Towards a dynamic, probabilistic, and attribute-wise model of intertemporal choice

Junyi Dai and Jerome R. Busemeyer

Indiana University, Bloomington

Abstract

Most theoretical and empirical research on intertemporal choice assumes a deterministic perspective, leading to the widely adopted delay discounting paradigm. As a form of preferential choice, however, intertemporal choice might well be probabilistic in nature. Two empirical studies were conducted to demonstrate this property, in which the delay amount effect, common difference effect and magnitude effect in intertemporal choice were revealed in a probabilistic manner. The results, especially those associated with the delay amount effect, challenge the traditional deterministic view and call for alternative approaches. Consequently, a number of probabilistic models were explored and fitted to the choice response data, including one alternative-wise random utility model, two alternative-wise diffusion models, and six attribute-wise diffusion models employing the general framework of decision field theory. The alternative-wise models were derived from the traditional hyperbolic discount function while the attribute-wise models were built upon direct and/or relative differences in money and delay amounts. Furthermore, response times for intertemporal choice were recorded for the first time and the diffusion models, which assume a dynamic structure, were also fitted to the response time data so that more information can be utilized to find a better model. The results showed that attribute-wise diffusion models involving only direct differences performed the best and were able to account for all three intertemporal effects. In addition, the empirical relationships between choice proportions and response times are consistent with diffusion models and thus favor a dynamic instead of static model structure.

Introduction

Intertemporal choice refers to the situation in which people need to choose among two or more payoffs occuring at different points in time. We can find numerous examples from the economic world and our daily lives which constitute such a scenario. For instance, a decision to deposit part of one's income in a bank instead of spending the money immediately can be interpreted as an intertemporal choice. In this case, one option is to purchase some goods with the money to fulfill one's needs instantly, while the other is to save it in order to get more money and satisfaction in the future. Another example of intertemporal choice may occur when you are waiting at a city bus stop. Sometimes an overdue bus crowded with passengers arrives and you need to decide whether to get on the bus to save time or wait for another bus that is less crowded.

Intertemporal choice has long been an intriguing topic that draws attention of economists, psychologists, and decision scientists. It was introduced by Rae[31]when addressing the issue of interest and later on elaborated by Fisher[10], leading to the well-known discounted utility (DU) model[34]. Psychologists have since put a considerable amount of effort to revise the model from a behavioral perspective. One of the most influential fruits from this endeavor is the hyperbolic discounting model[1, 2, 4, 23, 29], which differs from the DU model mainly in terms of their prediction on time-consistency. At the same time, Loewenstein, Prelec, and Thaler[19, 20, 21, 22] explored intertemporal choice in an attempt to expand related economic models to accommodate a variety of behavioral anomalies revealed in empirical studies. It is clear that both psychological and economic research contribute substantially to our knowledge base of this important topic.

Despite a long history of intensive investigation and a rich literature, we are still far from a good understanding of the underlying mechanism of intertemporal choice, i.e. the emotional and cognitive processes that lead to our intertemporal decisions. Furthermore, some important properties of intertemporal choice, such as its probabilistic nature, have not been systematically probed yet. As a result, the current paper will explore this critical property of intertemporal choice, and at the same time offers a description of the underlying emotional and cognitive processes that account for this property. A brief review of the traditional approaches to intertemporal choice and relevant findings will serve as a good starting point for that purpose.

Traditional approaches to intertemporal choice and relevant findings

Intertemporal choice from a discounting perspective

Most of the traditional studies on intertemporal choice focus on the way people assign subjective values or utilitites to immediate or delayed payoffs. The underlying assumption is that people make an intertemporal choice by first calculating the value of each option and then choosing the option with the higher value. Since a rational person would probably prefer getting a certain amount of money right away over having it postponed into the future, it is plausible to assume that the value of a payoff decreases when it is delayed as opposed to when it fulfills immediately. In other words, the value of a delayed payoff is discounted, which leads up to the concept of delay discounting. Consequently, discovering the appropriate form of discount function that links the objective value of a payoff to its discounted value due to delay became a pivotal issue in the study of intertemporal choice. It turned out that two classes of discount functions attract most attention of the researchers in this area, which we will talk about in what follows.

Exponential and hyperbolic discount functions

As mentioned earlier, intertemporal choice is an important topic in both economics and psychology. Scholars in these two areas, however, have been exploring it from quite different perspectives. Economists prefer to build up their theory and models on abstract axioms followed by rigorous mathematical derivations[10], while psychologists try to generate descriptive models that fit empirical data better[2, 23, 18]. Therefore, the DU model and related exponential discount function became the most popular tool for economists studying economic activities with a temporal component, while the hyperbolic discount function became the favorite apparatus of psychologists to describe the actual pattern of human intertemporal choice. According to the DU model, the appropriate discount function is

$$D(t) = exp(-\delta' t) = \delta^t \tag{1}$$

in which t represents the amount of delay associated with a payoff, and δ is a parameter representing the rate of discounting. The value of δ is typically between 0 and 1 to guarantee delay discouting. One important corollary of this form is the property of dynamic consistency, which suggests that people's preference between two intertemporal options should remain the same if both options are brought forward by the same amount of time. That is to say, the preference between the two options should not change as time passes. Dynamic consistency is usually a required condition in economic analysis because it is assumed to be a demonstration of rationality.

In reality, however, human beings usually do not behave in a consistent way as suggested by economic theories. When intertemporal choice is of concern, it means that people tend to alter their preference between two intertemporal options when both of them become closer or are further delayed to the same degree. Consequently, a different discount function is necessary to describe this pattern of inconsistency. One of the candidates is the widely adopted hyperbolic discount function as follows

$$D(t) = \frac{1}{1 + kt} \tag{2}$$

in which t has the same meaning as in Eq. 1 and k is the counterpart of δ. The value of k should be positive to ensure delay discouting. Unlike the exponential discount function, the hyperbolic discount function predicts dynamic inconsistency of intertemporal choice, which is typically found in empirical data[3, 9, 12, 27, 38]. A number of similar but different models were also investigated, including two-parameter hyperboloid model [26, 13] and two-parameter hyperbola model[30]. See [24] for a comparison of the aforementioned models.

Several important effects in intertemporal choice

Besides the phenomenon of dynamic inconsistency, researchers in various area have also investigated some other effects regarding intertemporal choice. For example, Kirby and Marakovic [18]studied the impact of reward magnitude on the discounting parameter in the hyperbolic discount function (i.e. Eq. 2) and found that it is a decreasing function of the size of the delayed reward. Similar results were reported by [14] and [8]. This relationship between the discounting parameter in the hyperbolic discount function and the size of delayed reward is usually termed as the magnitude effect in intertemporal choice. In an attempt to put intertemporal and risky choice under a common theoretical framework, Prelec and Loewenstein [28] summarized the analogy between these two research topics in terms of a number of related effects and proposed a set of assumptions upon which plausible accounts for these effects can be generated. The magnitude effect in intertemporal choice is just such an effect, whose counterpart in risky choice is the so-called "peanuts effect", i.e. risk taking for small gains and risk aversion for large gains. In order to explain the magnitude effect in intertemporal choice, the authors put forward an assumption of increasing proportional sensitivity, which suggests that multiplying the value of a specific attribute across all alternatives will shift more weight to the attribute. Consequently, for example, if one is indifferent between receiving 10 dollars now and receiving 20 dollars in 20 days, he/she should prefer receiving 200 dollars in 20 days to receiving 100 dollars now. This is due to the fact that the reward amount is multiplied by a constant, i.e. 10, and thus it becomes more important in the decision. As a result, an option with a higher value on the attribute will be more attractive than the other one. If we calculate the discounting rate for the two delayed rewards, clearly the discounting rate for the latter will be lower.

The assumption of increasing proportional sensitivity can be applied to the attribute of delay amount as well. Specifically, according to the assumption, if the delay amounts of both options in an intertemporal choice problem are increased by a common multiplicative constant, delay amount will become more decisive and the option with a shorter delay (i.e. a more desirable option in terms of delay amount) will be more preferable. For example, if one is indifferent between receiving 10 dollars in 10 days and receiving 20 dollars in 20 days, he/she should prefer receiving 10 dollars in 20 days to receiving 20 dollars in 40 days. We label this as the delay amount effect to differentiate it from the (reward) magnitude effect discusses above. This effect has not be as intensively studied as the magnitude effect, but it deserves a close look if we want to obtain a comprehensive understanding of intertemporal choice and develop cognitive models accordingly. In addition, it turns out to be a useful tool to demonstrate the probabilistic nature of intertemporal choice, which is one of the major goals of this paper.

Another effect in intertemporal choice explored in [28] is the common difference effect, which can be viewed as a generalized case of the immediacy effect. Again, suppose that one has no preference between getting 10 dollars in 10 days and getting 20 dollars in 20 days. According to the common difference effect, if we reduce both delays by the same amount, say, 5 days, then the person should prefer the option with a shorter delay. That is to say, he/she will choose getting 10 dollars in 5 days as opposed to getting 20 dollars in 15 days. If, by reducing both delays, the early option will take place immediately, the resultant change in preference is actually a demonstration of the immediacy effect. To account for these two effects, Prelec and Loewenstein added an assumption of decreasing absolute sensitivity to their general framework. Specifically, the assumption suggests that increasing the absolute magnitude of an attribute on both alternatives by a common additive constant will make the attribute less significant for the decision. In other words, to make an attribute more decisive, we need to decrease the absolute magnitude of the attribute by a common (positive) constant. In the current situation, it means that if both options are brought forward by the same amount of time, the earlier option will appear more desirable. Obviously this is also related to the phenomenon of dynamic inconsistency mentioned above: When the delays are substantially increased or decreased to the same degree, a preference reversal will show up according to the common difference effect and the assumption of decreasing absolute sensitivity.

Alternative approaches to intertemporal choice

Although intertemporal choice has long been studied from a discounting perspective and a variety of effects have been explored, leading to several distinct models, most of the conclusions and interpretations are based on a specific approach to this important phenomenon. First of all, a large majority of existing models on intertemporal choice, including the hyperbolic discounting model and its variants, assume a deterministic view on human choice behavior. According to this view, when required to make choices between the same pair of options repeatedly, an individual will always have the same preference and thus choose the same option. Second, to the best of our knowledge, all existing models are static in the sense that they do not provide a description of the underlying processes giving rise to the explicit responses. For the same reason, none of them ever attempts to address a very important measure in psychological research, i.e. the response time associated with a specific choice. Third, most of the current models presume that intertemporal choices are accomplished in an alternative-wise manner (see [35] for a discussion). Such an approach demands that people first figure out the utility of each option independently and then make an explicit choice by comparing their utilities. This concept is one of the core assumptions of the discounting perspective on intertemporal choice; indeed, delay discounting is such an important facet of intertemporal choice that the two terms have become interchangeable. In a nutshell, most traditional models of intertemporal choice assume a deterministic, static, and alternative-wise view on the topic, which might impose an unnecessary constraints. Therefore, the current article is intended to transcend these boundaries to introduce a different, and potentially more realistic account of intertemporal choice. First, let us explore several potential alternatives to the traditional approach.

Deterministic vs probabilistic approaches

A probabilistic approach to intertemporal choice, which does not assume perfect consistency in people's preference between a pair of options presented repeatedly, is a reasonable alternative to the traditional approach. Although the probabilistic nature of intertemporal choice has not been intensively examined in the past, it is not hard to find its counterpart in risky choice scenarios. Ever since the early days of behavioral studies on risky choice, strong support for its probabilistic nature has been reported. For example, Mosteller and Nogee [25]demonstrated that individuals were often inconsistent in their preferences for simple gambles over repeated occasions. According to a deterministic perspective on risky choice, the preference function should assume a step form with a leap at which the two options are indifferent for an individual. The empirical preference function, however, was strikingly similar to a psychometric function typically obtained from a perception study. This suggests that a deterministic model will be insufficient to account for the complexity of human risky choice. It might well be the case that the same pattern will occur in intertemporal choice. Furthermore, when the method of limits is utilized to obtain indifferent pairs of intertemporal options as in classical psychophysical studies on perception, individuals' cutoff points tend to vary across occasions and presenting orders. In other words, people's preference between certain pairs of intertemporal options might change back and forth, which is beyond the scope of a deterministic model. The current paper will explore this property of intertemporal choice and develop models accordingly.

Static versus dynamic approaches

Another deficit of the traditional approach to intertemporal choice is its failing to offer an account of the underlying processes which culminate in one's explicit choice. Although static models of human decision making are relatively easier to construct and are capable of explaing a variety of empirical results, their omission of the underlying emotional and cognitive processes renders them less competitive than dynamic ones. In addition, without a process description, static models of decision making have nothing to say about response time and its distribution. Nevertheless, response time obviously provides important information about how people make explicit responses and can be utilized to distinguish between models with similar predicitions on choice probabilities. A number of dynamic models of decision making have hence been proposed and applied to empirical results that are beyond the means of static models. Among them, decision field theory (DFT) by Busemeyer and Townsend[6, 7] may be the most successful one. It is a dynamic model in the sense that it describes in detail the deliberation process when several competing options are presented and one needs to determine which to choose. Given the similarity between intertemporal choice and risky choice ([28]), it is quite likely that the deliberation process assumed by DFT can also be utilized to account for the effects regarding intertemporal choice.

