from choice. If similarity were measured concurrently with choice, the similarity model would likely perform even better.

Understanding the contextual basis of similarity judgments could provide key insights into apparent violations of discounting model predictions. Many discounting models must change discount rates with not only the magnitude and sign of the reward but also the direction of the reward sequence (improving sequences are preferred over declining sequences) and the reward domain (monetary outcomes are discounted more steeply than health outcomes). Here, I demonstrate that similarity judgments can capture how the contexts of reward magnitude and sign influence intertemporal choice. This finding raises the possibility that similarity judgments may also account for other effects of context on intertemporal choices.

In summary, similarity is highly context dependent. Yet, its context dependence offers a powerful test of the similarity model. We can make predictions about how the variation within and between individuals in similarity judgments will influence within- and between-individual variation in intertemporal choices. Combining the rich literature on similarity with process models of decision making could open new avenues of future research on the similarity model and the process of making intertemporal choices.

References

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, *6*(1), 37–66. Retrieved from `doi:10.1007/BF00153759`

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

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, *18*(1), 133–165. Retrieved from `doi:10.1400/140334`

Bolker, B., & R Development Core Team. (2012). *bbmle: Tools for general maximum likelihood estimation.* (R package version 1.0.5.2)

Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference*. New York: Springer.

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. Retrieved from `doi:10.1006/jmps.1999.1282`

Chongsuvivatwong, V. (2012). *epicalc: Epidemiological calculator*. (R package version 2.15.1.0)

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

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42–46.

Dahl, D. B. (2013). *xtable: Export tables to LaTeX or HTML*. (R package version 1.7-1)

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. Retrieved from `doi:10.1037/a0035976`

de Leeuw, J. (2001). *Reproducible research: The bottom line* (Tech. Rep.). Los Angeles: Department of Statistics Papers, UCLA.

Doyle, J. R. (2013). Survey of time preference, delay discounting models. *Judgment and Decision Making*, *8*(2), 116–135.

Doyle, J. R., & Chen, C. (2012). *The wages of waiting and simple models of delay discounting.* Unpublished manuscript, Social Science Research Network: http://papers.ssrn.com/abstract=2008283, Rochester, NY.

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 and Cognition*, *34*(4), 914–928.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (Second ed.). Thousand Oaks, CA: Sage.

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

Goldstone, R. L., & Son, J. (2005). Similarity. In K. J. Holyoak & R. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 13–36). Cambridge, UK: Cambridge University Press.

Green, L., Fristoe, N., & Myerson, J. (1994). Temporal discounting and preference reversals in choice between delayed outcomes. *Psychonomic Bulletin and Review*, *1*(3), 383–389.

Green, L., Myerson, J., & Macaux, E. W. (2005). Temporal discounting when the choice is between two delayed rewards. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*(5), 1121–1133. Retrieved from `doi:10.1037/0278-7393.31.5.1121`

Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory and Cognition*, *25*(5), 715–723.

Hahn, U., & Chater, N. (1998). Similarity and rules: Distinct? Exhaustive? Empirically distinguishable? *Cognition*, *65*(2–3), 197–230. Retrieved from `doi:10.1016/S0010-`

0277(97)00044-9

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

Harrell, F. E., with contributions from Charles Dupont, & many others. (2014). *Hmisc: Harrell miscellaneous.* (R package version 3.14-4)

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

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–1633. Retrieved from doi:10.1038/nn2007

Kacelnik, A. (1997). Normative and descriptive models of decision making: Time discounting and risk sensitivity. In G. R. Bock & G. Cardew (Eds.), *Characterizing human psychological adaptations* (pp. 51–70). Chichester, UK: John Wiley & Sons.

Killeen, P. R. (2009). An additive–utility model of delay discounting. *Psychological Review*, *116*(3), 602–619. Retrieved from doi:10.1037/a0016414

Kirby, K. N. (1997). Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, *126*(1), 54–70.

Kirby, K. N., & Herrnstein, R. J. (1995). Preference reversals due to myopic discounting of delayed reward. *Psychological Science*, *6*(2), 83–89.

Kirby, K. N., & Maraković, N. N. (1995). Modeling myopic decisions: Evidence for hyperbolic delay-discounting within-subjects and amounts. *Organizational Behavior and Human Decision Processes*, *64*(1), 22–30. Retrieved from doi:10.1006/obhd.1995.1086

Komlos, J., Smith, P. K., & Bogin, B. (2004). Obesity and the rate of time preference: Is there a connection? *Journal of Biosocial Science*, *36*(2), 209–219. Retrieved from doi:10.1017/S0021932003006205

Laibson, D. I., Repetto, A., & Tobacman, J. (1998). Self-control and saving for retirement. *Brookings Papers on Economic Activity*, *29*(1), 91–196.

Leisch, F. (2002). Sweave: Dynamic generation of statistical reports using literate data analysis. In W. Härdle & B. Rönz (Eds.), *Compstat 2002–Proceedings in Computational Statistics* (pp. 575–580). Heidelberg: Physica Verlag.

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

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

Luhmann, C. C. (2013). Discounting of delayed rewards is not hyperbolic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1274–1279. Retrieved from doi:10.1037/a0031170

Marewski, J. N., & Olsson, H. (2009). Beyond the null ritual: Formal modeling of psychological processes. *Journal of Psychology*, *217*(1), 49–60. Retrieved from doi:10.1027/0044-3409.217.1.49

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In

M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: The effect of delay and of intervening events on reinforcement value* (Vol. 5, pp. 55–73). Hillsdale, NJ: Lawrence Erlbaum Associates.

