Noisy coding of time and discounting for money*

Ferdinand M. Vieider†[1,2]

[1] *RISLαβ, Department of Economics, Ghent University*
[2] *RISLαβ Africa, AIRESS, University Mohammed VI Polytechnic*

March 30, 2023

Abstract

Pronounced discounting of future rewards has been explained by either preferences or a variety of psychological motives, ranging from the 'miseries of self-denial' to difficulties imagining future utilities. Here, I depart from these traditional explanations by modelling reward discounting as noisy perception of time delays between a smaller-sooner and a larger-later reward. The model resolves a long-standing puzzle in economics—why substantial discounting is routinely observed in experiments using monetary payoffs instead of consumption. The model predicts discounting to be stationary except when the sooner outcome is truly immediate, in which case it predicts present-bias. The level of patience that is observed, however, is predicted to systematically depend on the length of the time delays used to measure discounting. The model builds on the intuition that people react to the complexities of the choice situation by quickly and approximately trading off time delays against relative rewards. Three experiments provide support for the stylized behavioural patterns predicted by the model. Denominating equivalent delays in days rather than weeks substantially increases impatience. Providing a visual aid that focuses attention on time delays reduces present-bias. These findings cannot be organized based on either psychological or preference-based explanations of discounting, which treat time as *objective*. They are, however, predicted by the noisy numerical comparison framework underlying the model I present.

*I acknowledge financial support from the Research Foundation—Flanders (FWO) under the project "Causal Determinants of Preferences" (G008021N). The experiments have received ethics approval from the ethics committee at the Faculty of Economics and Business, Ghent University, with number *FEB 2021 M.* I am grateful to Ranoua Bouchouicha for help in executing experiment I. I am further indebted to Mohammed Abdellaoui, Benjamin Enke, Thomas Epper, Xavier Gabaix, Thomas Graeber, Olivier L'Haridon, Nick Netzer, Ryan Oprea, Peter Wakker, and Leeat Yariv for helpful comments and discussions. All errors remain my own.

†Sint-Pietersplein 6, 9000 Ghent, Belgium; Email: ferdinand.vieider@ugent.be

1 Motivation

Humans as well as animals have long been known to substantially discount future rewards. To use an example from experiment I, an outcome received 6 weeks from now is considered only as good as 86% of that same outcome in the present (see point A in figure 1). This implies an annualized discount rate in excess of 300% assuming continuous compounding. Such pronounced discounting of future rewards has been explained either by preferences, or attributed to a variety of psychological motives. Such psychological motives include the 'misery of delayed gratification', visceral forces, and difficulties imagining future utilities, among others (see Frederick, Loewenstein and O'Donoghue, 2002, sec. 2, for a review).