Alternative-wise versus Attribute-wise approaches

As mentioned earlier, the traditional approach to intertemporal choice assumes an alternative-wise perspective, which is consistent with the concept of utility maximization frequently invoked in numerous theories on choice behavior, such as the expected utility theory [39] and cumulative prospect theory [16, 37]. However, some computational models of risky choice suggest that people actually employ an attribute-wise approach to make a decision. For example, the priority heuristic [5] suggests that decision makers first compare the minimum outcomes between two gambles of gains and choose the one with a higher minimum outcome if the corresponding difference is no less than one tenth of the maximum gain. If the stopping condition at the previous stage is not satisfied, a comparison of probabilities will follow. In other words, according to the priority heuristic, people reach a decision by comparing different options along specific attributes. Similarly, DFT also assumes an attribute-wise approach in the sense that differences within various attributes across options are the building blocks of the evidence accumulation process. Given the apparent analogy between intertemporal and risky choice, it may well be the case that people adopt an attribute-wise strategy when presented with a pair of intertemporal options with different gains and delays. In fact, Scholton and Read (2010) proposed an attribute-wise but deterministic model to intertemporal choice which accommodates some anomalies that the alternative-wise discounting approach cannot address. In this article, we will explore a different family of attribute-wise models which assume a probabilistic perspective as well. Specifically, we will develop diffusion models on intertemporal choice utilizing the general framework of DFT. Such models might provide us with a better description of the comparison process when choosing between options occuring at different time points.

A brief review of DFT and Proportional Difference (PD) Model

Basics of DFT

Since DFT will be the major tool for developing cognitive models of intertemporal choice in the current paper, it is necessary to review the critical features of the theory before exploring more technical details. First of all, DFT belongs to the broad class of sequential sampling models and it was the first application of such models to decision making under risk and uncertainty. One of the common assumptions of sequential sampling models is an evidence accumulation process upon which a final decision is made. Specifically, information or evidence for and against each option is collected and accumulated sequentially during the deliberation stage until the strength of evidence for one of the options reaches a threshold. At that time, a decision is made to choose the exact option that reaches the threshold, and the deliberation time, along with the time for other non-decisional components, determine the actual response time. There are two general types of sequential sampling models in terms of the stopping rule used in the accumulation process: models with a relative stopping rule and those with an absolute stopping rule. For a sequential sampling model with a relative stopping rule, evidence in favor of one option is evidence against the other option, while models assuming an absolute stopping rule usually presume that there are separate accumulators or counters for different alternatives and only favorite evidence is accumulated. In other words, the amount of evidence accumulated for various options do not need to be correlated. Another criterion to categorize sequential sampling models is the time scale involved in the accumulation process. Some models assume that evidence is sampled at discrete time points while others are more general in that evidence can be collected and accumulated continuously along time. DFT turns out to be a model with a relative stopping rule and continuous time scale, i.e., a diffusion model, although in some applications it can be approximated by a discrete-time random walk model. See [32] for a review and comparison of sequential sampling models for binary decisions.

Usually there are five parameters involved in the implementation of DFT to binary choice tasks. The first one is the drift parameter, d, which reflects the mean rate of information accumulation. The larger the drift parameter, the less time it takes to a make a decision on average. The second one is the threshold on evidence strength, θ, which determines when the accumulation process should stop. The higher the threshold, the longer it takes to make a choice. The third parameter is the initial level of evidence in the accumulation process, which can be viewed as a measure of bias towards a specific response. The fourth one, T_{er}, represents the amount of time associated with non-decisional components in a specific task; it is required for fitting response time distributions. Finally, the diffusion parameter, σ, reflects the amount of variance in the instantaneous rate of evidence accumualtion within the diffusion process. If there is no variability in instantaneous drift rate at all, the trajectory of accumulated evidence will constitute a straight line and the response and response time will be deterministic. Therefore, a non-zero σ is necessary for a probabilistic model of binary choices. It is more often than not thatσcan be treated as a scaling parameter and thus it is unnecessary to estimate it explicitly. However, in our application of DFT to intertemporal choice, σ will be determined from attribute values and other parameters, and thus not a trivial one. See for more details about potential diffusion models on intertemporal choice.

Successful applications

Although developed initially as a cognitive model of decision making under risk and uncertainty, DFT was later on generalized to a variety of other task environments with considerable success. For example, [33] proposed a multiple-alternative version of DFT to account for three well-established contextual effects concerning triadic choices, i.e., the similarity effect, the compromise effect, and the attraction effect. Before that, a number of distinct theories had been put forward to account for these effects separately, but none of them can provide a consistent explanation for all the three effects as DFT. To the contrary, the revised version of DFT can predict all the effects simultaneously with a parsimonious and consistent set of assumptions about the underlying deliberation process. Similarly, [15]developed a sequential value-matching model based on DFT indifference model to account for various preference reversal phenomena that had previously eluded a general theory. All these fruitful applications of DFT suggest that it really captures the fundamental processes involved in human choice behavior. As a result, we can apply the general framework of DFT to intertemporal choice tasks to compare to other models.

Basics of the PD model

The Proportional Difference (PD) model [11] is another probabilistic model of choice behavior based on which the current work was developed. This is an attribute-wise model capable of explaining several important violations of normative axioms of decision making. According to the PD model, when people need to choose between a pair of options with multiple attributes, they first compute the proportional difference within each attribute and then rely on a linear combination of these differences to obtain a general evaluation of each option. Consider two options defined as $A = (a, p)$ and $B = (b, q)$, where a and b are values of the first attribute, and p and q are values of the second. Also let us suppose that $a > b$ and $p < q$. In the simplest version of the model, the proportional difference within the first attribute is defined as $\frac{max\{|a|,|b|\}-min\{|a|,|b|\}}{max\{|a|,|b|\}}$ and similarly the proportional difference in the second attribute is defined as $\frac{max\{|p|,|q|)-min\{|p|,|q|\}}{max\{|p|,|q|\}}$. Because option A is superior to option B on the first attribute but less attractive on the second, subtracting the second difference from the first one will result in a general evaluation of option A. This general evaluation will then serve as the mean of a normal distribution, based on which the choice probability of option A can be determined. The other two quantities required for determining the probability of choosing option A are a parameter on personal decision threshold ,δ, and the standard deviation, σ, of the normal distribution. Specifically,

$$P(A) = Pr(z \leq (d - \delta)/\sigma)$$

in which $d = \frac{max\{|a|,|b|\}-min\{|a|,|b|\}}{max\{|a|,|b|\}} - \frac{max\{|p|,|q|)-min\{|p|,|q|\}}{max\{|p|,|q|\}}$. The rationale for the equation is that, according to the PD model, option A will be chosen if and only if the general evaluation plus a normally distributed disturbance exceeds the persoanl decision threshold. Consequently, we can interpret σ as the standard deviation of the random disturbance on the general evaluation.

Distinctions between DFT and the PD model

The PD model is distinct from DFT in two major aspects. First, it is a static model without any account of the underlying deliberation process and thus unable to predict response time distributions. This is a severe drawback of static models which makes them less valuable than dynamic ones like DFT. Second, although both models are attribute-wise in the sense that differences within attributes are the raw material from which a general preference level can be generated, DFT usually utilizes direct differences between options ($a - b$ and

$p - q$ in the example above) while the PD model relies on proportional or relative differences. The second distinction might lead to quite different predictions on choice probability between the two models. One important property of the PD model is the constant ratio rule, which states that increasing the attribute values of both options by the same porportion will not produce a change in choice probability. To the contrary, DFT usually predicts a different choice probability in this case. One of the goals of the current article is to examine the validity of both types of differences in intertemporal choice and develop models accordingly.

Basics of random utility models

Although diffusion processes are widely used to model probabilistic phenomena, several other ways exist for the same purpose. For example, we can also develop probabilistic models on intertemporal choice by introducing random components into the traditional deterministic models to generate corresponding random utility models. Historically, random utility models might be the first class of models explored to account for the probabilistic nature of preferential choice. The major difference between random utility models and deterministic ones lies in the way that the utility of a given payoff is assigned. According to a deterministic interpretation of utility, any option or payoff is associated with a fixed utility across trials. Therefore, when people are required to choose between a pair of options repeatedly, their preference should not change from trial to trial. To the contrary, a random utility model assumes that the utility of a given option might vary across trials and thus people's preference between the same pair of options may change from time to time. It is worthy noting that, although random utility models differ from deterministic ones in terms of utility variability, the decision rules are the same for these two types of models. That is to say, both classes of models assume that people always choose the option with a higher utility at a given instant, no matter whether it is a realization of a random variable at that moment or a constant value across time. Mathematically, the probability of choosing option A from a pair of options {A, B} equals

$$P(A|\{A,B\}) = P(U_A > U_B|\{A,B\}) \tag{3}$$

in which U_A and U_B are the random utility of options A and B respectively. Equation 3 is the general form of random utility models. In order to apply random utility models in real situations, we still need to specify the joint distribution function of the random utilities so that choice probabilities can be determined accordingly. One commonly used distribution is the multivariate normal distribution. In this case, the utility of each option follows a univariate normal distribution and the utilites jointly follow a multivariate normal distribution. If we can further assume that the utilities are independent of one another and the variance of marginal distributions are the same, the resultant random utility model is actually Thurstone's Case V model. Specifically, this model assumes that

$$U_A - U_B \sim N(\mu_A - \mu_B, \sigma^2)$$

and thus

$$P(A|\{A,B\}) = \Phi(\frac{\mu_A - \mu_B}{\sigma}) \tag{4}$$

in which μ_A and μ_B represent the average utility of options A and B respectively, and σ is a measure of the variability in choice response. A Thurstone Case V model on intertemporal choice will be explored and compared to diffusion models in this article to show the advantage of the latter.

Purposes of the study

The current paper is intended to fulfill five major goals. First, it will demonstrate that, just like in risky choice scenarios, people's preference for intertemporal options are probabilistic in essence. Previous research on this topic simply ignored the issue of preference variability in empirical data or treated it as a trivial component which could be minimized by using more complicated models. It is very likely that we might miss precious information revealed in actual data due to our neglect of the uncertainty in intertemporal preference. Second, dynamic and probabilistic models will be developed to account for the probabilistic nature of intertemporal choice and to explain the relevant effects accordingly. We will show that sampled differences in the evidence accumulation process play a critical role in these interpretations. Third, the same diffusion models will be fitted to the response time data so that dynamic models can be further compared to one another and the resultant winning model could obtain more support. To our best knowledge, this is the first time that response time in intertemporal choice or delay discounting tasks has been seriously examined and modelled. Fourth, both attribute-wise and alternative-wise models using the general framework of DFT will be explored and compared. Most existing models on intertemporal choice assume an alternative-wise perspective but an attribute-wise perspective may actually perform better. Finally, the models based on DFT will be compared to a random utility model extended from the traditional hyperbolic model to show the advantage of dynamic models over static ones.

Study 1

To demonstrate the probabilistic nature of intertemporal choice, an experiment was conducted using a broad range of intertemporal choice questions. The stimuli were so chosen that the delay amount effect, common difference effect, and magnitude effect in intertemporal choice could be measured in a probabilistic manner. That is to say, the actual choice proportion of either the smaller-but-sooner (SS) option or the larger-but-later (LL) option would change gradually while attribute values varied as required by those effects. After establishing the probabilistic nature of intertemporal choice, a variety of cognitive models were fitted to the empirical data on choice response. Diffusion models were also fitted to data on choice response and response time simultaneously so that we can gain more confidence in the resultant winning model.

Method

Materials

Because previous research on delay discounting revealed substantial individual difference in discounting rate, it is necessary to generate distinctive intertemporal choice questions for each subject. In this way, actual choice proportions will vary in a wide range rather than getting stuck to extreme values. Therefore, for every subject, an adjustment procedure was used to generate three pairs of approximately indifferent options, one for each of the aforementioned effects in intertemporal choice. In each case, three of the four attribute values involved in a pair of intertemporal options were fixed and the remaining one varied from trial to trial according to subjects' responses to the previous question. For the questions related to the delay amount effect, the shorter delays were fixed at 20 days; the longer delay were fixed at 60 days; and the larger rewards were always 36 dollars. The remaining attribute value, i.e., the smaller reward amount, was initially set at 20 dollars and altered contingent on subjects' responses. For example, if one chose the LL option for a specific pair of options, the smaller reward amount would increase in the next question, and vice versa. Similarly, for the questions related to the common difference effect, the shorter delays were fixed at 20 days; the longer delays were fixed at 60 days; and the larger reward amounts were always 32 dollars. The smaller reward amount started from 16 dollars and again changed according to subjects' previous response. Finally, for the questions concerning the magnitude effect, the smaller rewards, the larger rewards, and the short delays were fixed at 20 dollars, 40 dollars, and 8 days respectively. The longer delay was initially set at 20 days and changed in the same manner as the other questions.

For each effect, 600 formal questions were then created based on the indifferent pair generated using the adjustment procedure. Specifically, in the formal questions concerning the delay amount effect, the longer delays were always three times as long as the shorter ones, which ranged between 1 day and 40 days. For each pair of delays, associated reward amounts were then jiggled from those in the indifferent pair to generate 15 questions that were a little different from one another but practically the same. The purpose of this manipulation was to avoid having subjects respond to the same choice questions repeatedly to minimize the potential impact of memory. Furthermore, the tens digits of the smaller rewards in these questions were always the same, and this was also the case for the larger rewards. (See Appendix A for exemplar questions.) All in all, 600 (40 by 15) questions were generated based on the indifferent pair in this way. The very method was employed to generate the same number of questions for the common difference effect and magnitude effect respectively. See Appendix A for a sample of the questions a typical subject answered in Study 1. See Figure 1 for a screenshot of the experimental software used in the study.