McKerchar, T. L., Green, L., Myerson, J., Pickford, T. S., Hill, J. C., & Stout, S. C. (2009). A comparison of four models of delay discounting in humans. *Behavioural Processes*, *81*(2), 256–259. Retrieved from doi:10.1016/j.beproc.2008.12.017

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Myerson, J., & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. *Journal of the Experimental Analysis of Behavior*, *64*(3), 263–276.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* Cambridge, UK: Cambridge University Press.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*(10), 421–425. Retrieved from doi:10.1016/S1364-6613(02)01964-2

R Development Core Team. (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Rachlin, H. (1970). *Introduction to modern behaviorism.* San Francisco, CA: W. H. Freeman and Co.

Rachlin, H. (2000). *The science of self-control.* Cambridge, MA: Harvard University Press.

Rachlin, H. (2006). Notes on discounting. *Journal of the Experimental Analysis of Behavior*, *85*(3), 425–435.

Rachlin, H., Raineri, A., & Cross, D. (1991). Subjective probability and delay. *Journal of the Experimental Analysis of Behavior*, *55*(2), 233–244. Retrieved from doi:10.1901/jeab.1991.55-233

Read, D. (2004). Intertemporal choice. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 424–443). Oxford: Blackwell.

Revolution Analytics, & Weston, S. (2014). *foreach: Foreach looping construct for R.* (R package version 1.4.2)

Robles, E., & Vargas, P. A. (2007). Functional parameters of delay discounting assessment tasks: Order of presentation. *Behavioural Processes*, *75*, 237–241.

Rubinstein, A. (1988). Similarity and decision-making under risk (Is there a utility theory resolution to the Allais paradox?). *Journal of Economic Theory*, *46*(1), 145–153. Retrieved from doi:10.1016/0022-0531(88)90154-8

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

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

Sarkar, D. (2008). *Lattice: Multivariate data visualization with R.* New York: Springer.

Sarkar, D., & Andrews, F. (2013). *latticeExtra: Extra graphical utilities based on lattice.* (R package version 0.6-26)

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

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.

Retrieved from doi:10.1111/cogs.12104

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323. Retrieved from doi:10.1126/science.3629243

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2009). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248–1284. Retrieved from doi:10.1080/03640210802414826

Sozou, P. D. (1998). On hyperbolic discounting and uncertain hazard rates. *Proceedings of the Royal Society of London, Series B*, *265*(1409), 2015–2020.

Stephens, D. W. (2002). Discrimination, discounting and impulsivity: A role for an informational constraint. *Philosophical Transactions of the Royal Society of London, Series B*, *357*(1427), 1527–1537. Retrieved from doi:10.1098/rstb.2002.1062

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory.* Princeton: Princeton University Press.

Stern, N. (2008). The economics of climate change. *American Economic Review*, *98*(2), 1–37. Retrieved from doi:10.1257/aer.98.2.1

Stevens, J. R. (2010). Intertemporal choice. In M. D. Breed & J. Moore (Eds.), *Encyclopedia of animal behavior* (pp. 203–208). Oxford: Academic Press.

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

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352. Retrieved from doi:10.1037/0033-295X.84.4.327

Vlaev, I., Chater, N., Stewart, N., & Brown, G. D. (2011). Does the brain calculate value? *Trends in Cognitive Sciences*, *15*(11), 546–554. Retrieved from doi:10.1016/j.tics.2011.09.008

Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, *40*(1), 1–29.

Wittmann, M., & Paulus, M. P. (2008). Decision making, impulsivity and time perception. *Trends in Cognitive Sciences*, *12*(1), 7–12.

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

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

Zeileis, A., & Grothendieck, G. (2005). zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software*, *14*(6), 1–27.

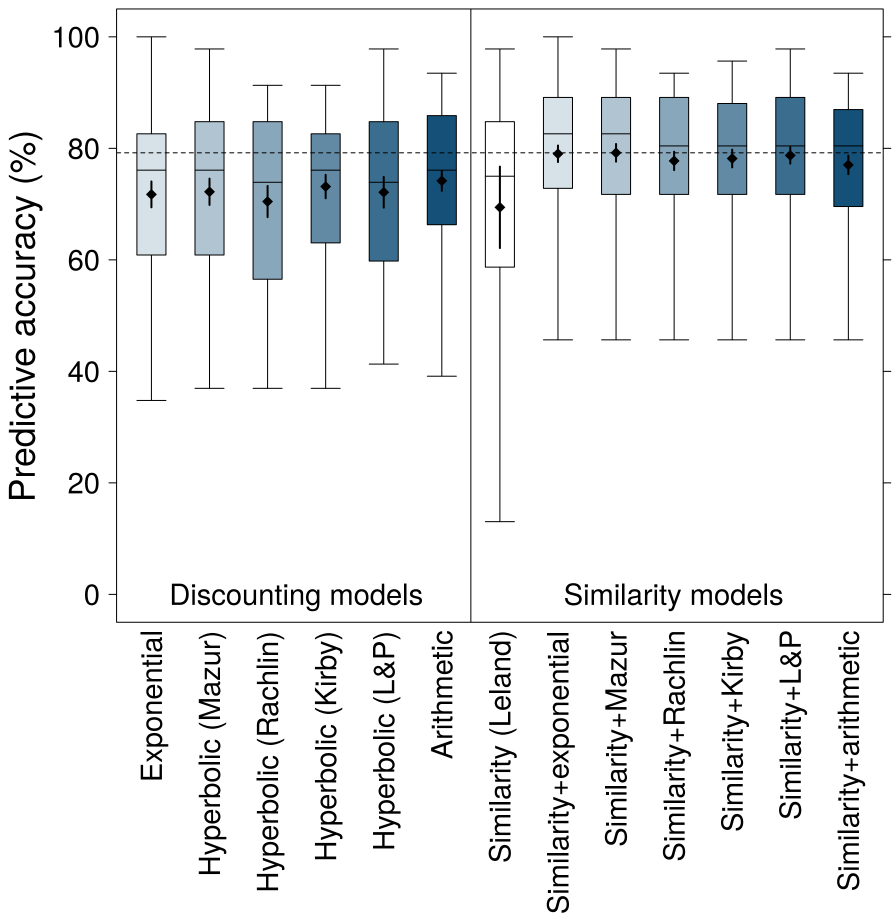
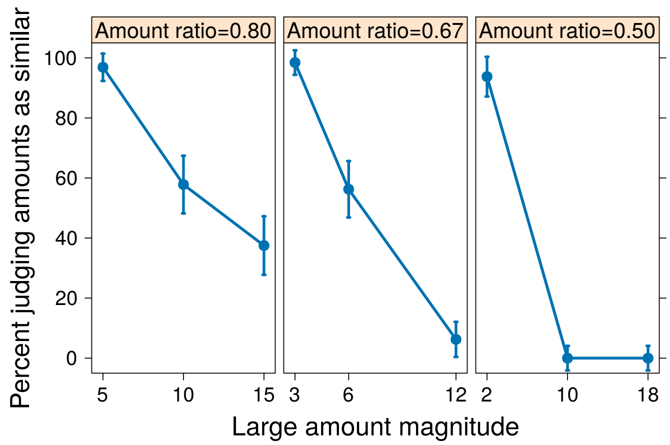