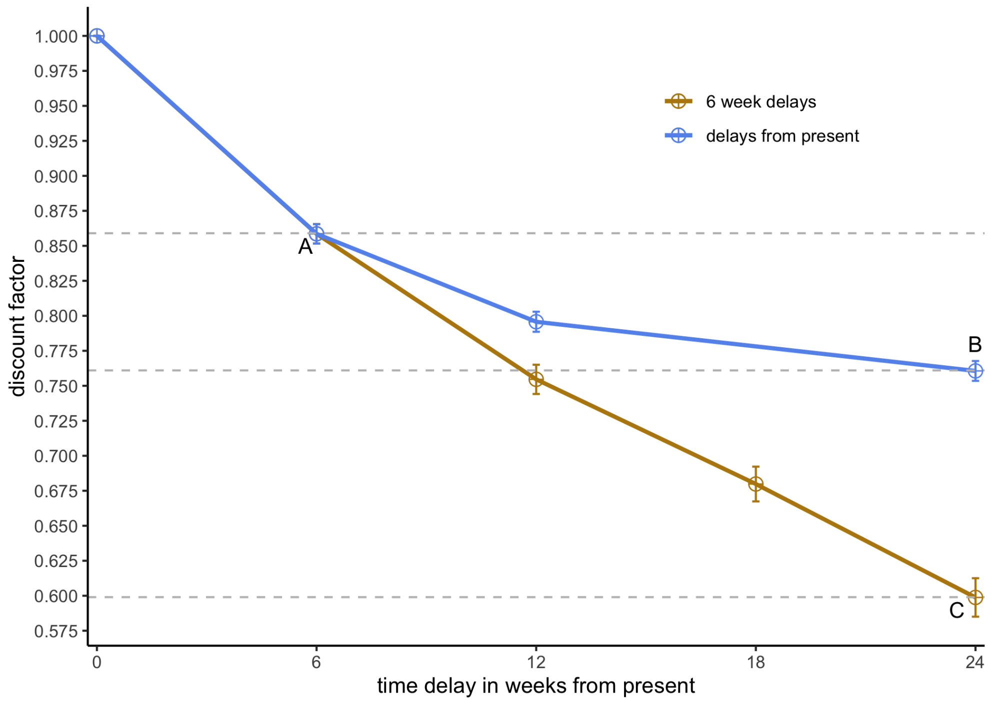


Figure 1: Nonparametric discount functions based on different time delays

Patterns shown are based on an incentivized classroom experiment with 175 students, using tradeoffs between smaller-sooner and larger-later rewards (see section 3). The figure shows discount functions obtained based on different types of choice stimuli. The function labelled 'delays from present' uses delays of 6, 12 and 24 weeks from the present. The '6 week delays' function is obtained by using delays of 6 weeks from the present, 6 weeks from 6 weeks, 6 weeks from 12 weeks, and 6 weeks from 18 weeks, so that the discount factor for a 12-week delay, δ_{12}, is calculated as $\delta_{12} = \delta_{0,6} \times \delta_{6,12}$, where the subscripts indicate the sooner and later time delays attached to the smaller and larger payoffs, respectively. Following the same procedure, we obtain $\delta_{18} = \delta_{0,6} \times \delta_{6,12} \times \delta_{12,18}$, and $\delta_{24} = \delta_{0,6} \times \delta_{6,12} \times \delta_{12,18} \times \delta_{18,24}$.

These explanations, however, leave some central issues unaddressed. Models enshrining the motives mentioned above explain discounting for *consumption*.

Point A in figure 1, however, has been obtained with an experiment using *monetary* payoffs. Substantial discounting for money is puzzling from an economic point of view, since one would need to assume immediate consumption of any monetary payouts to explain it (Cubitt and Read, 2007; Cohen, Ericson, Laibson and White, 2020).[1] This important puzzle remains unaddressed, with most experiments on time discounting using monetary payoffs instead of consumption.

Here, I present a model of delay discounting based on the premise that people infer the true time delays between payouts from noisy signals by optimal combination with a Bayesian prior. While the information aggregation process is optimal, noise arising from the quick and approximate comparison of time delays and relative payoffs implies that the inferred time delays will be systematically distorted. The noisiness of the process derives from limited attention being paid to the time dimension in quick comparative assessments of smaller-sooner versus larger-later rewards. Since the distortion affects the perception of *time delays*, the model explains pronounced discounting for money.[2] Principles of computational efficiency inspired by neuroscience furthermore result in a prediction of present-bias, which is thus based on a distinct mechanism. As a given delay between the sooner and later reward is pushed farther into the future, however, the model predicts stationary behaviour. In this sense, the model provides micro-foundations for quasi-hyperbolic discounting (Laibson, 1997; Imai, Rutter and Camerer, 2021).

At the same time, the model predicts that the measured discount rate will systematically depend on the length of the delay used to measure it—a phenomenon I call *delay-dependence*. This happens because distortions will apply to *delays between different rewards*. The same mechanism generating the iconic hyperbolic pattern in delays from the present and leading up to point B in figure 1, which has

[1] In particular, discounting measured from monetary payoffs can in general only be used to infer discounting for consumption when discount rates fall between the market borrowing and lending rates. Cubitt and Read (2007) show that for discount rates exceeding the market borrowing rate—as is clearly the case for point A in the figure—no inferences can be drawn on time preferences at all if one assumes agents to abide by standard principles of economic rationality. Discounting for money could, however, arise from the fundamental uncertainty of the future (Halevy, 2008; Chakraborty, Halevy and Saito, 2020; Epper and Fehr-Duda, 2023)—a point to which I will return below.