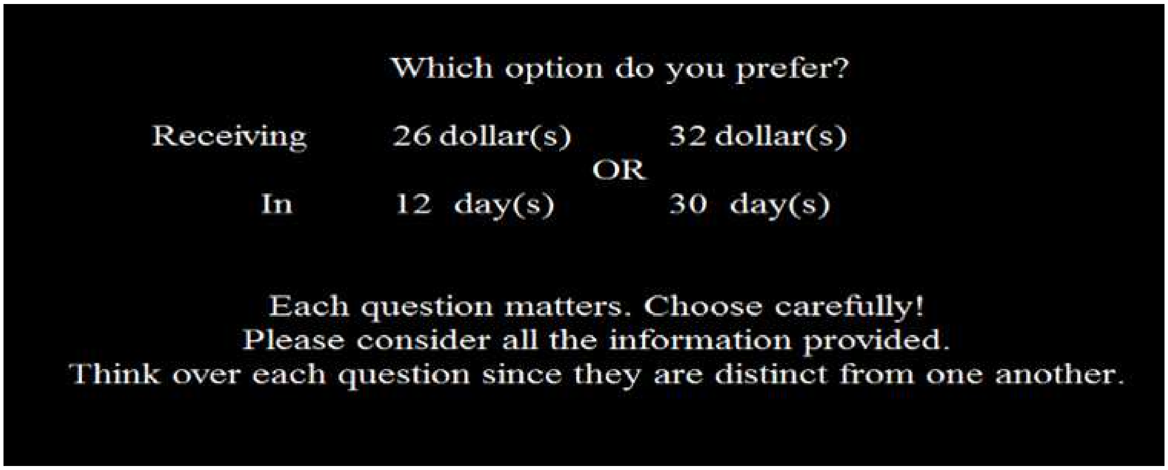

Figure 1. A screenshot of the experimental software used in Studies 1 and 2. Participants reported their decisions by clicking the left or right mouse button for options at the corresponding position. The SS options were always shown on the left, and the LL options were always shown on the right.

Participants

Ten participants (5 females and 5 males) with an average age of 27 were recruited at a public research university via advertisement on notice boards. All the participants received payment for their participation. Specifically, for each participant, one intertemporal choice question was randomly picked from his/her question set and the person was paid the amount of money he/she chose in that question. In addition, there was a baseline payment of 16 dollars in addition to the amount contingent on the randomly selected question. The date of payment was also determined by the time delay participants chose in the randomly selected question. The average payment was about 42 dollars.

Procedure

This study consisted of four sessions, with subsequent sessions for each subject one week apart. All instructions and questions were presented on a computer screen and subjects used a mouse to make responses. The experimental setting was realized by a set of programs in Matlab together with the Cogent toolbox to record both choice responses and response times. In each session, subjects needed to answer 450 intertemporal questions by indicating their choice with a mouse click. To minimize the fatigue effect, two major breaks and contingent short breaks were inserted into each session. Questions associated with the three intertemporal effects were included in each session, and they were divided into the four sessions in such a way that questions in different sessions were comparable to one another and each session contained the same number of questions of each type. In each session, different types of questions were presented in a random order so that successive questions appeared to be irrelevant to one another. The adjustment procedure to generate indifferent pairs was run in the first session before all formal questions were created and presented. Furthermore, a practice section was provided at the beginning of the first session so that subjects could get familiar with the intertemporal choice task. Subjects were instructed throughtout the whole study to take into account all pieces of information involved in each question and think them over before making a choice. They were also reminded of the payment plan that one question would be randomly picked and they would get the chosen amount of money at the specified delay. In the first two sessions, subjects were instructed to make careful choices, while in the last two sessions, they were instructed to make careful but quick responses. If their responses were too fast in the first two sessions or too slow in the last two sessions, a warning sign would pop up. The thresholds on acceptable response time were mainly determined from subjects' performance in the practice section and thus varied among subjects. In the first two sessions, the lower threshold was typically above 3000ms, while in the last two sessions it was fixed at 1500ms. The latter was chosen to guarantee enough time for participants to read the attribute values. There was also an upper limit in the last two sessions, which was at least 3000 ms. Finally, one filler question with a dominated option, i.e., a smaller reward with a longer delay, was presented after each set of 25 formal questions. If subjects chose the dominated option in a filler question, a warning sign would pop up, which asked for more attention and at the same time provided subjects with a short break if they wanted.

Model specification and fitting method

A total of nine distinct models of intertemporal choice will be explored in this article. Eight of them are diffusion models and the remaining one is a random utility model. Six of the diffusion models utilize the general framework of DFT, which assumes a dynamic and attribute-wise approach. The other two diffusion models are alternative-wise and have the hyperbolic discount function as their core. As mentioned in Section , we need to determine the values of five parameters (d, σ, θ, z and Ter) when modeling a specific choice behavior with DFT. When intertemporal choice is of concern, it means that we shall relate these parameters to either the intertemporal options presented to participants or the actual data on choice response and response time. According to DFT, the evidence accumulation process is a stochastic process and the amount of information sampled at a specific time is a random variable due to attentional shift. In the current applications of DFT, we assume that at a given time, people attend to either the money amounts or the delay amounts and take the difference between options as the evidence for or against each option. We can interpret the evidence as affective impact pushing people towards one option or the other. The evidence will then be accumulated until a certain threshold is reached. When either direct or relative differences are considered in a model, we need a free parameter, w, to represent the attention weight to money amount (the attention weight to delay amount will be $1 - w$ by definition). When both direct and relative differences are considered, it is assumed that subjects will sample one of the four possible differences (direct/relative by money/delay) at a specific time, and thus we need three attention weight parameters which are free to vary. Furthermore, utility functions are necessary for transforming the objective values of money and delay into their subjective utilities before differences are sampled. This will introduce additional parameters into our models. Both attention weight and utility parameters, together with attribute values in a specific question, will be used to determine the values of d and σ. The remaining three parameters, θ, z, and Ter will be treated as free parameters and estimated from the data. To sum up, in our models of intertemporal choice based on DFT, parameters d and σwill be replaced by attention weight and utility parameters and parameters θ, z, and Ter will have the same meaning as usual.

Since the building blocks of the six diffusion models based on DFT are the direct and/or relative differences in the two attributes involved in intertemporal choice, i.e., money and delay, they will be referred to as weighted additive difference models from now on. The first and simplest model involves only the direct differences and uses identical utility functions on both attributes. As a result, there are only four parameters (w, θ, z, and Ter) in this model, and it will serve as a baseline for subsequent model comparison. The second model again only involves the direct differences but it uses power utility functions, leading to a sum of six parameters ($w, \alpha, \beta, \theta, z$, and Ter). The third model is identical to the second one except for an additional scaling parameter in the utility function of time delay. Consequently, the utility function for time delay in this model is $U(x) = cx^{\beta}$rather than simply $U(x) = x^{\beta}$. The fourth model considers only the relative differences in money and delay amounts, resulting in a sum of six parameters ($w, \gamma, \delta, \theta, z$, and Ter). This model generalizes the concept of proportional difference in the PD model and incorporates it into the general framework of DFT. Therefore, we can examine the usefulness of relative differences in intertemporal choice tasks by comparing the performance of this model to that of others. Finally, the last two models are the most comprehensive ones which involve both direct and relative differences. As a result, more parameters are necessary in these model. Specifically, the fifth model involves a total of 10 parameters ($w_{1,} w_2, w_3, \alpha, \beta, \gamma, \delta, \theta, z$, and Ter), while the sixth model contains an additional scaling parameter on time delay when direct difference is of concern.

Two alternative-wise diffusion models are explored as well so that their performance can be compared to that of attribute-wise diffusion models. In this case, we assume that drift parameter is determined by the difference in discounted utility and at the same time, retain an evidence accumulation perspective as in other diffusion models. Specifically, in both alternative-wise models, the traditional hyperbolic discount function is utilized to calculate the discounted utility for both payoffs in a specific question and the corresponding difference is assigned to the drift parameter, d. The hyperbolic model is chosen as the foundation of these alternative-wise models because it appears to be the most popular deterministic model so far, at least among psychologists.The diffusion parameter, σ, was figured out differently in these two models. In one of the models, we treated it as a free parameter, while in the other model, we set σ^2 proportional to the sum of delay amounts. For the former, we assume that the amount of uncertainty associated with each discounted utility is the same across payoffs, while for the latter, we suppose that the longer a payoff is delayed, the more uncertain its discounted utility will be. We also assume that the utilities of the two payoffs in a question vary independently in both models. Conventionally, when the hyperbolic discount function is used, no utility function is applied to either the money amount or the delay amount. To make the models a little more flexible and similar to other diffusion models based on DFT, we actually used a power utility function for the money amount. No utility function was applied to delay amount so that the hyperbolic discount function could retain its original form. All the other elements of diffusion models were implemented in the same way as before.

In order to compare dynamic versus static models on intertemporal choice, a random utility model extended from the hyperbolic model is also explored here. The extension is necessary because previous models on intertemporal choice are all deterministic and thus can neither account for its probabilistic nature nor make predictions on choice probabilities and response time distributions. There are three parameters involved in the model, and it will be used to predict choice probabilities so that its performance can be compared to those of the models based on DFT. The first parameter is the discounting parameter, k, as in the traditional discount function. The second parameter is a utility parameter, α, which represents the exponent of the power utility function for money. These two parameters will be used to calculate the discounted utility of an intertemporal option. The last parameter, σ, is a measure of the variability of choice response as in Equation 4. With all three parameters, we are able to calculate the choice probability of a specific option in an intertemporal choice question.

For all the aforementioned models, maximum likelihood estimation were used to estimate the relevant parameters and BIC index was calculated as an index for model selection. Since the models tested in this article differ in number of parameters, we need to take into account the issue of model complexity when comparing models. The BIC resolves this problem by introducing a penalty term for the number of parameters in a model. Consequently, a lower BIC value suggests a better balance between goodness-of-fit and model complexity and thus a more desirable model[36]. The models were fitted to individual data and the SIMPLEX algorithm was employed (using the fminsearch function in Matlab) to find the maximum likelihood estimates of the parameters for each subject. See Appendix B for more details on the model fitting procedure.

Results

Choice patterns for the questions related to various intertemporal effects

To demonstrate the probabilistic nature of intertemporal choice and examine whether various intertemporal effects actually show up in a probabilistic way, we first investigated the choice patterns for different types of questions. Because each pair of options was only presented once, it was impossible to estimate the choice probabilities for a single pair. However, we can combine different but similar questions together to obtain an approximate measure on choice probabilities. Consequently, questions in each session were divided into 15 subgroups, each of which included 30 questions associated with a specific intertemporal effect. The questions in each subgroup were similar to one another in attribute values and thus could be viewed as practically the same. The choice proportion of the LL options in each subgroup was then computed as an estimate of related average choice probability. The data from subsequent sessions were analyzed separately because the amounts of time pressure might differ. The upper left panel in Figure 2 shows a line graph illustrating the choice patterns of a typical subject for the questions in the first session. Three lines are depicted in the graph, each relating to a specific intertemporal effect. It is readily seen that in general choice proportions do not change abruptly as suggested by a deterministic perspective. For example, the line with circle markers is associated with the delay amount effect and it shows that when the delays were increased proportionally, the choice proportions of the LL options declined gradually from 0.9 to 0.1. Similarly, the line with square markers corresponds to the magnitude effect and it indicates that, when the reward amounts were increased proportionally, the choice proportions of the LL options rose progressively from 0 to 0.87. However, the line associated with the common difference effect (with diamond markers) does not have a clear positive slope as suggested by that effect. The choice patterns of other subjects were qualitatively the same, and results from other sessions were similar to those in the first one. For the questions regarding the delay amount effect, 30 out of the 40 (10 subjects × 4 sessions) sequences of choice proportions are inconsistent with the prediction of a deterministic interpretation of the effect. Likewise, for the questions regarding the magnitude effect, 27 out of the 40 sequences demonstrate a violation of the deterministic assumption on intertemporal choice. The other three panels in Figure 2 illustrate the average results across subjects in different sessions and for different effects. It is clear that the delay amount effect and magnitude effect were present in each session but not the common difference effect.

Three 5(subgroups)×4(sessions) within-subjects factorial ANOVAs were conducted to further test the intertemporal effects, and the graphs of choice proportions for the LL options are shown in Figure 2. For the questions concerning the delay amount effect, there is a significant difference in actual choice proportion among subsequent subgroups of similar questions (F = 11.28, p <.01, partial η^2=.556), while the difference among sessions is approaching significance (F = 3.441, p > .08). Furthermore, there is a significant linear trend in choice proportion (F = 15.10, p < .01, partial η^2=.627), which is consistent with a probabilistic interpretation of the delay amount effect. Similar results also occur for the questions concerning the magnitude effect. Specifically, there is a significant difference in choice proportion among different subgroups of questions with similar attribute values (F = 20.09, p < .001, partial η^2 = .691), and the difference among sessions is also significant (F = 7.77, p <.01, partial η^2 = .463). Besides, there is a linear trend in choice proportion among successive subgroups (F = 20.07, p < .01, partial η^2 = .69), following the probabilistic interpretation of the magnitude effect. Finally, there is not a significant difference in choice proportion among different subgroups of questions concerning the common difference effect, nor is a significant effect of session. To sum up, choice proportions typically change in a progressive rather than an abrupt manner as required by a deterministic perspective. Besides, the delay amount effect and magnitude effect were demonstrated in a probabilistic rather than deterministic way.