Figure 1. Predictive accuracy of intertemporal choice models in Experiment 1. The mean predictive accuracy per model varied across participants. Diamonds and error bars represent mean and within-subjects 95% confidence intervals. Boxplots show median, interquartile range, and range. Dashed line represents maximum predictive accuracy.

(a)



(b)

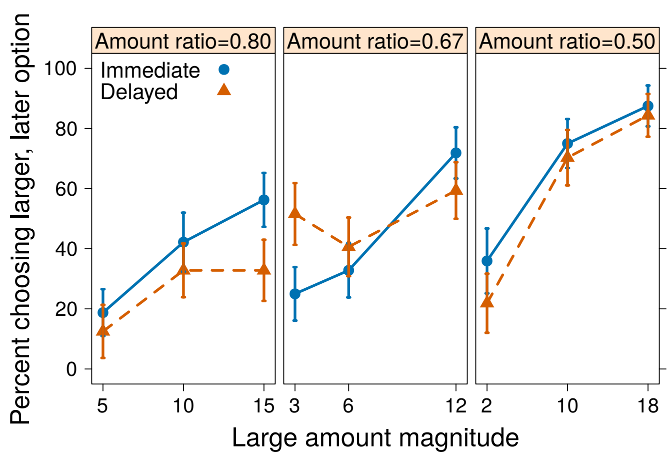


Figure 2. Tests of magnitude effect in Experiment 1. Each panel represents a block of questions with same amount ratio. (a) For the Immediate questions, the short delay is always 0 days (today). For the Delayed questions, the short delay ranges from 4-8 days. The percentage of participants who rated the amounts as similar decreased as the large reward magnitude increased. Similarity judgments were identical if the short delay was immediate or delayed, so a single line is drawn. (b) Choice for the larger, later option in the binary choices increased with the reward magnitude. Points and error bars represent means and within-subjects 95% confidence intervals.

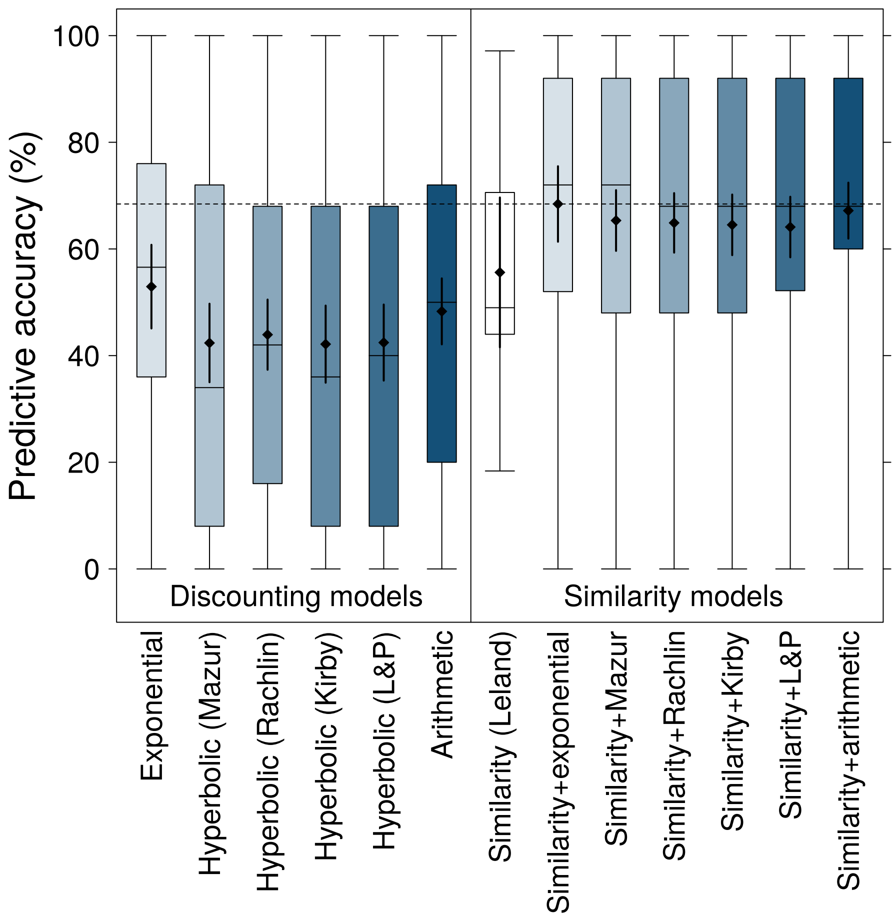


Figure 3. Predictive accuracy of intertemporal choice models in Experiment 2. The mean predictive accuracy per model varied across participants. Diamonds and error bars represent mean and within-subjects 95% confidence intervals. Boxplots show median, interquartile range, and range. Dashed line represents maximum predictive accuracy.