[2] Note that this does not imply that discounting must be *the same* for money and consumption, since additional patterns may emerge from outcome-distortions.

been obtained based on a 24-week delay from the present, thus also generates the delay-dependence leading up to point C, which has been obtained by multiplying the discount factors of subsequent 6-week delays from the present, from 6 weeks, from 12 weeks, and from 18 weeks.[3] The coexistence of these two discounting patterns cannot be organized by any of the standard discounting models, which predict *the same* patterns for the two types of measures. This holds for an exponentially decreasing discount function such as postulated in Paul Samuelson's (1937) discounted utility model, as well as for all functions from the hyperbolic family that have been proposed to account for the pattern leading up to point B, and for models emphasizing effects of the inherent uncertainty of the future (Halevy, 2008; Chakraborty et al., 2020; Epper and Fehr-Duda, 2023).[4]

I test the model in three experiments. Experiment I presents binary tradeoffs between smaller-sooner and larger-later rewards received with different time delays—the standard measurement paradigm in investigations of time discounting.[5] The results indicate the importance of present-bias and delay-dependence. I find no support for strongly decreasing impatience, which is not predicted by the model.[6] Predictive model tests based on cross-validation indicate that the noisy coding of time (NCT) model outperforms exponential discounting, as well

[3]At an annualized rate of close to 100% calculated from point B based on continuous compounding, it remains clearly beyond the range that would be compatible with discounting for money under standard explanations of discounting. The annualized rate obtained from point C, on the other hand, is close to the rate of 300% per annum obtained based on point A. Differences between the functions could, in principle, emerge based on implementation errors. The differences shown in figure 1 are, however, both too systematic and too substantial to derive from the typical 'white noise' attached to the deterministic decision models. I will discuss more sophisticated error models below.

[4]Read (2001), who first documented delay-dependence in discounting, makes the point that discount functions from the general hyperbolic family, such as Mazur's (1987) proportional discounting and Loewenstein and Prelec's (1992) hyperbolic discounting, were devised mainly to fit earlier evidence that was based purely on discounting measured by means of different delay lengths from the present. Several subsequent papers have used the same type of delays of varying length from the present (e.g., Ebert and Prelec, 2007; Zauberman, Kim, Malkoc and Bettman, 2009).

[5]Such tradeoffs are often collected into choice lists in applications, which nudge respondents towards indicating a *present equivalent* for delayed rewards. Here, I will use purely binary choice tasks, both because of their simplicity, and because of their consistency with the model I propose.

[6]Following Prelec (2004), I define strongly decreasing impatience as $(y, \tau_s) \sim (x, \tau_\ell) \rightarrow (y, \psi + \tau_s) \prec (x, \psi + \tau_\ell)$, where ψ, τ_s, τ_ℓ are nonnegative time delays, $\tau_s < \tau_\ell$, and $x > y$ are monetary outcomes. I define present-bias where the equivalent condition only holds when when the sooner outcome obtains immediately, i.e. $\tau_s = 0$.

as a large set of non-exponential discount functions plus additive noise used in the literature.

Experiment II tests a core prediction of the model—that time distortions are driven by the underlying numerical quantities, rather than by any inherently chronometric properties of time delays. Subjects are randomly assigned to see identical time delays denominated either in *weeks* or in *days*. Preference-based explanations of reward discounting predict no difference between the two conditions. The same holds true for explanations based on mechanisms such as 'self-control problems', 'visceral influences', 'miseries arising from delayed gratification', or difficulties in imagining future *utilities*, all of which model discounting for *consumption* and treat time delays as objectively perceived. Accounts of decreasing impatience based on the fundamental uncertainty of the future do also not predict any difference in this case, since they model the inherent properties of time while incorporating concerns about (contract) survival. The noisy coding of time model, however, predicts systematic differences based on the underlying numerical magnitudes. The reason for this lies in the observation that quick tradeoffs become more difficult and hence imprecise when delays are described in *days* rather than *weeks*. The experiment indeed shows sizeable differences in impatience depending on whether time delays are expressed in days or weeks in a way that is consistent with the model predictions.

Experiment III proposes a representational manipulation of time delays that uses a visual aid under the form of arrows with length proportional to the time delay. While making comparisons of time delays easier, thus implementing an opposite manipulation of the one in experiment II, experiment III aims specifically at increasing *attention* to the time dimension. This serves to test the idea that the mechanism generating present-bias, although conceptually distinct from the noisy perception mechanism manipulated in experiment II, is itself also driven by a lack of attention to the time dimension. The results indicate a decrease in noisiness, which results in the opposite effect of the one documented in experiment II. Concomitantly, present-bias is reduced considerably, which supports the attentional interpretation underlying the parameter capturing present-bias in the model.