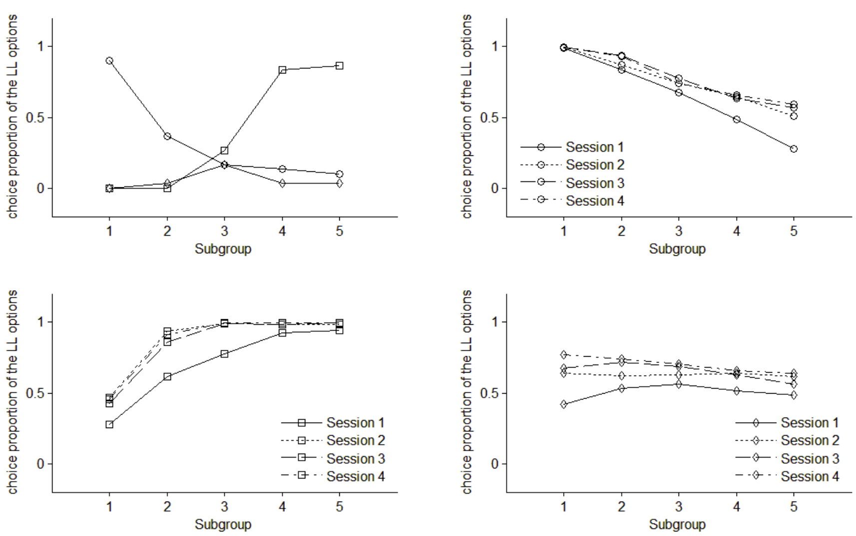

Figure 2. Choice proportions of the LL options for various intertemporal effects and sessions in Study 1. In these graphs, circle markers represent subgroups associated with the delay amount effect; square markers represent subgroups for the magnitude effct; and diamond markers represent subgroups for the common difference effect. Upper left panel: a graph for a typical subject in Session 1. Upper right panel: a graph demonstrating the delay amount effect. Bottom left panel: a graph demonstrating the magnitude effect. Bottom right panel: a graph for the common difference effect, which was absent in Study 1.

Relationships between choice proportions and response times

One critical prediction of diffusion models is that extreme choice probabilities are associated with short response times. Empirically, this entails an inverted U-shaped relationship between the choice proportions of LL options and average response times within subgroups of similar questions. Since a variety of diffusion models will be explored in this article, it is important to first test this prediction to provide evidence for using diffusion processes in these models. To show the existence of the inverted U-shaped relationship, actual choice proportion of the LL options and average response time were calculated for each subgroup. In this way, each subject contributed 15 data points (5 for each intertemporal effect), and the data points across all subjects were then categorized into five groups in terms of actual choice proportion. Specifically, the first group contained data points with a choice proportion between 0 and 0.2, the second contained data points with a choice proportions between 0.2 and 0.4, and so on. Table 1 shows the mean response time within each group. It is clear that for extreme choice proportions (i.e., below 0.2 or above 0.8), the related average response times tended to be shorter, while the average response times associated with moderate choice proportions (i.e., between 0.2 and 0.8) were relatively longer. The mean response time associated with choice proportions below 0.2 or above 0.8 was 3.54s, while the mean response time related to moderate choice proportions was 4.48s. The difference was statistically significant (t = - 6.84, p<.01).

Table 1 Relationship between actual choice proportions of the LL options and average response times within subgroups of similar questions in Study 1.

Actual choice proportion of the LL options	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1
Average response time	3.88s	4.46s	4.72s	4.36s	3.45s

Besides the macrolevel relationship between choice proportions and response times examined above, diffusion models also predict that, for each choice question, the less likely choice response takes more time on average than the other more likely one. This constitutes another critical test on the validity of diffusion models for intertemporal choice. Because each question was only presented once in the current study, it is impossible to estimate its choice probability. Consequently, we categorized questions into subgroups as before and calculated the actual choice proportions and mean response times of the SS and LL options within each subgroup. The resultant choice proportions and mean response times could be viewed as an approximation of the choice probabilities and mean response times of the SS and LL options for each single question in a subgroup if they had been asked repeatedly. The mean response times were then divided into two groups in terms of the popularity of related choice responses. Specifically, one group contained mean responses times for options more likely to be chosen in a subgroup, and the other contained mean response times for less likely options. It turned out that there was a significant difference between these two groups ($M_1 = 4532.7ms, M_2 = 5336.5ms, t = -5.35, p < .001$) and the direction of difference followed the prediction of diffusion models. In other words, the microlevel prediction of diffusion models on the relationship between choice proportions and response times also hold for the current empirical data on intertemporal choice.

Results of model fitting and comparison

Since empirical data suggest that, just like risky choice, intertemporal choice is also probabilistic in nature, we need to develop probabilistic models to account for this property. Eight diffusion models and one random utility model are explored here. All of them are probabilistic models but only those diffusion models are dynamic and thus able to predict response time distributions. Consequently, we will first fit the models to the choice response data so that the random utility model can be compared to the diffusion models. Table 2 shows the results of model fitting in terms of average BIC values and counts of lowest BIC values for each model and session. In all sessions, the performance of Models 3 is the best in terms of average BIC value while Models 2, 5, and 6 follow. On the other hand, Model 4 is always associated with the highest average BIC value and thus least desirable. The two alternative-wise diffusion models based on the hyperbolic discount function, i.e., Models 7 and 8, are always inferior to attribute-wise diffusion models with power utility functions (i.e., Models 2, 3, 5, and 6), but superior to the random utility model. For all the sessions except Session 3, the simplest diffusion model, i.e., Model 1, fits the data better than the random utility model based on the hyperbolic discount function (Model 9) in terms of average BIC value. Overall, Model 3 has the lowest average BIC value and the highest count of lowest BIC values across subjects and session, while Model 4 performs the worst in terms of both average BIC value and count of lowest BIC values. To sum up, when choice response is of concern, Model 3 is the best model and in general diffusion models perform better than the random utility model. See Appendix B for the details of the fitting procedure.

Table 2 Results of fitting probabilistic models to the empirical data on choice response from Study 1

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Session 1	311.4(2)	289.2(0)	279.9(3)	444.7(0)	286.1(3)	288.9(0)	304.3(1)	308.4(0)	337.1(1)
Session 2	239.2(0)	221.0(1)	220.1(2)	364.7(0)	228.1(1)	231.4(0)	242.1(2)	246.3(1)	256.4(3)
Session 3	232.8(2)	197.3(3)	175.5(4)	353.2(0)	192.4(0)	195.1(0)	208.3(0)	212.3(0)	226.3(1)
Session 4	206.9(2)	190.2(2)	183.0(3)	343.6(0)	196.4(0)	203.1(0)	209.7(2)	216.2(0)	226.4(1)
Overall	247.6(6)	224.4(6)	214.6(12)	376.6(0)	225.7(4)	229.6(0)	241.1(5)	245.8(1)	261.5(6)

Note. Model 1: diffusion model involving direct differences and identical utility functions; Model 2: diffusion model involving direct differences and power utility functions without a scaling parameter on time; Model 3: diffusion model involving direct differences and power utility functions with a scaling parameter on time; Model 4: diffusion model involving relative differences and power utility functions; Model 5: diffusion model involving both direct and relative differences and power utility functions without a scaling parameter on time; Model 6: diffusion model involving both direct and relative differences and power utility functions with a scaling parameter on time when direct difference is of concern; Model 7: diffusion model based on the hyperbolic discount function with σ as a free parameter; Model 8: diffusion model based on the hyperbolic discount function with σ proportional to the sum of delay amounts; Model 9: random utility model based on the hyperbolic discount function, which is static but probabilistic. Average BIC value is shown for each model and session, and the counts of lowest BIC values are shown in parentheses.

To further compare the diffusion models among themselves and explore their capability of fitting response times, Models 1 - 8 were fitted to choice responses and response times simultaneously. Specifically, the defective probability density of making a specific response within a certain amount of time was utilized in the model fitting procedure to obtain maximum likelihood estimates of parameters and calculate BIC values accordingly. Since both choice responses and response times were taken into account, more information from the data was exploited to distinguish among the competing diffusion models. Table 3 lists the results of model fitting. Again, Model 3 performs the best in describing the empirical data when average BIC value is used as criterion. The performances of Models 2, 5, and 6 fit the data reasonably well, while Model 4 does not. The performances of Models 7 and 8 are only better than that of Model 4. When count of lowest BIC is of concern, Model 2 is comparable to Model 3 and performs better than other diffusion models. All in all, the comparison results are generally the same as before when the models were only fitted to the data on choice responses.

Table 3 Results of fitting diffusion models to both choice response and response time data in Study 1

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Session 1	2014.6(0)	1955.8(3)	1925.0(3)	2092.8(0)	1942.8(0)	1935.5(2)	1986.5(0)	1984.6(2)
Session 2	1891.7(0)	1791.0(3)	1769.7(5)	1896.3(0)	1792.8(0)	1788.6(0)	1806.1(1)	1807.6(1)
Session 3	1088.4(0)	985.0(4)	967.1(3)	1108.8(0)	991.8(0)	995.3(0)	1005.8(1)	1006.9 (2)
Session 4	1011.5(0)	928.1(4)	915.0(3)	1042.3(0)	941.1(0)	946.8(0)	959.6(2)	961.0 (1)
Overall	1501.6(0)	1415.0(14)	1394.2(14)	1535.0(0)	1417.1(0)	1416.6(2)	1439.5(4)	1440.3(6)

Note. Model 1: diffusion model involving direct differences and identical utility functions; Model 2: diffusion model involving direct differences and power utility functions without a scaling parameter on time; Model 3: diffusion model involving direct differences and power utility functions with a scaling parameter on time; Model 4: diffusion model involving relative differences and power utility functions; Model 5: diffusion model involving both direct and relative differences and power utility functions without a scaling parameter on time; Model 6: diffusion model involving both direct and relative differences and power utility functions with a scaling parameter on time when direct difference is of concern; Model 7: diffusion model based on the hyperbolic discount function with σ as a free parameter; Model 8: diffusion model based on the hyperbolic discount function with σ proportional to the sum of delay amounts; Model 9: random utility model based on the hyperbolic discount function, which is static but probabilistic. Average BIC value is shown for each model and session, and the counts of lowest BIC values are shown in parentheses.

Model predictions

Another way to evaluate the performance of a specific model in fitting empirical data is to make predictions on choice responses and response times using the estimated values of model parameters and then compare the predictions with the actual data. Since Model 3 performs the best, followed by Models 2, 5, and 6, we first compared actual choice proportions with these models' predictions when they were fitted to choice response data. Figure 3 shows the scatter plots of the average choice probabilities predicted by these models and the actual choice proportions. Each point in the scatterplots is associated with a subgroup of similar questions concerning a specific intertemporal effect as before. Clearly, there is a strong correlation between the predicted average choice probabilities and the actual choice proportions for each of the models (for Model 2, r = .96, p<.001; for Model 3, r = .98, p<.001; for Model 5, r = .98, p<.001; for Model 6, r = .99, p<.001). In other words, we can use these models to make reasonably good predictions on the actual choice proportions. Similar results showed up when we used the parameter values estimated from fitting these models to both the choice response data and response time data. Figure 4 shows the scatterplots of these models' predictions on response times and the actual response times. For each question, we first figured out the distribution of response time given the actual choice response and then used the expected value of that distribution as a point estimation. After that, the actual mean response time and predicted mean response time for each subgroup of questions were calculated, resulting in a specific data point in the scatterplots. The general pattern remained the same when other measures of central tendency were explored or the unconditional mean response time was utilized. It is readily seen that the predictions of these models match the actual data quite well (for Model 2, r =.78, p <.001 ; for Model 4, r =.81 , p <.001)

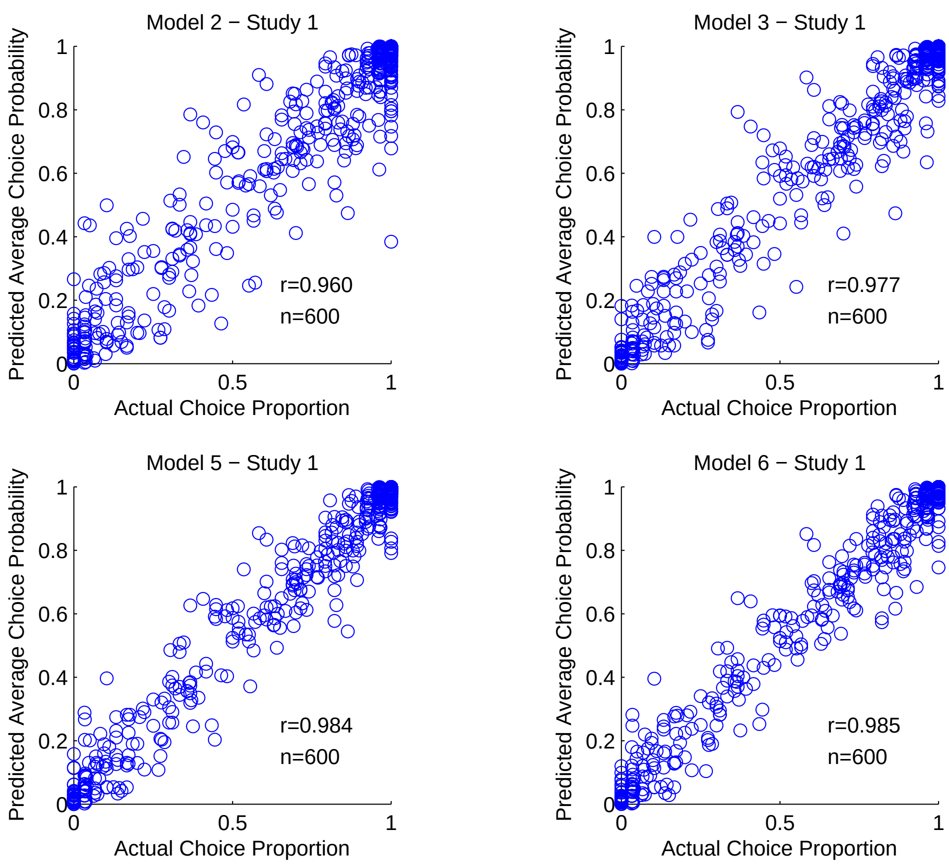

Figure 3. Scatter plots of average predicted choice probabilities and the actual choice proportions for Models 2, 3, 5, and 6 in Study 1. Each point in the scatterplots is associated with a subgroup of similar questions.