Supplementary Materials

Table S1

Questions and Mean Responses for Experiment 1

Short delay	Long delay	Small amount	Large amount	Mean choice for LL	Standard deviation
0	10	2	3	0.43	0.50
0	10	4	6	0.31	0.47
0	10	8	12	0.49	0.50
0	12	4	5	0.08	0.27
0	12	8	10	0.22	0.42
0	12	12	15	0.24	0.43
0	15	1	2	0.14	0.35
0	15	4	6	0.18	0.39
0	15	5	10	0.67	0.48
0	15	9	18	0.80	0.40
0	22	1	10	0.94	0.24
0	22	3	10	0.82	0.39
0	22	5	10	0.59	0.50
0	22	7	10	0.16	0.37
0	22	7	10	0.25	0.44
3	33	8	12	0.24	0.43
3	33	8	12	0.25	0.44
3	33	8	12	0.25	0.44
3	33	8	12	0.25	0.44
3	33	8	12	0.27	0.45
3	33	8	12	0.37	0.49
4	16	2	3	0.16	0.37
4	16	4	6	0.24	0.43
4	16	8	12	0.65	0.48
4	22	7	10	0.37	0.49
5	25	1	2	0.29	0.46
5	25	5	10	0.69	0.47
5	25	9	18	0.84	0.37
5	38	6	8	0.10	0.30
5	38	6	8	0.12	0.33
5	38	6	8	0.12	0.33
5	38	6	8	0.12	0.33
5	38	6	8	0.14	0.35
5	38	6	8	0.22	0.42
8	22	4	5	0.08	0.27
8	22	7	10	0.53	0.50
8	22	8	10	0.31	0.47
8	22	12	15	0.47	0.50
12	17	8	12	0.90	0.30
12	22	7	10	0.55	0.50
15	22	1	10	1.00	0.00
15	22	3	10	0.98	0.14
15	22	5	10	0.94	0.24
15	22	7	10	0.76	0.43
16	22	7	10	0.86	0.35
20	22	7	10	0.96	0.20

Table S2

Magnitude Effect Questions for Experiment 1

Short Delay (Days)	Long Delay (Days)	Small Amount (€)	Large Amount (€)	k	Delay Ratio	Delay Difference (days)	Amount Ratio	Amount Difference (€)
0	15	1	2	0.07	0.00	15	0.50	1
0	15	5	10	0.07	0.00	15	0.50	5
0	15	9	18	0.07	0.00	15	0.50	9
5	25	1	2	0.07	0.20	20	0.50	1
5	25	5	10	0.07	0.20	20	0.50	5
5	25	9	18	0.07	0.20	20	0.50	9
0	10	2	3	0.05	0.00	10	0.67	1
0	10	4	6	0.05	0.00	10	0.67	2
0	10	8	12	0.05	0.00	10	0.67	4
4	16	2	3	0.05	0.25	12	0.67	1
4	16	4	6	0.05	0.25	12	0.67	2
4	16	8	12	0.05	0.25	12	0.67	4
0	12	4	5	0.02	0.00	12	0.80	1
0	12	8	10	0.02	0.00	12	0.80	2
0	12	12	15	0.02	0.00	12	0.80	3
8	22	4	5	0.02	0.36	14	0.80	1
8	22	8	10	0.02	0.36	14	0.80	2
8	22	12	15	0.02	0.36	14	0.80	3

Table S3

Median Parameter Estimates for Models

Model	Experiment 1	Experiment 2	Experiment 3 Gains	Experiment 3 Losses
Exponential	$\delta = 0.016$	$\delta = 0.009$	$\delta = 0.011$	$\delta = 0.008$
Hyperbolic (Mazur)	$k = 0.02$	$k = 0.01$	$k = 0.02$	$k = 0.01$
Hyperbolic (Rachlin)	$k = 0.03$, $\sigma = 1.2$	$k = 0.05$, $\sigma = 0.76$	$k = 0.06$, $\sigma = 0.75$	$k = 0.13$, $\sigma = 0.55$
Hyperbolic (Kirby)	$k = 0.07$, $\mu = -0.59$	$k = 0.02$, $\mu = -0.3$	$k = 0.05$, $\mu = -0.56$	$k = 0.03$, $\mu = -0.58$
Hyperbolic (Loewenstein & Prelec)	$\alpha = 0.04$, $\beta = 0.03$	$\alpha = 0.08$, $\beta = 0.03$	$\alpha = 45.55$, $\beta = 2.22$	$\alpha = 2.71$, $\beta = 0.23$
Arithmetic	$\lambda = 0.14$	$\lambda = 0.03$	$\lambda = 0.07$	$\lambda = -0.05$

Table S4

Questions and Mean Responses for Experiment 2

Short delay	Long delay	Small amount	Large amount	k	Mean choice for LL	Standard deviation
18	27	7	10	0.33	0.72	0.45
81	117	7	10	0.33	0.44	0.50
88	127	7	10	0.33	0.33	0.48
40	58	7	10	0.50	0.67	0.48
61	88	7	10	0.50	0.65	0.48
75	108	7	10	0.50	0.44	0.50
82	118	7	10	0.50	0.37	0.49
89	128	7	10	0.50	0.44	0.50
73	105	7	10	0.60	0.50	0.50
87	125	7	10	0.60	0.37	0.49
36	52	7	10	0.75	0.65	0.48
50	72	7	10	0.75	0.56	0.50
71	102	7	10	0.75	0.50	0.50
85	122	7	10	0.75	0.37	0.49
141	202	7	10	0.75	0.26	0.44
55	79	7	10	1.00	0.50	0.50
76	109	7	10	1.00	0.43	0.50
83	119	7	10	1.00	0.43	0.50
97	139	7	10	1.00	0.39	0.49
118	169	7	10	1.00	0.41	0.50
146	209	7	10	1.00	0.26	0.44
153	219	7	10	1.00	0.24	0.43
160	229	7	10	1.00	0.20	0.41
195	279	7	10	1.00	0.28	0.45
216	309	7	10	1.00	0.26	0.44