The NCT model I present rest on a dual approach to the model of Gabaix and Laibson (2017). The latter model a perfectly patient Bayesian decision-maker who produces noisy simulations of future *utilities*. If the simulation noise increases linearly with the time delay to the consumption outcome, the combination with a zero-mean prior results in a proportional discount function as proposed by Mazur (1987). The model of Gabaix and Laibson (2017) thus builds microfoundations for explanations of discounting based on difficulties imagining future utilities. Time delays, on the other hand, are perceived *objectively* according to the model. In contrast, the NCT model rests on the noisy perception of *time delays*. Given the difference in setup and predictions, and the similarity in formalism, the two models can be thought of as complementing each other.[7]

Formally, the model shares a common underpinning with several recent papers modelling the effect of noisy, but otherwise optimal, cognitive processes in complex choice environments (Natenzon, 2019; Khaw, Li and Woodford, 2021; Vieider, 2021). It also shares a common intuition with evolutionary explanations of decision-making (Robson, 2001; Netzer, 2009; Netzer, Robson, Steiner and Kocourek, 2021). The NCT model builds on the idea that the tradeoffs between the multiple quantities involved in choices between smaller-sooner and larger-later options trigger approximate comparison strategies that may introduce systematic bias into the decision process. This intuition has affinities with recent contributions that have showcased the importance of 'cognitive uncertainty' (Enke and Graeber, 2019). The intuition of noisy cognition in complex decision situations is furthermore consistent with experiments that have documented how patterns that had been attributed to risk attitudes may in reality be triggered by the complexity of the choice situation (Oprea, 2022). Implementing a design imitating the complexities of intertemporal tradeoffs in an atemporal setting, Enke, Graeber

[7] For instance, the NCT model does not predict strongly decreasing impatience, in the sense of impatience decreasing for identical time delays as they are pushed farther into the future. While the experiments in this paper do no support strongly decreasing impatience, this may change when outcomes are consumption goods and when delays become very long, being measured e.g. on a time scale of years instead of weeks or months. In such cases, a combination of the two models may be desirable, which is straightforward given the models' technical similaries.

and Oprea (2023) have documented results that closely resemble what one may consider 'typical discounting patterns'. The modelling approach I take in this paper captures precisely this intuition, whereby choice patterns are driven by quick comparisons in complex tradeoffs, rather than anything inherently chronometric.

This paper proceeds as follows. Section 2 introduces the model. Section 3 describes experiment I, and section 4 presents experiment II, with section 5 presenting experiment III. Section 6 provides a discussion of the results and concludes the paper.

2 The noisy coding of time model

Modelling preliminaries

I model the condition under which a decision-maker chooses a larger-later reward x, paid at time τ_ℓ, over a smaller-sooner amount y, paid at time τ_s. I start from a choice rule devoid of subjective parameters[8], under which the larger-later amount will be chosen whenever

$$e^{-(\tau_\ell - \tau_s)} > \frac{y}{x}. \tag{1}$$

This choice rule emphasizes the tradeoff of the time delay between the two rewards against their relative size. It is optimal inasmuch as discounting is stationary, resulting in consistent choice patterns over time. The substantial discounting of 100% per time unit reflects the fundamental intuition underlying the model that patience—and perhaps the *the very meaning of time*—needs to be learned from experience. This take is consistent with extremely pronounced discounting observed in children (Mischel and Ebbesen, 1970; Mischel, Shoda and Rodriguez, 1989; Bettinger and Slonim, 2007), and with the emphasis put on education in economic models endogenizing discounting (Doepke and Zilibotti, 2014; 2017).

The central idea underlying the model is that choice quantities are not perceived directly, but are encoded by a signal that needs to efficiently represent these

[8]Note that this choice rule is used without loss of generality. Augmenting the choice rule by a normatively low discount rate or by a concave utility function does not affect any of the conclusions derived below. I thus focus on a minimalistic setup deprived of any further motives.

quantities in a way suitable for neural calculations. I will thus assume that the choice rule in (1) is logged twice, and that the objective time delay $t \triangleq \tau_\ell - \tau_s$ is replaced by the posterior inference on the time delay drawn from a signal r:

$$E\left[ln(t)\,|\,r\right] < ln\left(-ln\left(\frac{y}{x}\right)\right), \tag{2}$$

where I treat rewards as being perceived objectively and without distortion for simplicity (including noisy inferences on outcome ratios does not affect the conclusions drawn from the model—see online appendix C for a generalization to include outcome distortions). The double-logging of the choice rule serves to put it on the scale of the neural computations, which are supposed to take place on a logarithmic scale. This adds neuro-biological realism, since it puts all subsequent computations on a linear scale. The conclusions below, however, do not depend on this assumption, and deriving the model on the scale of the choice rule in (1) results in model predictions that are empirically indistinguishable from the ones derived below—see online appendix B for such an alternative derivation.