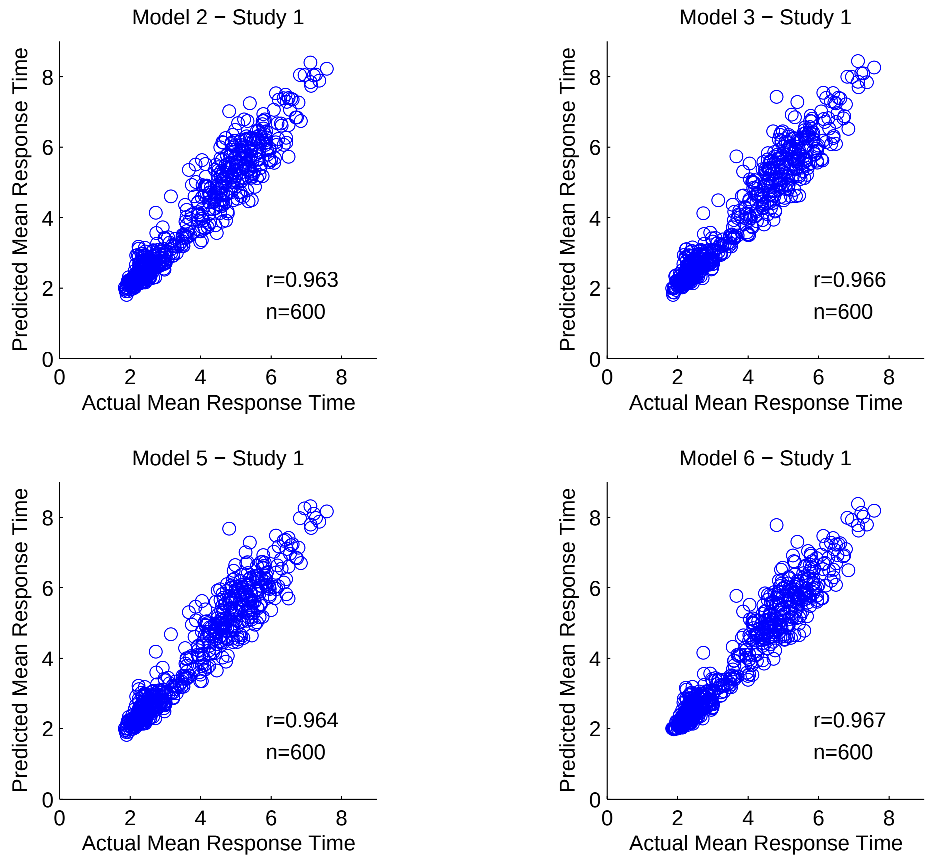

Figure 4. Scatter plots of predicted mean response times and actual mean response times for Models 2, 3, 5, and 6 in Study 1. Each point in the scatterplots is associated with a subgroup of similar questions. Conditional mean response time given the actual choice response is used as a point estimation for each question.

Finally, we can test the validity of the diffusion models by examining their predictions on the impact of experimental manipulation on actual choice proportions. It has been shown that, by systematically changing attribute values, the delay amount effect and magnitude effect were revealed in the empirical data but not the common difference effect (Fig. 2). If the diffusion models actually capture the essense of the underlying processes leading to the explicit responses, their predictions should replicate the empirical choice patterns. Following are line graphs illustrating the predictions of Model 3 on average choice probabilities of the LL options for different effects and sessions. It is readily seen that the general pattern in the predicted results is almost the same as that in the empirical results. The line graphs for Models 2, 5 and 6 are similar to those for Model 3.

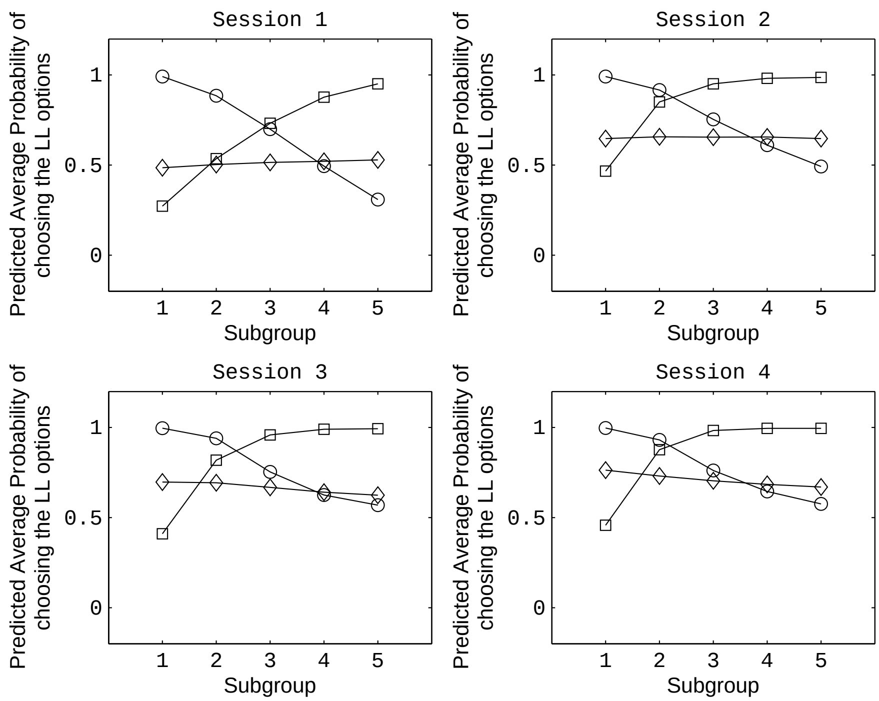

Figure 5. The predictions of Model 3 on average choice probabilities of the LL options for various intertemporal effects and sessions in Study 1. In these graphs, circle markers represent subgroups associated with the delay amount effect; square markers represent subgroups for the magnitude effct; and diamond markers represent subgroups for the common difference effect. In each session, Model 3 predicted the delay amount effect and magnitude effect but not the common difference effect.

Discussion

Probabilistic nature of intertemporal choice

The data from Study 1 suggest that intertemporal choice is also probabilistic in nature just like risky choice. Specifically, subjects' responses to the questions devised to test various intertemporal effects reveal a violation of the deterministic assumption on this phenomenon. More importantly, choice pattern for the questions concerning the delay amount effect appears to be beyond the means of any existing account of this important topic and thus constitutes a severe challenge to the deterministic assumption. According to the delay amount effect, when the time delays of both options are multiplied by a common constant, people's preference should shift towards the SS option. That is to say, when the ratio of delay amounts is constant, the longer the delays are, the less preferable the LL option will be. If we assume further a deterministic perspective on intertemporal choice, it can prove that there exists a single cutoff point on the shorter delay (or equivalently on the longer delay) where people are just indifferent between the two options. Moreover, such a perspective also stipulates that when the shorter delay amount is below the cutoff point, people will choose the LL option for sure, and vice versa. In other words, the choice probability of the LL options should change abruptly from one to zero across the cutoff point. This is actually a typical pattern predicted by a deterministic view. Furthermore, it can be shown that such a pattern is consistent with both the exponential discount function and the hyperbolic discount function as long as a deterministic approach is assumed, and it remains true even when the magnitude effect is taken into account (See Appendix C for mathematical proofs of these properties.) When similar questions are combined into subgroups as in Study 1, the deterministic approach entails that only one subgroup could have an actual choice proportion that is between 0 and 1, and the actual proportions should be either monotonic increasing or monotonic decreasing (although not necessarily strictly monotonic). However, a majority of subjects in Study 1 demonstrated a gradual rather than abrupt shift in choice proportion for the questions related to the delay amount effect. This forms a strong piece of evidence against the deterministic assumption of intertemporal choice.

Model comparisons

After showing the inherent randomness in intertemporal choice, a number of probabilistic models are fitted to the data so that more sophisticated models can be developed and tested. It turned out that the diffusion models are generally better than the random utility model in describing the empirical data collected in Study 1. It should not be surprising since the relationships between choice proportions and response times are in favor of the dynamic structure involved in diffusion models. On the other hand, it is worth noting that only one quite simple random utility model, i.e., a Thurstone Case V model expanded from the hyperbolic discount function, was explored. It is possible that more complicated random utility models may perform better when choice response is of concern. Diffusion models, however, are essentially superior to random utility models when we want to account for response times as well. In this scenario, it is more appropriate to explore models that are embedded with a dynamic structure.

Another important finding is that the diffusion models considering only direct differences in reward amount and delay amount perform much better than that involving only relative diffferences. Actually the latter is even worse than the random utility model which does not have a dynamic structure. This implies that we cannot use only relative or proportional differences to explain human intertemporal preference. Since the concept of relative or proportional difference originates from the PD model, it suggests that the simplest version of the PD model (i.e., fixed σ and δ across choice questions) will perform poorly when fitted to the current data. When both direct and relative differences are included, the resultant models are still generally inferior to those only involving direct differences, especially the model with a scaling parameter on time (i.e., Model 3).

Besides the six attribute-wise diffusion models based on direct and/or relative differences, two alternative-wise diffusion models expanded from the hyperbolic discount function were also examined. In this way, we put the traditional hyperbolic discounting model and the new weighted additive difference models on the same structural level. It turned out that the attribute-wise models are still better than the alternative-wise ones when both of them assume a dynamic structure. On the other hand, Models 7 and 8 perform better than Model 9, the static model, which suggest that the involvement of a dynamic structure does help improve the performance of relevant models. To sum up, the dynamic models perform better than the static one when the hyperbolic discount function is at the core; attribute-wise models seem to be better than alternative-wise models when fitting empirical data; and relative or proportional difference seems to be an unnecessary component for a good model on intertemporal choice.

Drawbacks of Study 1

Although Study 1 reveals the probabilistic nature of intertemporal choice and provides a platform on which various models can be explored, there are several drawbacks in the study which may weaken the validity of the results. First of all, the sample size of the current study, i.e., 10, was relatively small compared to other studies on intertemporal choice. Consequently, we might need a larger sample so that our conclusions can be built upon a more solid foundation. Second, the subjects needed to finish four sessions in total which were quite similar in structure and stimuli. Although successive sessions were administrated at least one week apart, it was possible that subjects had a subtle memory of what happened in previous sessions or gradually formulated fixed strategies in later sessions. Both possibilities constituted a violation of the assumption of independence between responses, which is necessary for the fitting procedure to estimate model parameters. In fact, the responses of most subjects became more extreme in later sessions. In other words, their responses turned more predictable when they got more experience with the stimuli. This may well be the reason why the BIC values decrease in later sessions as listed in Tables 1 and 2. Consequently, it may be helpful to reduce the length of the study so that subject only needs to go through one instead of four sessions. Finally, the time pressure involved in the first two sessions seem to be inappropriate in the sense that most subjects needed to postpone their responses sometimes to avoid the warning sign. As a result, in Study 2, we changed the lower time limit accordingly to avoid this undesirable side-effect.

Study 2

Given the drawbacks of Study 1, a new experiment was conducted to obtain more confidence in previous results. Specifically, more subjects were recruited in Study 2, which involved only one session that lasted as long as a single session in Study 1. Besides, the lower limit on response time was reduced so that it was less likely for subjects to postpone their responses. The main purpose of Study 2 was to replicate the results of Study 1 with a better experimental design.

Method

Materials

As in Study 1, an adjustment procedure was first employed to generate three approximately indifferent pairs of intertemporal options for every subject, with one pair for each intertemporal effect. In each case, three of the four attribute values were fixed and the remaining one varied from trial to trial according to subjects' responses to the previous question. For the questions related to the delay amount effect, the shorter delay was fixed at 20 days; the longer delay was fixed at 40 days; and the larger reward amount was fixed at 35 dollars. The remaining attribute value, i.e., the smaller reward amount, was initially set at 20 dollars and changed on the basis of subjects' responses. Similarly, for the questions related to the common difference effect, the shorter delay was fixed at 20 days; the longer delay was fixed at 50 days; and the larger reward amount was fixed at 32 dollars. The smaller reward amount started from 16 dollars and again changed according to subjects' previous response. Finally, for the questions concerning the magnitude effect, the smaller reward, the larger reward, and the shorter delay were fixed at 20 dollars, 40 dollars, and 12 days respectively. The longer delay was initially set at 30 days and changed in the same manner as for the other questions.

For each effect, 160 formal questions were then created based on the indifferent pair generated from the adjustment procedure. Fewer questions were required here because of a reduction in the length of the study. Specifically, in the formal questions concerning the delay amount effect, the longer delays were always two times as long as the shorter delays, which ranged between 1 day and 40 days. Note that the ratio of delays was different from that in Study 1. For each pair of delays, associated reward amounts were then jiggled from the ones in the indifferent pair to generate 4 questions that were a little different from one another but practically the same. Again the purpose of this manipulation was to avoid presenting subjects with the same choice questions multiple times to reduce possible impact of memory. Furthermore, the tens digits of the smaller rewards in these similar questions were always the same, and the same applied to the longer rewards. This was intended to make them look more similar to one another. Overall, 160 (40 by 4) questions were generated based on the indifferent pair in this way. The same method was used to generate another 320 questions associated with the common difference effect and magnitude effect.

Participants

Forty-six participants (29 females and 17 males) with an average age of 23 were recruited at a national research university via advertisement on notice boards. Nine of them generated abnormal indifferent pair(s) of intertemporal options (i.e., with one option dominating the other) and thus their data were analyzed separately. Among the remaining 37 participants, 25 were women and 12 were men, and the average age was about 21. All the participants were reimbursed for doing the study. Specifically, for each participant, one intertemporal choice question was randomly picked from his/her question set and the person was paid one-fourth of the money he/she chose in that question. Besides, there was a baseline payment of 4 dollars in addition to the payment contingent on the randomly selected question. The date of payment was also determined by the amount of time delay participants chose in the randomly selected question. The average payment was about 11 dollars.