Table S5

Questions and Mean Responses for Experiment 3

Short delay	Long delay	Small amount	Large amount	Choice LL (gain)	SD (gain)	Choice LL (loss)	SD (loss)
0	20	32	55	0.58	0.50	0.21	0.42
0	20	40	70	0.86	0.35	0.21	0.42
0	25	40	55	0.39	0.49	0.46	0.51
0	40	25	35	0.18	0.38	0.36	0.49
0	50	30	85	0.74	0.44	0.11	0.31
10	20	10	18	0.61	0.49	0.29	0.46
10	25	15	35	0.77	0.42	0.14	0.36
10	27	40	65	0.56	0.50	0.11	0.31
10	30	30	35	0.09	0.29	0.36	0.49
10	30	40	62	0.49	0.50	0.61	0.50
10	35	25	34	0.21	0.41	0.36	0.49
10	37	21	30	0.16	0.37	0.57	0.50
10	37	65	75	0.14	0.35	0.39	0.50
10	40	67	85	0.35	0.48	0.46	0.51
10	65	45	70	0.33	0.48	0.50	0.51
10	85	21	30	0.04	0.19	0.46	0.51
20	25	10	12	0.46	0.50	0.36	0.49
20	27	20	26	0.47	0.50	0.18	0.39
20	37	27	30	0.12	0.33	0.64	0.49
20	40	32	45	0.53	0.50	0.36	0.49
20	43	34	35	0.04	0.19	0.61	0.50
20	50	47	60	0.28	0.45	0.86	0.36
20	50	83	85	0.02	0.13	0.54	0.51
20	65	48	55	0.07	0.26	0.57	0.50
20	85	30	35	0.05	0.23	0.46	0.51
30	37	10	12	0.33	0.48	0.14	0.36
30	37	20	24	0.40	0.49	0.25	0.44
30	37	48	55	0.70	0.46	0.29	0.46
30	40	15	19	0.28	0.45	0.29	0.46
30	50	32	43	0.33	0.48	0.25	0.44
30	55	40	50	0.28	0.45	0.25	0.44
30	60	32	55	0.60	0.49	0.75	0.44
30	60	53	55	0.02	0.13	0.32	0.48
30	65	16	24	0.18	0.38	0.25	0.44
30	70	16	30	0.44	0.50	0.39	0.50
30	70	24	55	0.58	0.50	0.18	0.39
30	70	50	80	0.46	0.50	0.29	0.46
30	80	47	58	0.19	0.40	0.50	0.51
30	85	53	55	0.04	0.19	0.68	0.48
30	100	50	74	0.21	0.41	0.39	0.50