The choice rule above assumes that the signal r encodes the time delay between two rewards. This captures the intuition of noise arising from the approximate comparisons of the time delays attached to the rewards being traded off against the ratio of the rewards themselves. Expressing the signal r on the logarithmic scale serves to avoid an unbounded increase in the cognitive resources needed for the representation of larger numbers (Dehaene and Changeux, 1993; Dayan and Abbott, 2001; Dehaene, 2003). Gold and Shadlen (2001) showcase the neural advantages of logarithmic coding in a comparative setting. Howard and Shankar (2018) have shown theoretically that logarithmic compression is optimal for adaptation to diverse environments. This conclusion is general, and holds *independently from the statistical distribution of stimuli in the environment.*[9]

Given limits on neuronal resources, as well as limits imposed by the precision with which such a signal can be optimally decoded (Dayan and Abbott, 2001,

[9]Behaviourally, Zauberman et al. (2009) have shown that the dependence of subjectively perceived delays on objective delays from the present may be best fit by a logarithmic function.

chapter 3), the mental signal will typically be noisy. Noise will be especially relevant when comparisons and tradeoffs between multiple dimensions in complex choice situations are made quickly and intuitively, rather than being based on precise calculations (Whalen, Gallistel and Gelman, 1999). Lim, O'Doherty and Rangel (2011) show that in such comparative settings neuronal activity correlates with the value difference between items being compared, thus providing direct physiological evidence for comparative coding. The model presented below explicitly builds on the idea that in complex choice situations quantities relevant to the decision will often be gauged approximately, potentially yielding systematic biases in the assessments of the choice quantities.

Logarithmic representation of time delays incurs into issues when a time delay τ becomes small. Since $\lim_{\tau \to 0} ln(\tau) = -\infty$, the resources needed to represent small delays would increase unboundedly, thus contradicting the very resource-saving rationale underlying the logarithmic representation. Howard and Shankar (2018) argue that this resource-saving rationale dictates that a number κ be added to the objective quantity. I will thus substitute $s \triangleq \kappa + \tau_s$ if $\tau_s > 0$ and $\ell \triangleq \kappa + \tau_\ell$ for the objective time delays to model mental representations of the objective delays in order to prevent the numerical representation to become boundless as $\tau_s \to 0$. While this transformation is inconsequential for $\tau_s > 0$, since the additive κ will drop out of the time delay, i.e. $\ell - s = \tau_\ell - \tau_s$, it will have interesting implications for $\tau_s \equiv 0$ and when time delays τ_ℓ are small. This may be best thought of as an arbitrary numerical fix to a computational problem that is itself subject to adjustment whenever the importance of the situation warrants this. In this sense, κ may arise from inattention to some aspects of the decision problem when decisions are taken quickly and tradeoffs are gauged approximately.

Bayesian inference on time delays

Below, I formalize how signals are encoded and decoded to arrive at posterior inferences on the delay that triggered the signal. For simplicity, I treat rewards as being perceived objectively and without distortion. Allowing for the noisy distortion of outcomes reinforces the intuition of noise arising from quick and approximate tradeoffs, but does not otherwise affect the conclusions drawn from the model (see online appendix C for a generalization to include outcome distortions). I model the mental signal as a single draw from the following distribution:

$$r \sim \mathcal{N}(\, ln(t)\,,\, \nu^2\,), \tag{3}$$

where the parameter ν quantifies coding noise, and is assumed to be locally independent of t.[10] Physiological evidence directly suggests the use of a normal distribution (Nieder and Miller, 2003; Dehaene, 2003). The normal distribution is furthermore plausible because of its analytical advantages for the mind, allowing for the derivation of a precise posterior based on only 3 quantities to be stored in memory. More general, non-conjugate functions, on the other hand, would require thousands of points to obtain an accurate approximation of the posterior.[11]

The information contained in the signal r needs to be decoded to make inferences about the outside stimulus that has generated the signal. To do this, the mind builds generative models of the world, enshrined in probabilistic prior distributions learned over time, which enable causal inference about the underlying triggers of the signals. This noisy signal r is subsequently decoded by combination with a prior indicating the probability of different time delays in the environment. Such a process is indeed optimal given the noise contained in the signal, thus reducing error. I will assume a conjugate prior distribution from the normal family:

$$ln(t) \sim \mathcal{N}(\, \mu\,,\, \sigma^2\,). \tag{4}$$

The log-normal form of the prior enshrines the observation that most delays one

[10] Local independence as defined here implies that the coding noise does not change with the time delays used. Note that this assumption is only expected to hold *locally*, in the sense that one should not expect it to hold when transitioning between different orders of magnitude, or indeed for different numerical representations of identical time delays.