Procedure

As was mentioned above, Study 2 only contained one session. As in Study 1, all the instructions and questions were presented on a computer screen and subjects used a mouse to make responses. This was carried out by a Matlab program together with the Cogent toolbox to record both choice responses and response times. In the current study, subjects were presented with 480 intertemporal questions and required to indicate their preferences. To lessen the fatigue effect, two major breaks and contingent short breaks were inserted into the study. Different types of questions were presented in a random order so that successive questions appeared to be irrelevant to one another. The adjustment procedure to generate indifferent pairs was run before all formal questions were created and presented. Furthermore, a practice section was provided before subject generated indifferent pairs. Subjects were instructed throughtout the whole study to take into account all pieces of information involved in each question and think it over before making a choice. They were also reminded of the payment plan that one question would be randomly selected and their payment would be contingent on the specific choice they made in that question. In this study, subjects were required to make careful responses. If they responded too fast, a warning sign would pop up. The lower threshold on response time was set at 1500ms, which was more lenient than that in Study 1, where the lower threshold was determined by subjects' responses in the practice section and was typically above 3000ms. Finally, one filler question with a dominated option was presented after each set of 40 formal questions. If subjects chose the dominated option in a filler question, a warning sign would pop up, which asked for more attention and at the same time provided subjects with a short break if they wanted.

Results

The results reported here are mainly from the 37 subjects who generated intertemporal questions without a dominating option. At the end of this section, the results from those subjects who answered abnormal questions will be briefly discussed.

Choice patterns for the questions related to various intertemporal effects

As before, we first combined questions into subgroups and calculated the actual choice proportion for each subgroup in order to demonstrate the probabilistic nature of intertemporal choice. In Study 2, each subgroup contained 32 questions associated with a specific intertemporal effect. The left panel in Figure 6 shows a line graph illustrating the choice patterns of a typical subject in the current study. As in Figure 2, three lines are depicted in the graph, each relating to a specific intertemporal effect. As in Study 1, the choice proportions do not change abruptly as suggested by a deterministic perspective. For example, the line with circle markers is associated with the delay amount effect and it suggests that when the delays were increased proportionally, the actual choice proportion of the LL option decreased gradually from 1 to about 0.3. Similarly, the pattern for the questions concerning the magnitude effect is generally the same as in Study 1. For the questions regarding the delay amount effect, all 37 sequences of choice proportions are inconsistent with the prediction of a deterministic interpretation of the effect. Likewise, for the questions regarding the magnitude effect, 36 out of the 37 sequences of choice proportions suggest a violation of the deterministic assumption of intertemporal choice. The right panel in Figure 6 also shows the average situation across subjects, which is quite similar to that for the typical subject. The only difference between Studies 1 and 2 lies in the common difference effect. It can be seen from the right panel in Figure 6 that the common difference effect seems to be revealed in Study 2.

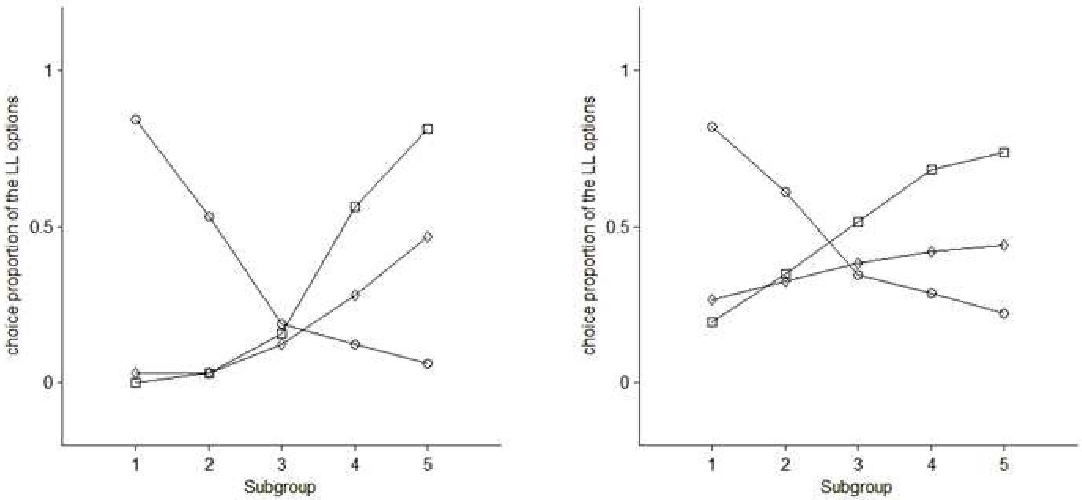

Figure 6. Choice proportions of the LL options for various intertemporal effects in Study 2. In these graphs, circle markers represent subgroups associated with the delay amount effect; square markers represent subgroups associated with the magnitude effct; and diamond markers represent subgroups associated with the common difference effect. Left panel: a line graph for a typical subject. Right panel: a line graph demonstrating the average results across subjects for the three intertemporal effects.

Three within-subjects ANOVAs were conducted to further test the intertemporal effects respectively. For the questions concerning the delay amount effect, there is a significant difference in choice proportion among different subgroups of similar questions ($F = 67.99$, $p <.001$, partial η^2=.654). Furthermore, there is a significant linear trend in choice proportion ($F = 103.3$, $p < .001$, partial η^2=.742), which is consistent with the probabilistic prediction of the delay amount effect. Similar results show up for the questions concerning the magnitude effect: the difference in choice proportion among different subgroups is significant ($F = 49.03$, $p < .001$, partial $\eta^2 = .577$), as well as a linear trend in choice proportion among successive subgroups ($F = 79.82$, $p < .001$, partial $\eta^2 = .689$). Finally, there is a significant difference in actual choice proportion among different subgroups of questions concerning the common difference effect ($F = 4.75$, $p < .05$, partial η^2=.117), and the corresponding linear trend is also significant ($F = 6.09$, $p <.05$, partial η^2=.145). To sum up, choice proportions typically changed in a progressive rather than abrupt manner as required by a deterministic approach. In addition, on average, all three intertemporal effects were revealed in a probabilistic manner.

Relationships between choice proportion and response time

As in Study 1, actual choice proportion of the LL options and average response time were calculated for each subgroup to show the inverted U-shaped relationship between choice proportions and response times within subgroups of similar questions. In the current study, however, each subgroup contained 32 rather than 30 questions. Again, the resultant data points were further divided into five groups in terms of actual choice proportion. Table 4 shows the mean response time for each group. As before, for extreme choice proportions (i.e., below 0.2 or above 0.8), the related mean response times tended to be shorter. while the response times associated with moderate choice proportions (i.e., between 0.2 and 0.8) were relatively longer. The mean response time associated with choice proportions below 0.2 or above 0.8 was 3.84s, while the mean response time associated with moderate choice proportions was 4.32s. The difference was statistically significant as in Study 1 ($t = - 4.81$, $p<.01$).

Table 4 Relationship between actual choice proportions of the LL options and mean response times within subgroups of similar questions in Study 2.

Actual choice proportion of the LL options	0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1
Average response time	3.91s	4.35s	4.28s	4.31s	3.73s

The microlevel relationship between choice proportions and response times were also investigated as in Study 1. Again, we first calculated the actual choice proportions and mean response times of the SS and LL options within each subgroup. The mean response times were then divided into two groups in terms of the popularity of related choice responses. Specifically, one group contained mean responses times for options more likely to be chosen in a subgroup, and the other contained mean response times for less likely options. The same significant result showed up when these two groups of mean response times were compared to each other ($M_1 = 4236.4ms, M_2 = 5205.1ms, t = -6.65, p < .001$) and the direction of difference followed the prediction of diffusion models. Therefore, the microlevel prediction of diffusion models on the relationship between choice proportions and response times was established in Study 2 as well.

Results of model fitting and comparison

After reconfirming the probabilistic nature of intertemporal choice with the new data set, we proceeded to fit the probabilistic models to the data. As before, we first fit the models to the choice response data so that the random utility model can be compared to the diffusion models. Table 5 shows the fitting results in terms of average BIC values and counts of lowest BIC values for each model. As in Study 1, the performance of Models 2, 3, 5, and 6 are superior to that of the remaining models, including the random utility model (i.e., Model 9), in terms of average BIC value, and the diffusion model involving only relative differences (i.e., Model 4) performs the worst. When the counts of lowest BIC values are of concern, Model 2 performs the best, followed by Model 1. If we consider both choice response and response time data, Models 2, 3, 5, and 6, again perform the best in terms of average BIC value. Furthermore, Model 2 leads the competition when the count of lowest BIC values is of concern, followed by Models 3 and 5. To sum up, in Study 2, Model 2 performs the best while Model 4 performs most poorly.

Table 5 Results of fitting probabilistic models to the empirical data from Study 2

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Fitting choice response data	385.0(6)	327.6(12)	326.9(1)	466.9(0)	332.2(5)	338.1(1)	365.3(4)	361.9(3)	392.7(5)
Fitting both choice response and response time data	1886.8(3)	1833.2(13)	1831.7(6)	1923.2(2)	1817.2(6)	1822.4(1)	1878.3(5)	1864.0(1)	

Note. Model 1: diffusion model involving direct differences and identical utility functions; Model 2: diffusion model involving direct differences and power utility functions without a scaling parameter on time; Model 3: diffusion model involving direct differences and power utility functions with a scaling parameter on time; Model 4: diffusion model involving relative differences and power utility functions; Model 5: diffusion model involving both direct and relative differences and power utility functions without a scaling parameter on time; Model 6: diffusion model involving both direct and relative differences and power utility functions with a scaling parameter on time when direct difference is of concern; Model 7: diffusion model based on the hyperbolic discount function with σ as a free parameter; Model 8: diffusion model based on the hyperbolic discount function with σ proportional to the sum of delay amounts; Model 9: random utility model based on the hyperbolic discount function, which is static but probabilistic. Average BIC value is shown for each model, and the counts of lowest BIC values are shown in parentheses.

Model predictions

As in Study 1, we first examined the performance of Models 2, 3, 5, and 6 when they were fitted to the choice response data since they performed best in terms of the BIC index. Figure 7 shows the scatter plots of average choice probabilities predicted by these models and the actual choice proportions. Each point in the plots is associated with a subgroup of similar questions. Once again, there is a strong correlation between the average predicted choice probabilities and the actual choice proportions for each of the models (for Model 2, r =.961, p<.001; for Model 3, r = .968, p<.001 ;for Model 5, r =.981, p<.001; for Model 6, r = .982, p<.001). We also calculated the predicted probabilities when these four models were fitted to both choice response data and response time data. Figure 8 shows the scatterplots of model predictions on mean response times and the actual mean response times. For each question, we first figured out the distribution of response time given the actual choice response and then used the expected value of that distribution as a point prediction. Each point in the plots corresponds to a subgroup of similar questions. The general pattern remains the same when other measures of central tendency were explored or the unconditional mean response time was utilized. As in Study 1, the predictions of all these models match the actual data quite well (for Model 2, $r = .961$, $p < .001$; for Model 3, $r = .968$, $p < .001$; for Model 5, $r = .981$, $p < .001$; for Model 6, $r = .982$, $p < .001$).

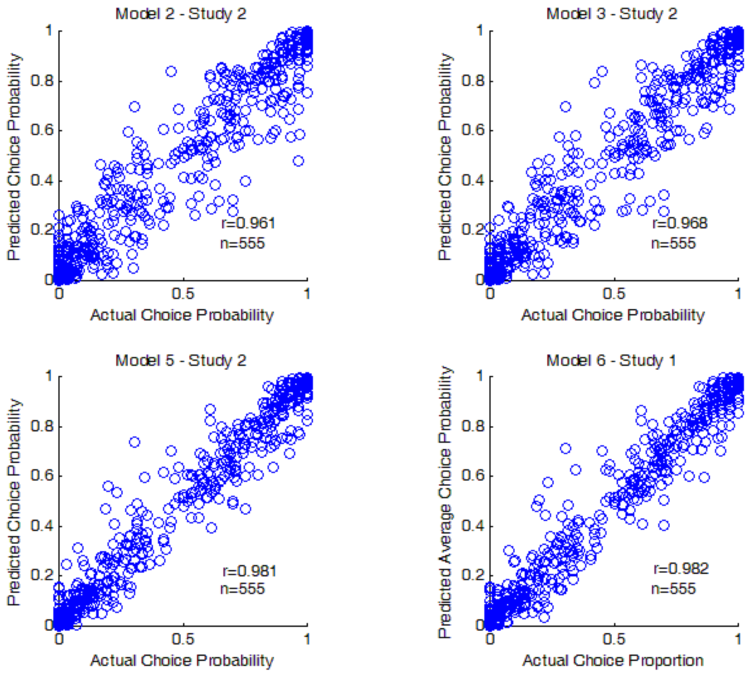

Figure 7. Scatter plots of average predicted choice probabilities and the actual choice proportions for Models 2, 3, 5 and 6 in Study 2. Each point is associated with a subgroup of similar questions.