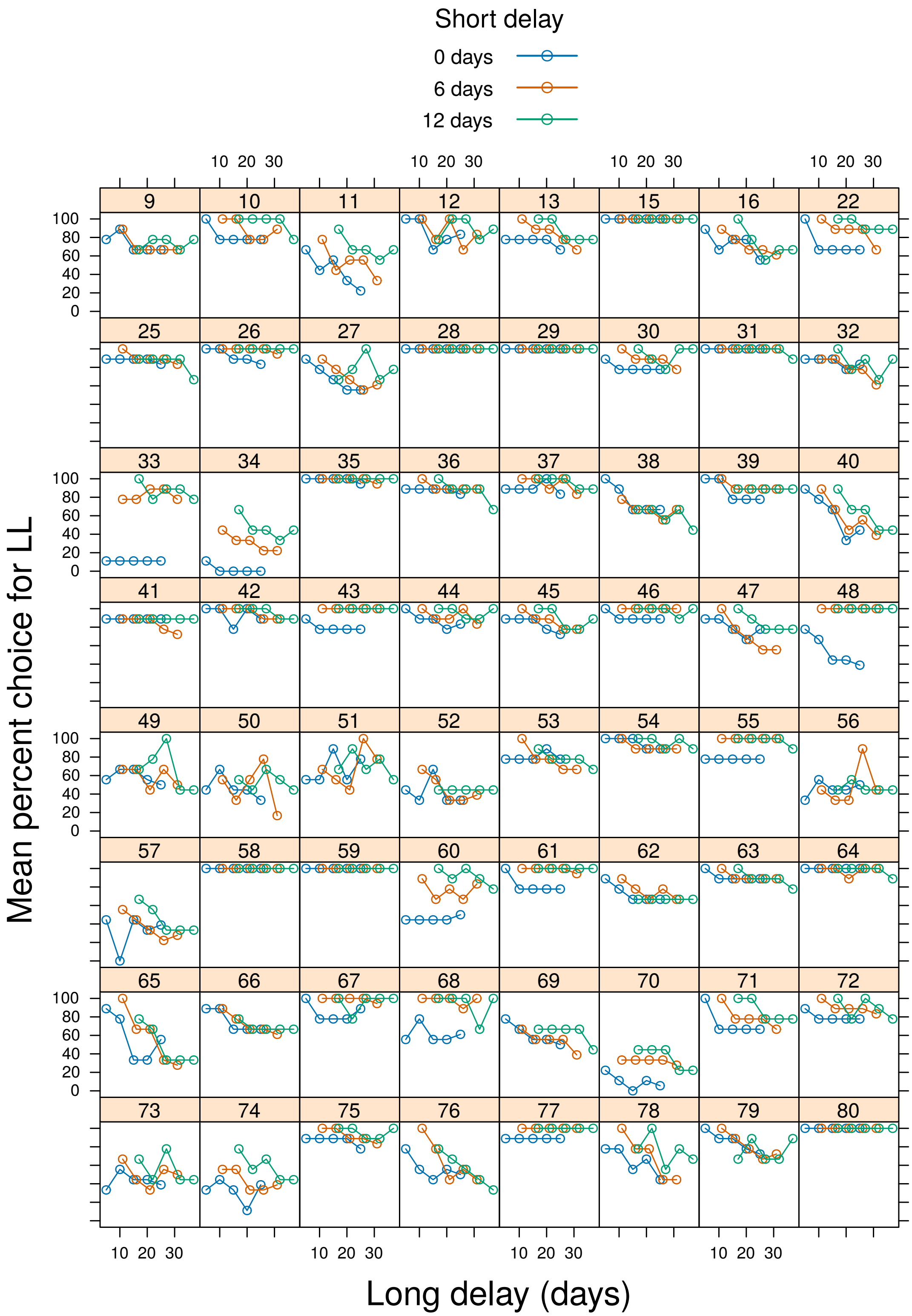


Figure S1. Mean choice percentages for staircase data in Experiment 1. Participants experienced three short delays and five long delays in the staircase phase of Experiment 1.

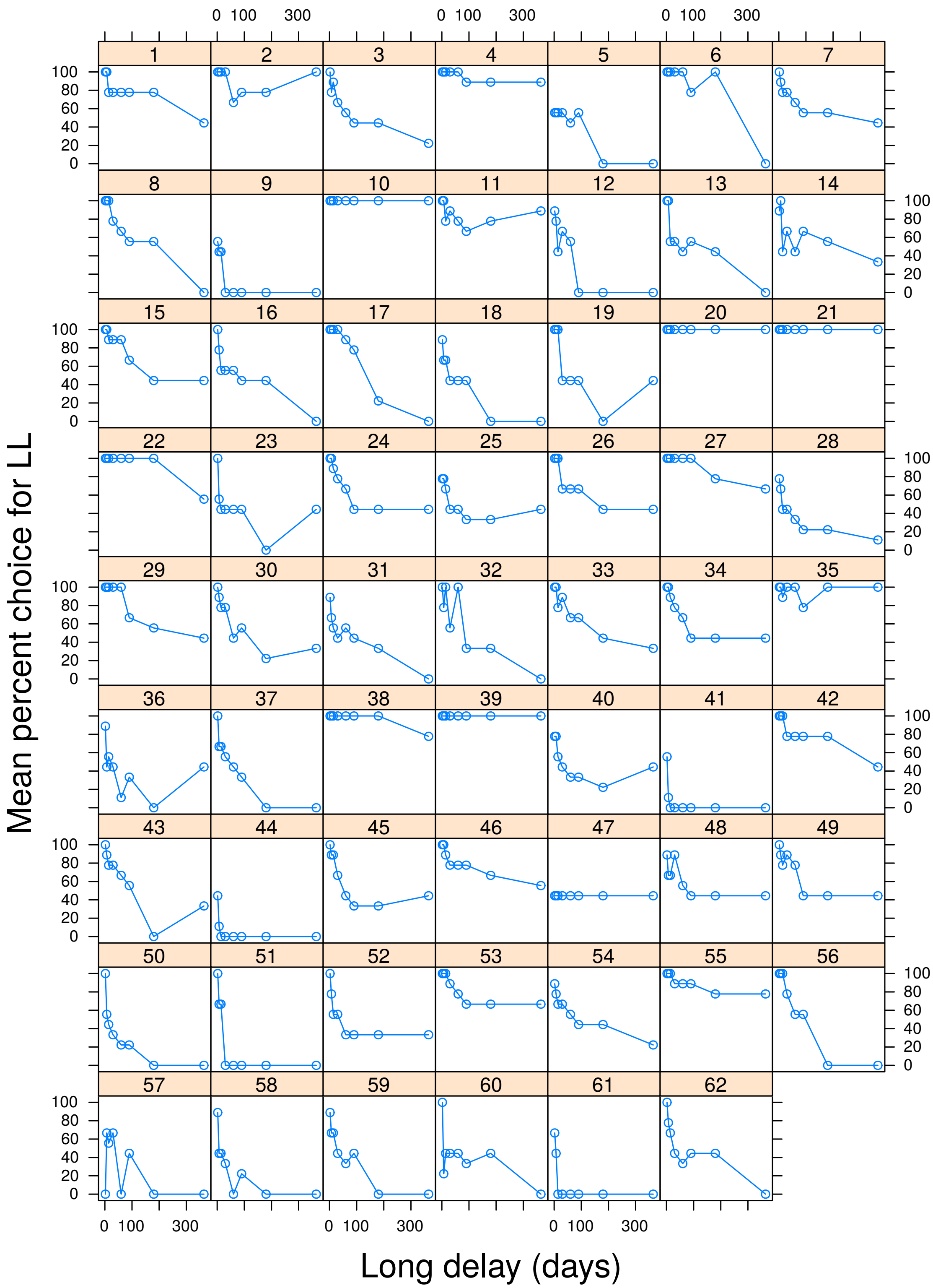
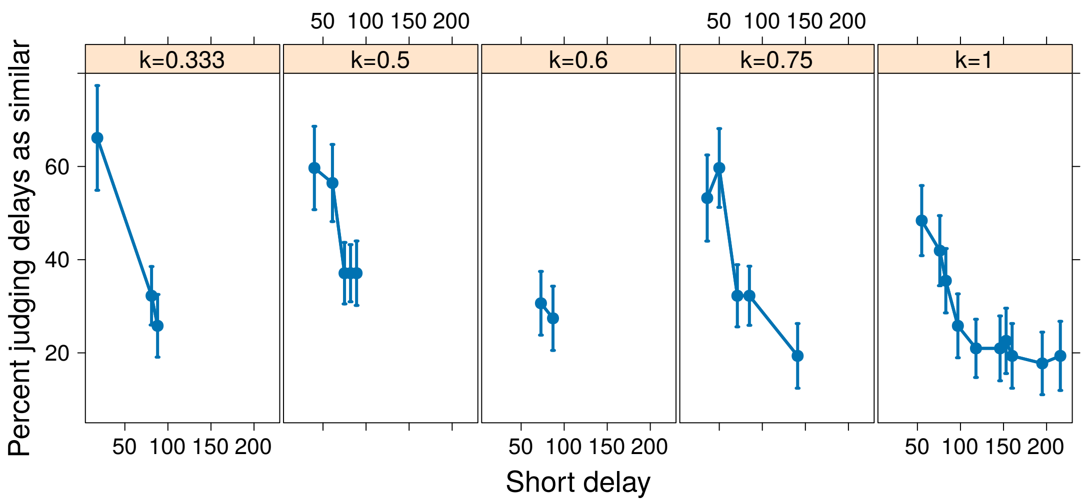