[11] It is worth noting that the central results do not hinge on the normality assumption. Showcasing the efficiency of logarithmic coding in neural representations, Gold and Shadlen (2001) showed how the posterior inferences obtained based on the normal setup hold for a wide class of symmetric and asymmetric distributions.

faces are short while some are very long, and naturally reflects the non-negative nature of time delays. The conjugate form to the normal likelihood further has important analytical advantages, and allows to quickly access the posterior.

Combining the likelihood in equation 3 with the prior in equation 4 by Bayesian updating (see online appendix A for details), we obtain the posterior inference on the time delay that is expected to have caused the signal:

$$E[t \,|\, r] = \beta r + ln(\alpha), \tag{5}$$

where $\alpha \triangleq exp\,(\,(1-\beta)\times\mu\,)$, and the Bayesian shrinkage weight $\beta \triangleq \frac{\sigma^2}{\sigma^2+\nu^2}$ governs the degree of regression of the posterior mean to the mean of the mental prior, μ. Unexpected, noisy signals falling far from the mean of the prior will thus be discounted more heavily than expected signals characterized by high precision. For values of $\beta < 1$, any noisy signal will be shrunk towards the prior mean μ, with the direction of the shrinkage determined by whether r is smaller or larger than μ. Intuitively, signals smaller than the mean will thus be increased, suggesting time delays that are longer than they actually are, and thus leading to increased impatience. Signals larger than the prior mean, on the other hand, will be reduced, thus suggesting time delays that are shorter than the objective delay, and thus increased patience. Whether an individual is relatively patient or impatient for a given delay will thus depend on the interplay between β and α.

The stochastic choice rule

The posterior expectation of the time delay can now be used to inform decisions. Substituting the posterior expectation in (5) into the mental choice rule in (2) and rearranging yields a threshold equation capturing the conditions under which the larger-later reward will be chosen over the smaller-sooner reward:

$$r < \beta^{-1}\left[ln\left(-ln\left(\frac{y}{x}\right)\right) - ln(\alpha)\right]. \tag{6}$$

This threshold equation fully describes the condition under which the larger-later amount will be chosen. It shows how the signal about the time delay is traded off against the reward ratio, which in turn is transformed by the subjective parameters emerging from the Bayesian updating process.

To result in testable predictions we need to obtain a choice rule devoid of the mental signal r, which is unknown to external observers. Obtaining the z-score for r from the likelihood in (3) and comparing it to the z-score of the threshold equation (see online appendix A for details), we obtain the following probabilistic choice rule:

$$Pr[(x,\tau) \succ (y,0)] = \Phi\left(\frac{ln\left[-ln\left(\frac{y}{x}\right)\right] - [ln(\alpha) + \beta\, ln(t)]}{\beta\, \nu}\right), \tag{7}$$

where $Pr[(x,\tau) \succ (y,0)]$ represents the probability of the larger-later reward being chosen, and Φ represents the standard normal distribution. In contrast to traditional discounting models, which typically combine a deterministic preference model with an independently and arbitrarily chosen stochastic choice model (He, Golman and Bhatia, 2019), the probabilistic setup used to derive the noisy coding model produces both the choice parameters *and* the stochastic choice setup, thus resulting in an inherently stochastic model of inter-temporal choice.

Behavioural implications

When modelling delays from the present, i.e. for $\tau_s \equiv 0$ and $t = \ell$, equation (7) can be seen as providing stochastic microfoundations for the constant sensitivity discount function of Ebert and Prelec (2007). This is most easily seen by zooming in on the equality condition in the numerator of (7) while momentarily ignoring the stochastic aspect. Defining $\delta_\ell \triangleq \frac{y}{x}$ and exponentiating the expression in the numerator twice yields the following expression at the point of indifference:

$$\delta_\ell = exp(-\alpha \ell^\beta). \tag{8}$$

Ebert and Prelec (2007) show that α creates a demarcation between the near and the far future. Discounting for delays shorter than $\frac{1}{\alpha}$ will increase in time insensitivity relative to the exponential benchmark, whereas discounting for delays longer than $\frac{1}{\alpha}$ will decrease. This insight resonates with the interpretation of the NCT model, whereby α captures regression of time perception to the mean of the prior. Ebert and Prelec (2007) show experimentally that time sensitivity β increases when a visual aid for the length of the time delay is provided, and decreases under time pressure. Both results are perfectly aligned with an account whereby attention determines the level of coding noise in the NCT model.