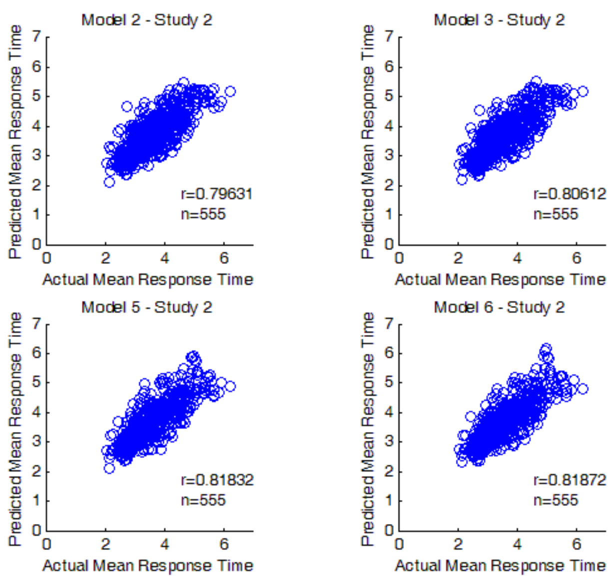

Figure 8. Scatter plots of predicted mean response time and actual mean response time for Models 2, 3, 5, and 6 in Study 2. Each point is associated with a subgroup of similar questions.

Finally, we tested the validity of the diffusion models by examining their predictions on the impact of experimental manipulation on actual choice proportions as in Study 1. All three intertemporal effects were revealed in the current study (Fig. 6). If the diffusion models actually capture the essense of the underlying processes leading to the explicit responses, their predictions should replicate the empirical choice patterns.

Following are line graphs illustrating the predictions of Models 2, 3, 5, and 6 on average choice probabilities of the LL options associated with different effects. It is readily seen that the general pattern in the predicted results is the same as that in the empirical results.

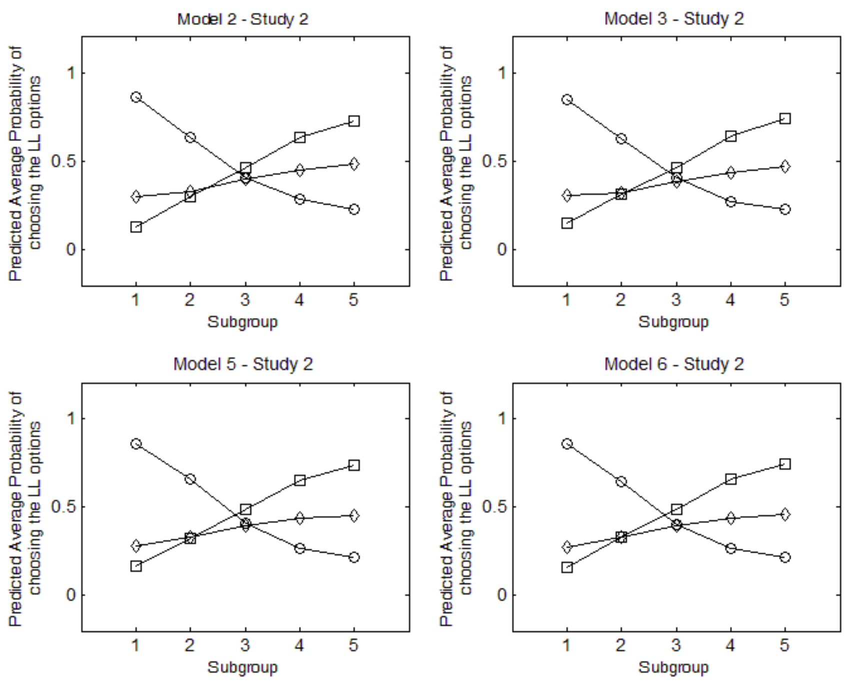

Figure 9. The predictions of Models 2, 3, 5, and 6 on average choice probabilities of the LL options for various intertemporal effects in Study 2. In these graphs, circle markers represent subgroups associated with the delay amount effect; square markers represent subgroups for the magnitude effect; and diamond markers represent subgroups for the common difference effect. All these models predicted the three intertemporal effects.

Results of subjects answering abnormal questions

As was mentioned above, nine subjects generated at least one indifferent pair of intertemporal options that contained dominating and dominated options (e.g., the reward amounts were the same but one option had a shorter delay). In this case, it is conceivable that most of the time subjects would choose the dominating options, leading to quite extreme choice proportions. Moreover, any diffusion model explored so far will predict a choice probability of one for the dominating option. If a subject happened to choose the dominated option, the likelihood of the corresponding data set would be zero and its logarithm would be negative infinity. Actually, seven out of the nine subjects did occasionally choose dominated options. Consequently, in order to avoid breakdown of the model fitting procedure, whenever a dominated option was chosen, a very small probability (e.g., .01) should be assigned to the response instead of 0. With this modification, it is possible to fit the diffusion models and the random utility model to the abnormal data.

It turned out that the data of these nine subjects were generally the same as those of the other subjects. First of all, for the questions concerning the delay amount effect, there was a trend consistent with the probabilistic interpretation of the effect and it was approaching statistical significance ($F = 4.49$, $p = .067$). For the questions concerning the magnitude effect, a linear trend consistent with the probabilistic demonstration of the effect was found ($F = 16.96$, $p < .01$, partial η^2=.679). There was not a significant linear trend associated with the questions concerning the common difference effect. When various models were fitted to the data, Model 2 again performs better than the other models.

Discussion

In general, the results from Study 2 replicated those from Study 1 with a refined experimental design and thus further corroborated the probabilistic nature of intertemporal choice. As in Study 1, both the delay amount effect and magnitude effect were revealed in a probabilistic rather than deterministic manner. Moreover, the advantage of diffusion models based on weighted additive difference, especially those involving only direct differences, was again manifested with the large sample. This suggested that direct differences in reward and delay amounts play a significant role in determining human intertemporal preference, while relative differences may have some unique but small contribution beyond that of direct differences. The performance of alternative-wise models built upon the traditional hyperbolic discount function was inferior in general, especially the specific random utility model which lacks a dynamic structure. The poor performance of the static model was not unexpected since the relationships between choice proportion and response time in Study 2 again favored a dynamic approach.

One important new finding in Study 2 was that the common difference effect also showed up in a probabilistic way, which was obscured in Study 1 probably due to the small sample size. Another possible reason for the weak demonstration of the effect was that the specific range of delays explored in the current two studies might severely abate the effect. Previous research on the common difference effect typically involved delays that differed by at least 1 year when human subjects were required to indicate their preference between different payoffs. In order to make the real payment more credible to the subjects, we intentionally limited the range of delay amounts so that they could be fulfilled within 6 months. Consequently, the common difference effect might be too weak to bring about a statistically significant result. It will be helpful to widen the range of delay amounts in future studies to demonstrate the common difference effect in a probabilistic manner.

General Discussion

Probabilistic nature of intertemporal choice

In the past several decades, research on intertemporal choice has been dominated by the delay discounting paradigm which assumes a deterministic view on human intertemporal preference. According to the paradigm, when people need to choose between two payoffs occuring at different points in time, they first assign a psychological value or subjective utility to each option with a specific rule or discount function and then choose the one with a higher utility. That is to say, given a certain discount function, the utilities are deterministic and thus people's preference between different intertemporal options. As an important initial step towards an understanding of intertemporal choice, this paradigm has born numerous fruits, including various candidates for the discount function and feasible interpretations of the intertemporal effects discussed in this article. The main advantage of this perspective lies in its simplicity which puts more complicated issues aside so that researchers can focus on the most fundamental aspects of the phenomenon.

Although the deterministic perspective has long been popular among psychologists and economists interested in intertemporal choice, it is obviously not the only way to look into this important topic. Considering the large amount of evidence in favor of the probabilistic nature of preferential choice shown in this article, it is very likely that intertemporal choice is essentially probabilistic just like risky choice. This important property of intertemporal choice, however, has long been neglected, explicitly or implicitly. For example, previous studies on the form of discount function usually treated the deviations of actual data points from the fitting line as non-systematic errors and tried to estimate the corresponding parameters by minimizing the sum of sqaured errors. In this way, the emphasis was on the best form of the fitting line while the deviations were regarded as a nuisance component which should be eliminated. To the contrary, the probabilistic nature of intertemporal choice revealed in this article implies that the deviations actually demonstrated this inherent property which deserves the same amout of attention as the fitting line itself. Furthermore, the conclusions of previous studies on the form of discount functions may also be misleading due to their neglect of the probabilistic nature of intertemporal choice. For instance, the advantage of the hyperbolic discount function over the exponential discount function in describing empirical data (i.e., producing a higher R^2) may indeed be the consequence of the robustness of the hyperbolic form against the randomness of human intertemporal preference. All in all, a probabilistic view on intertemporal choice may shed new light on this substantial research area and therefore change our understanding in a fundamental way.

Dynamic versus static models

After showing the probabilistic nature of intertemporal choice, a number of models were explored to find a more comprehensive description of the empirical data. Specifically, eight diffusion models and a random utility model were tested and compared to one another. Although both types of models are probabilistic, the diffusion models assume a dynamic approach while the random utility model is still static just like traditional delay discounting models. The primary advantage of a dynamic model over a static one lies in its explicit description of the emotional and cognitive processes via which a final decision is reached. Such a description makes dynamic models more informative than static ones and, as a result, capable of predicting choice response time as well. Practically, researchers can exploit more empirical data to examine various models and make more convincing conclusions. It is clear that both the theoretical and practical superiority of a dynamic model renders it more desirable as a candidate model on intertemporal choice. The results of model comparison also suggested that the diffusion models are generally better than the random utility model in fitting the empirical choice response data. In addition, dynamic models are always preferable to static ones when response times are of concern.

Attribute-wise versus alternative-wise models

Another difference among the probabilistic models lies in their assumption on the order of information search and integration. The six weighted additive difference models assume an attribute-wise perspective while the other models based on the hyperbolic discount function are alternative-wise just like traditional models. An attribute-wise model suggests that people search and compare information on a certain attribute across options at a time and then integrate the comparison results across different attributes to make a decision. On the contrary, an alternative-wise perspective requires that people evaluate the attractiveness of a specific option at a time and then reach a conclusion by comparing and/or accumulating the results of evaluation. The results of current studies suggested that an attribute-wise approach is more suitable than an alternative-wise approach for describing empirical data on intertemporal choice. Given the success and popularity of the hyperbolic discount function in previous research, however, one may wonder why weighted additive difference models are superior to models built upon the discount function. There are four possible explanations for this result. First, as discussed in Scholton and Read (2010), the discounting approach was unable to account for a number of anomalies which an attribute-wise model can handle. In other words, the hyperbolic discount function may be inherently defective as the core of a model on intertemporal choice. Second, previous research in favor of the hyperbolic discount function only examined the indifferent pairs and thus ignoring a large portion of the information in the dataset. While all the data are taken into account as in the current studies, the hyperbolic model may become less useful. Third, the probabilistic nature of intertemporal choice was never explored in previous studies on the appropriate form of discount function and thus the resultant success under the deterministic framework does not necessarily means a good model. Finally, almost all the questions in the current studies involved two delayed options rather than a combination of immediate and delayed payoffs, but the latter is the typical form of questions used in previous research on the form of discount function. Consequently, an attribute-wise strategy may be more efficient and feasible than an alternative-wise strategy in this case since the latter requires evaluating discounted utility twice. It will be valuable in the future to explore the performance of alternative-wise models when the traditional form of questions are asked.

Comparison among weighted additive difference models

Another important finding from model comparison is that models involving only direct differences are better than models involving relative differences in general no matter whether they were fitted to choice response data only or fiited to both choice response data and response time data simultaneously. Specifically, Models 2 and 3 perform better than Models 5 and 6, and Model 1 performs better than Model 4, which performs most poorly among all the models explored here, including the static random utility model. The success of Models 2 and 3 suggest that direct differences are the essential components people consider when they face an intertemporal choice, and the inferiority of Models involving relative differences, especially that of Model 4, implies that relative difference itself is not sufficient to explain human intertemporal preference. In addition, Model 3 fitted the data in Study 1 a little better than Model 2, while Model 2 appeared more desirable than Model 3 in Study 2. The comparability between Models 2 and 3 in terms of the BIC index suggests that the scaling parameter on time may not be an integral part of a good model. Since Study 2 was based on a refined experimental design, Model 2 should become our choice if we want to have a simple model with reasonably high accuracy.

A general framework for explaining the three intertemporal effects simultaneously

In the result sections, we have shown that Models 2, 3, 5 and 6 are able to replicate the pattern of change in actual choice proportion when attribute values were manipulated as required by the three intertemporal effects. In fact, Models 2, 3, 5 and 6 all provide a general framework for explaining the delay amount effect, common difference effect, and magnitude effect simultaneously. Specifically, direct differences involved in these models set up a basis for accounting for the delay amount effect and magnitude effect, while the assumption that direct differences are sampled after applying utility functions make them capable of explaining the common difference effect. Besides, we can use the relative differences involved in Models 5 and 6 to explain the common difference effect as well.

Mathematically, the delay amount effect and magnitude effect are equivalent if we treat reward amount and delay amount as two independent and exchangeable attributes. This is actually one implicit assumption of the diffusion models. Therefore, when reward amounts or delay amounts of the two options are increased proportionally, the associated direct difference will increase for sure regardless of the specific value the related utility parameter assumes. This will in turn make the LL options (for magnitude effect) or the SS options (for delay amount effect) more attractive. That is to say, it becomes more likely that the relevant decision threshold will be reached first after the attribute values are changed according to the specification of the effects.

On the other hand, Models 2 and 3 offer an explanation for the common difference effect as well when its utility parameter on time delay is smaller than 1. Under this condition, increasing both delays by a common additive constant will reduce the (subjective) direct difference between the delay amounts. This will in turn make the LL option more attractive because its disadvantage due to a longer delay is lessened. Since Models 2 and 3 are nested in Models 5 and 6 respectively, the same mechanism can be invoked by the latter to account for the common difference effect. Additionally, the relative difference involved in Models 5 and 6 provides another approach to explaining the effect, since the relative difference in delay will also decrease if both delays are lengthened by a common additive constant. All in all, Models 2, 3, 5 and 6 are able to account for all the three important effects in intertemporal choice explored in the current paper with a single general framework based on evidence accumulation process. This again demonstrates the power of diffusion models in general and DFT models in particular to account for various effects of preferential choice simultaneously.