Figure S2. Mean choice percentages for staircase data in Experiment 2. Participants experienced eight long delays in the staircase phase of Experiment 2.

(a)



(b)

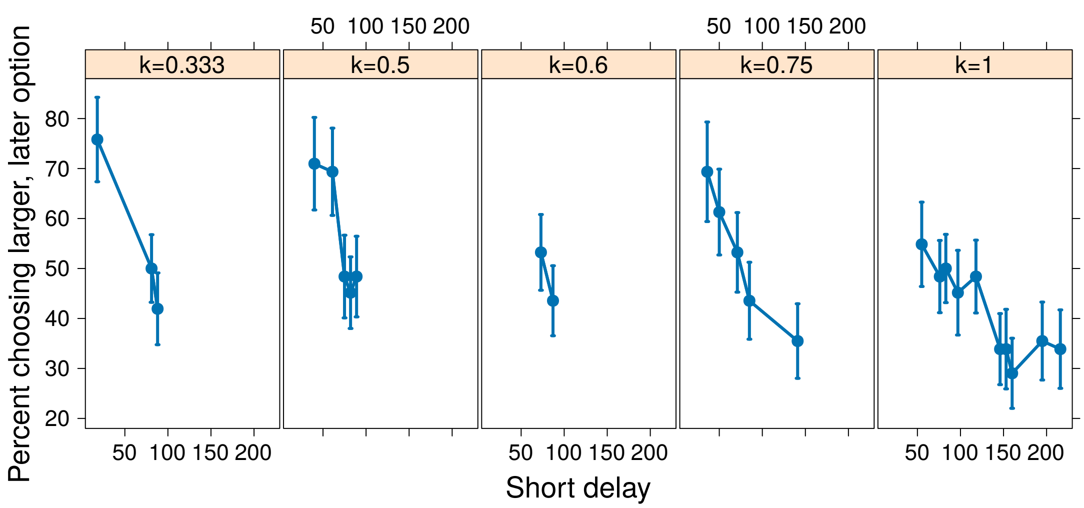


Figure S3. Tests effect of delay in Experiment 2. Each panel represents a block of questions with same k parameter at indifference. (a) The percentage of participants who rated the delays as similar decreased as the short delay magnitude increased. (b) Choice for the larger, later option in the binary choices decreased with the short delay magnitude. Points and error bars represent means and within-subjects 95% confidence intervals.