The NCT model, however, differs from the constant sensitivity function in several respects. The definition of the encoded time delay as $\ell \triangleq \kappa + \tau_\ell$ implies that the function will show a discontinuous drop in the vicinity of $\tau_\ell = 0$. That is, $\lim_{\tau_\ell \to 0} D(t) = exp(-\alpha\kappa^\beta)$, where $D(t)$ indicates the discount function. Assuming as usual that $D(0) = 1$, so that immediate payments are not discounted—in the absence of any delay being indicated, there is no time delay to be encoded—this introduces a drop in the vicinity of $\tau_\ell = 0$ which predicts present-bias—a preference for immediately received rewards over even slightly delayed rewards. Present-bias is thus captured by the number κ, introduced to fix a numerical overflow problem which may itself be subject to attentional modulation. This interpretation is consistent with approaches modelling costly self-denial (Gul and Pesendorfer, 2001). Other than in the constant sensitivity function, time discriminability β is closely linked to decision error, and the model is inherently stochastic.

Once up-front delays are introduced, the NCT predicts that *delays between the two options* are nonlinearly distorted, whereas the constant sensitivity model predicts transformations of the individual time delays from the present, so that $\delta_{\tau_s,\tau_\ell} = exp(-\alpha(\tau_\ell^\beta - \tau_s^\beta))$. Much like other functions from the non-exponential family, reviewed in detail in online appendix D, the constant sensitivity function can thus not account for delay-dependence. The NCT, on the other hand, predicts that longer delays are compressed more, which implies that impatience systematically depends on the time delays used to measure it. For a given delay length, however, the function is stationary, i.e. it is independent of the value taken by the up-front delay τ_s except in the special case where $\tau_s = 0$. The fact that this function obtains naturally from an intuitive optimal choice rule may explain the pervasiveness of delay-dependence (Dohmen, Falk, Huffman and Sunde, 2017; Enke et al., 2023), given the model's *predictive* interpretation. Delay-dependence is ultimately produced by a lack of attention in approximate judgments. If more resources are dedicated to the correct perception of time, in the limit as $\nu \to 0$ revealed impatience will be independent of the length of the time delays used.

Relation to Gabaix and Laibson (2017)

Gabaix and Laibson (2017) recently proposed a setup that is formally related to the model presented here. They assume Bayesian agents who are perfectly patient, but who perceive future *utilities* with some noise, so that $s_\tau \sim \mathcal{N}(u(x_\tau), \xi_\tau^2)$, where $u(x_\tau)$ is the utility of a reward x received at time τ, and s_τ is the noisy simulation of that utility. This simulation is combined with a prior $u(x_\tau) \sim \mathcal{N}(\widehat{\mu}, \zeta^2)$. This yields the posterior expectation $E[u(x_\tau)|s_\tau] = \widehat{\mu} + D(\tau)(s_\tau - \widehat{\mu})$. Assuming that the noisiness of the signal increases linearly in the time delay τ, i.e. $\xi_\tau^2 = \xi^2 \times \tau$, they obtain $D(\tau) \triangleq \frac{\zeta^2}{\zeta^2+\xi^2\tau} = \frac{1}{1+\frac{\xi^2}{\zeta^2}\tau}$, which takes the form of the proportional discount function proposed by Mazur (1987), $D(\tau) = \frac{1}{1+\gamma\tau}$, with $\gamma \triangleq \frac{\xi^2}{\zeta^2}$. Note that while future utilities are subject to noisy simulations, time delays are perceived objectively. The model thus predicts discounting to be proportional to the time delay from the present, but does not predict delay-dependence or present-bias. The model of Gabaix and Laibson (2017) provides micro-foundations for psychological accounts according to which future utilities are uncertain. The model thus presents a view that is highly complementary to the one presented here.

3 Experiment I: Patterns of reward discounting

3.1 Experiment

Bachelor students attending an introductory class in behavioural economics at Ghent University were invited to take part in a classroom experiment. The students had been exposed to the basics of expected utility theory, but had not covered time discounting yet. Students were told to bring a laptop or tablet to class to participate in an experiment. They were told that the anonymized aggregate data would be used to illustrate typical aggregate choice patterns for teaching purposes. They were also told that 10 students would be randomly extracted to play one of their choices for real money immediately after the experiment. Overall, 175 students participated in the experiment and provided a complete set of responses during the allocated time.