Conclusion

The current paper describes two empirical studies aimed at demonstrating the probabilistic property of intertemporal choice and explores a number of brand-new models to accomodate this important feature. The results of both studies strongly support the general conclusion that intertemporal choice is probabilistic in nature just like other preferential choice. Additionally, diffusion models involving weighted additive difference appeared to be better than both the random utility model and the diffusion models built upon traditional hyperbolic discount function. More importantly, a couple of weighted additive difference models also provide a single general framework for the delay amount effect, common difference effect, and magnitude effect in intertemporal choice simultaneously. Our results also suggest that Model 2 provides a parsimonious model with sufficient accuracy.

References

[1] Ainslie, G. (1974). Impulse control in pigeons. Journal of the experimental analysis of behavior, 21, 485-489.

[2] Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. Psychological bulletin, 82, 463-496.

[3] Ainslie, G., & Herrnstein, R. (1981). Preference reversal and delayed reinforcement. Animal Learning and Behavior, 9, 476-482.

[4] Ainslie, G. (1992). Picoeconomics: The strategic interaction of successive motivational states with the person. Cambridge, England: Cambridge University Press.

[5] Brandstatter E., Gigerenzer, G, & Hertwig, R. (2006). The priority heuristic: Making choices without tradeoffs. Psychological Review, 113, 409-432.

[6] Busemeyer, J., & Townsend, J. (1992). Fundamental derivations from decision field theory. Mathematical Social Sciences, 23, 255-282.

[7] Busemeyer, J., & Townsend, J. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. Psychological Review, 100, 432-59.

[8] Chapman, G. & Winquist, J. (1998). The magnitude effect: Temporal discount rates and restaurant tips. Psychonomic Bulletin & Review, 5, 119-123.

[9] Christensen-Szalanski, J. (1984). Discount functions and the measurement of patients' values. Medical Decision Making, 4, 47-58.

[10] Fisher, I. (1930). The Theory of Interest. Macmillian, New York.

[11] Gonzalez-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. Psychological Review, 109, 137-155.

[12] Green, L., Fisher, B., Perlow, S., & Sherman, L. (1981). Preference reversal and self-control: Choice as a function of reward amount and delay. Behaviour Analysis Letters, 1, 43-51.

[13] Green, L., Fry, A. F., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. Psychological Science, 5, 33-36.

[14] Green, L., Myerson, J., & McFadden, E. (1997). Rate of temporal discounting decreases with amount of reward. Memory & Cognition, 25, 715-723.

[15] Johnson, J., & Busemeyer, J. (2005). A dynamic, stochastic, computational model of preference reversal. Psychological Review, 112, 841-861.

[16] Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47, 263-292.

[17] Kirby, K.N. & Marakovic, N.N. (1995). Modeling myopic decisions: Evidence for hyperbolic delay-discounting within subjects and amounts. Organizational Behavior & Human Decision Processes. 64, 22-30.

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

[19] Loewenstein, G. (1988). Frames of mind in intertemporal choice. Management Science. 34, 200-214.

[20] Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal choice. The Journal of Economic Perspectives, 3, 181-193.

[21] Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: evidence and an interpretation. The Quarterly Journal of Economics, 107, 573-597.

[22] Loewenstein, G., & Prelec, D. (1993). Preferences for sequences of outcomes. Psychological Review, 100, 91-108.

[23] 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, 5, 55-73. Mahwah, NJ: Erlbaum.

[24] McKerchar, T. L., Green, L., Myerson, J., Pickford, T. S., Hill, J. D., & Stout, S.C. (2009). A comparison of four models of delay discounting in humans. Behavioral Processes, 81, 256-259.

[25] Mosteller, F. & Nogee, P. (1951). An experimental measurement of utility. Journal of Political Economy, 59, 371-404.

[26] Myerson, J. & Green, L. (1995). Discounting of delayed rewards: Models of individual choice. Journal of the Experimental Analysis of Behavior, 64, 263-276.

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

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

[29] Rachlin, H. (1989). Judgment, decision, and choice: A cognitive/behavioral synthesis. New York: Freeman.

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

[31] Rae, J. (1834). The sociological theory of capital. http://www.archive.org/details/sociologicaltheo00raej

[32] Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-chocie reaction time. Psychological Review, 111, 333-367.

[33] Roe, R., Busemeyer, J., & Townsend, J. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. Psychological Review, 108, 370-92.

[34] Samuelson, P. (1937). A note on measurement of utility. The Review of Economic Studies, 4, 155-161.

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

[36] Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461-464.

[37] Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. Journal of Risk and Uncertainty, 5, 297-323.

[38] Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. Economic Letters, 8, 201-207.

[39] von Neumann, J, & Morgenstern, O. (1944). Theory of Games and Economic Behavior. Princeton, NJ: Princeton University.

Appendices

Appendix A - Typical questions involved in the current studies

To generate a reasonable range of questions for each subject, an adjustment procedure was utilized at the beginning of both studies. Consequently, each subject answered a unique set of questions. However, due to the same procedure of generating the questions, question sets were similar to one another. Following is a table showing a sample of the questions for a typical subject in the studies.

	smaller reward (in dollars)	shorter delay (in days)	larger reward (in dollars)	longer delay (in days)
Delay amount effect	20	1	34	3
	20	1	35	3
	21	1	34	3
	21	1	35	3
	20	2	34	6
	20	2	35	6
	21	2	34	6
	21	2	35	6
	...	...	...	...
	20	40	34	120
	20	40	35	120
	21	40	34	120
	21	40	35	120
Common difference effect	23	1	31	41
	23	1	32	41
	24	1	31	41
	24	1	32	41
	23	2	31	42
	23	2	32	42
	24	2	31	42
	24	2	32	42
	...	...	...	...
	23	40	31	80
	23	40	32	80
	24	40	31	80
	24	40	32	80
Magnitude effect	1	7	2	83
	1	7	2	84
	1	8	2	83
	1	8	2	84
	2	7	4	83
	2	7	4	84
	2	8	4	83
	2	8	4	84
	...	...	...	...
	40	7	80	83
	40	7	80	84
	40	8	80	83
	40	8	80	84

Appendix B - Details of the model fitting procedure

The first step in the fitting process was to prune the data for each individual in terms of response time so that outliers were excluded from further analyses. Specifically, any question with a response time shorter than

1500ms or longer than 10000ms was removed. We set the lower bound to ensure that subjects had sufficient time to sample all the information presented on the screen, and chose the upper bound to avoid the effect of extremely long response times on parameter estimation when response time was taken into account. It turned out that most response times were below the upper bound. For the weighted additive difference models, the attentional weight parameters were constrained between .05 and .95 to avoid extreme values that were practically meaningless. Besides, the utility parameters for direct differences were limited between .01 and 2 and those for relative differences were no smaller than .01. Since objective relative differences tend to be quite small in value compared with objective direct differences, no upper limit was applied to the utility parameters for relative differences so that different types of subjective differences could be equally influential. For all the diffusion models, both decision threshold, θ, and initial position, z, were set as proportional to the diffusion parameter, σ, and the proportional constants, θ^* and z^*, were the actual free parameters we estimated. This is a common practice when DFT models are implemented. Furthermore, the non-decisional component of response time, T_{er}, was constrained by an upper limit equal to the shortest response time produced by an individual when the data of that individual were fitted. For the diffusion models and random utility model based on the hyperbolic discount function, the utility parameter, α, was limited between .01 and 2, the discounting parameter, k, was no smaller than .0001, and the variability parameter, σ, could assume any positive value. With these limitations in place, the nine models were fitted to individual choice response data and/or response time data and maximum-likelihood estimation was utilized to find the best parameter values. In addition, the BIC index was calculated for each model and subject so that model complexity could be taken into account when models were compared to one another.

Appendix C - Proof of the delay amount effect as an indication of the stochastic nature of intertemporal choice

In both studies, it was found that choice proportions of the LL options changed gradually when delay amounts were increased in a proportional way. This was interpreted as a demonstration of the probabilistic nature of intertemporal choice. We will show the validity of this deduction by proving its converse-negative proposition, that is, a deterministic approach entails an abrupt change in choice proportion under this circumstance given either the exponential discount function or the hyperbolic discount function. We will also show that the converse-negative proposition is still true even if the magnitude effect is taken into account. First of all, Suppose a decision maker is indifferent between two intertemporal options, (v_a, t_a) and (v_b, t_b), and $0 < v_a < v_b, 0 < t_a < t_b$. According to the exponential discount function which assumes a deterministic perspective, we have

$$U(v_a, t_a) = v_a \times \delta^{t_a} = v_b \times \delta^{t_b} = U(v_b, t_b)$$

in which $U(\cdot, \cdot)$ represents the discounted subjective utility of an option and $0 < \delta < 1$. If both delays are multiplied by a constant m greater than 1, the new SS option, (v_a, mt_a) will be more preferable to the new LL option, (v_b, mt_b), since

$$U(v_a, mt_a) = v_a \times \delta^{mt_a} = v_a \times \delta^{t_a} \times \delta^{(m-1)t_a} = v_b \times \delta^{t_b} \times \delta^{(m-1)t_a}$$

$$> v_b \times \delta^{t_b} \times \delta^{(m-1)t_b} = v_b \times \delta^{mt_b} = U(v_b, mt_b)$$

When the magnitude effect is taken into account, the same result will occur given the exponential discount function. In this case, the magnitude effect requires that $0 < \delta_a < \delta_b < 1$ since $v_a < v_b$, and the indifference between the two options implies that

$$U(v_a, t_a) = v_a \times \delta_a^{t_a} = v_b \times \delta_b^{t_b} = U(v_b, t_b)$$

Because $v_a < v_b$, we have $\delta_a^{t_a} > \delta_b^{t_b}$. Consequently, for the new pair of options (v_a, mt_a) and (v_b, mt_b),

$$U(v_a, mt_a) = v_a \times \delta_a^{mt_a} = v_a \times \delta_a^{t_a} \times \delta_a^{(m-1)t_a} = v_b \times \delta_b^{t_b} \times (\delta_a^{t_a})^{m-1}$$

$$> v_b \times \delta_b^{t_b} \times (\delta_b^{t_b})^{m-1} = v_b \times \delta_b^{mt_b} = U(v_b, mt_b)$$

In other words, the new SS option will again be preferred over the new LL option.

For the hyperbolic discount function, the indifference between (v_a, t_a) and (v_b, t_b) implies that

$$U(v_a, t_a) = \frac{v_a}{1 + kt_a} = \frac{v_b}{1 + kt_b} = U(v_b, t_b)$$

and thus

$$\frac{v_a}{v_b} = \frac{1 + kt_a}{1 + kt_b}$$

When both delays are increased proportionally,

$$\frac{U(v_a, mt_a)}{U(v_b, mt_b)} = \frac{v_a/(1 + kmt_a)}{v_b/(1 + kmt_b)} = \frac{1 + kt_a}{1 + kt_b} \times \frac{1 + mkt_b}{1 + mkt_a} = \frac{1 + k(t_a + mt_b) + mk^2 t_a t_b}{1 + k(t_b + mt_a) + mk^2 t_a t_b}$$

Because $t_a < t_b$ and $m > 1$, we have $t_a + mt_b > t_b + mt_a$. Therefore, $\frac{U(v_a, mt_a)}{U(v_b, mt_b)} = \frac{1+k(t_a+mt_b)+mk^2 t_a t_b}{1+k(t_b+mt_a)+mk^2 t_a t_b}$ is greater than 1, indicating that the new SS option is preferable to the new LL option. When the magnitude effect is taken into account, the indifference implies that

$$U(v_a, t_a) = \frac{v_a}{1 + k_a t_a} = \frac{v_b}{1 + k_b t_b} = U(v_b, t_b)$$

Because $v_a < v_b$,we have $1 + k_a t_a < 1 + k_b t_b$ and thus $k_a t_a < k_b t_b$. Consequently,

$$\frac{U(v_a, mt_a)}{U(v_b, mt_b)} = \frac{v_a/(1 + k_a mt_a)}{v_b/(1 + k_b mt_b)} = \frac{1 + k_a t_a}{1 + k_b t_b} \times \frac{1 + mk_b t_b}{1 + mk_a t_a} = \frac{1 + (k_a t_a + mk_b t_b) + mk_a k_b t_a t_b}{1 + (k_b t_b + mk_a t_a) + mk_a k_b t_a t_b} > 1$$

since $k_a t_a + mk_b t_b > k_b t_b + mk_a t_a$. All in all, increasing both delays proportionally will make the SS option more attractive, given people are indifferent between the original pair. Similar reasoning can be invoked for the situation where the original pair of options are not equally appealing. In this case, the ratio of discounted utilities will increase when both delays are increased proportionally. The monotonic change pattern guarantees that, when both delays are increased in a proportional way, there exists only one cutoff point on the shorter delay amount (or equivalently on the longer delay amount) at which people are indifferent between the SS and LL options. Besides, for any pair with shorter delays, the SS option should be chosen, and for any pair with longer delays, the LL option should be chosen. That is to say, choice probability of the LL option should change from 1 to 0 at the cutoff point. When similar questions are combined into subgroups as in current studies, only one subgroup can have a choice proportion that is between 0 and 1 given a deterministic view on intertemporal choice. Since most subjects did not produce this pattern, the probabilistic nature of intertemporal choice is self-evident.

Acknowledgements

The experiments reported in this article were realised using Cogent 2000 developed by the Cogent 2000 team at the FIL and the ICN and Cogent Graphics developed by John Romaya at the LON at the Wellcome Department of Imaging Neuroscience.