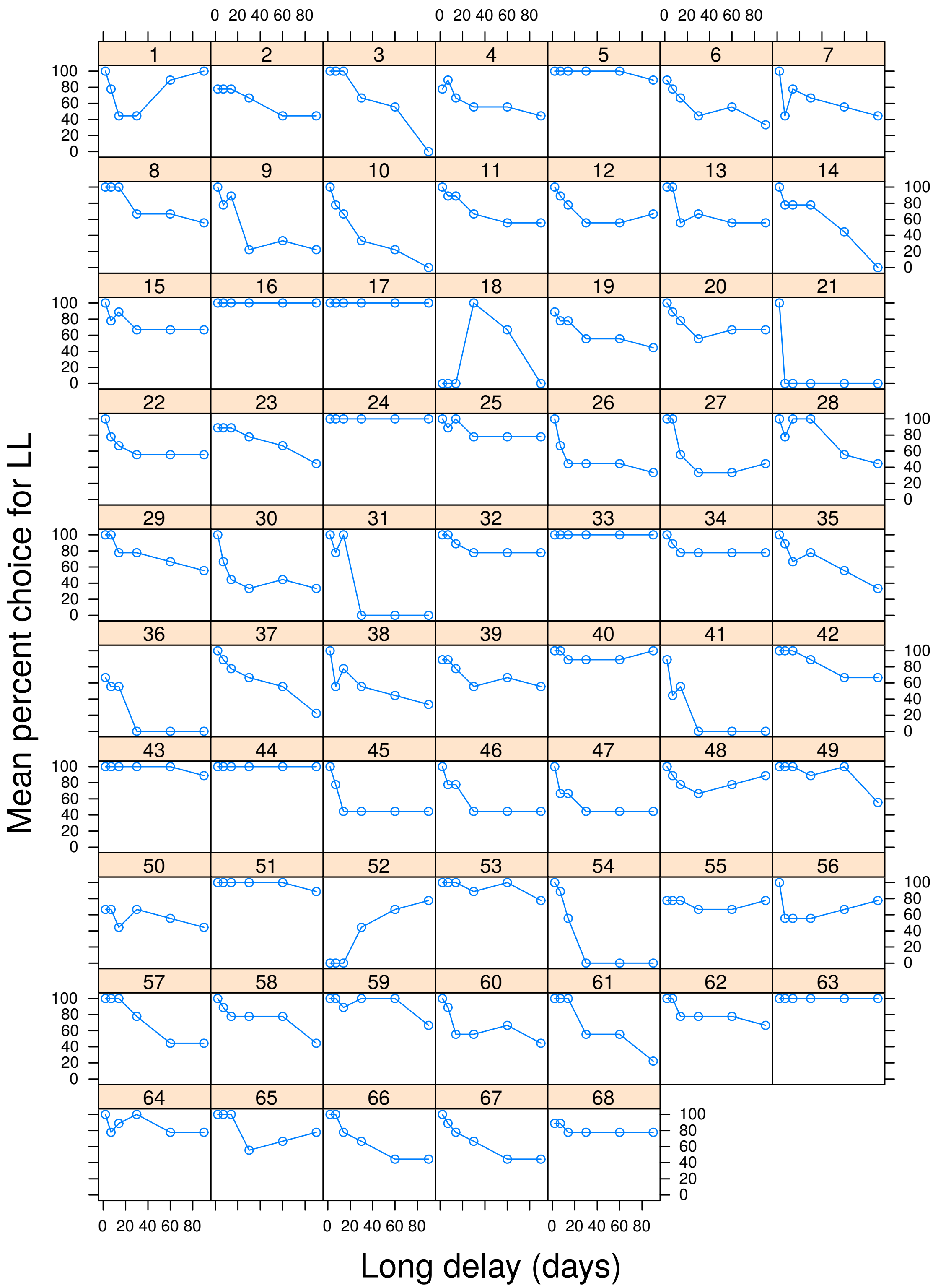


Figure S4. Choice percentages for staircase data in Experiment 3 gain condition. Participants experienced six long delays in the staircase phase of Experiment 3.

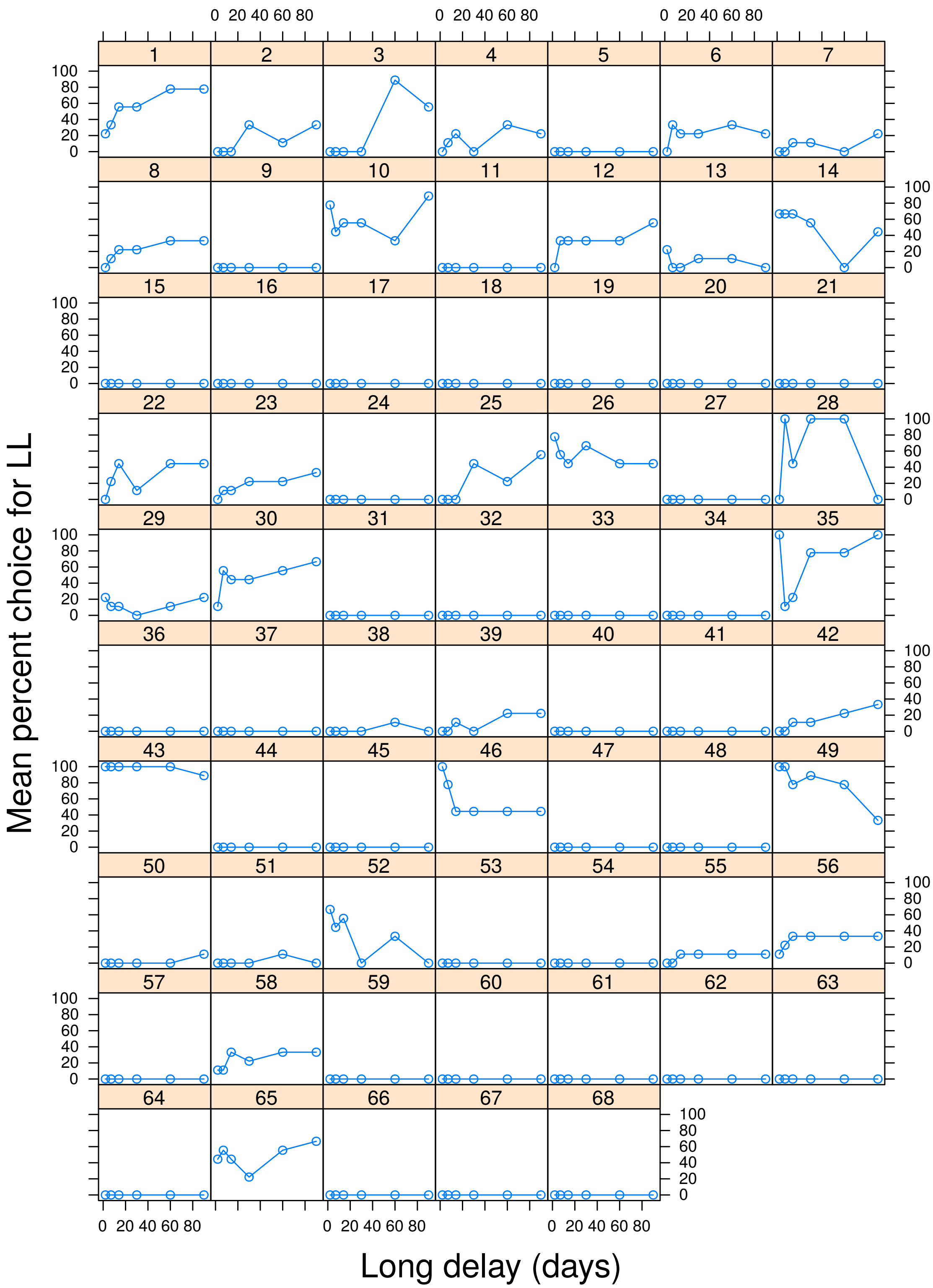


Figure S5. Choice percentages for staircase data in Experiment 3 loss condition. Participants experienced six long delays in the staircase phase of Experiment 3.