The time delays used in the experiment are depicted in figure 2. They were chosen using simulations to allow for optimal identification of all model components. Importantly, I chose the stimuli in such a way as to allow for the identification of patterns predicted by the model—such as present-bias and delay-dependence—as well as to allow for the identification of patterns *not* predicted by the model—such as strongly decreasing impatience. In particular, comparison of AB to BC and of AC to CE allows for the identification of present-bias. Comparison of AB and BC with AC, of CD and DE with CE, and of all the 6 week delays (AB, BC , CD , DE) and 12 week delays (AC , CE) with the full delay over 24 weeks (AE) allow for the identification of delay-dependence. Strongly decreasing impatience can be identified from the comparison of BC, CD, and DE.

The future outcome was fixed at €50. The choice of such a round, invariant amount was meant to ensure that outcomes are perceived objectively, rather than being subject to noisy perceptions themselves (see online appendix C for a discussion of robustness to this assumption). The earlier amounts ranged between €33 and €49 inclusive in steps of €1, allowing for discount factors between 0.64 and 1 (discount rates between 0 and 56%) per period. Each screen presented one single

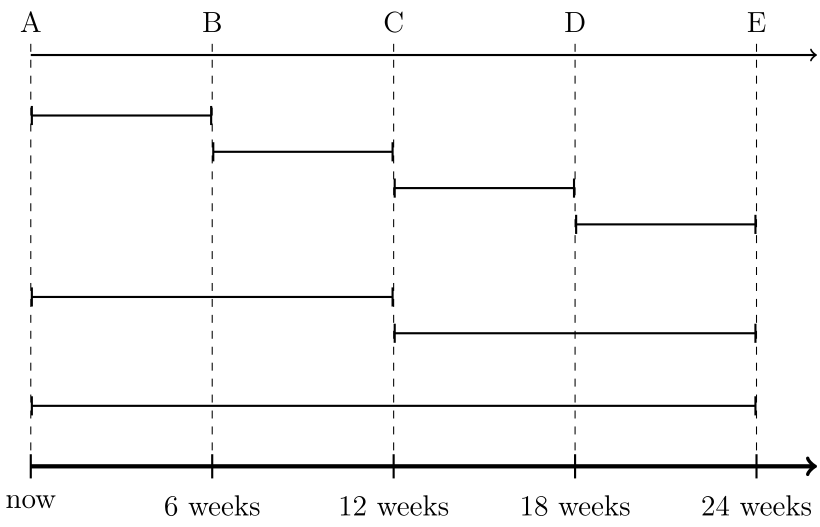


Figure 2: Time horizons used in experiment

Illustration of time delays used in the experiment. The maximum delay, indicated by AE, was 24 weeks. This delay was divided into 4 different sub-periods of 6 weeks, AB, BC, CD, and DE; and into 2 different sub-delays of 12 weeks, AC and CE. Comparison of AB to BC and of AC to CE allows for the identification of present-bias. Comparison of BC to CD and DE, and of the latter two, allow for the identification of strongly decreasing impatience. Comparison of long delays with their constituent parts allow for the identification of delay-dependence. Notice that, given 4 delays of 6 weeks, 2 of 12 weeks, and 1 of 24 weeks, the stimuli are well-fitted by a log-normal distribution.

choice, and the individual choice pairs were presented completely at random. This design was implemented to fit the discrete binary choice setup modelled. Subjects made 158 choices, which took about 20 minutes on average. All stimuli were presented at least once. In addition, 40 randomly selected stimuli were repeated. The repeated extractions were executed with replacement, so that the same stimulus may recur a number of times. Identification of decision noise, which plays a central role in the model, is thus assured by i) repeated observations of the same stimuli; and ii) monotonicity violations between similar stimuli.

Before the start of the experiment, the lecturer presented the instructions (online appendix G), to make sure that everybody had an understanding of the tasks. Since the tasks consisted of a binary choice paradigm, the actual explanations of the tasks were very simple. The lecturer, however, emphasized the procedural aspects of the payout mechanism. Both immediate and future payouts were made by bank transfer. Bank transfers between all major Belgian banks are immediate. This was emphasized in the instructions, and subjects were told that in case of an immediate payout being extracted to count for real pay, the lecturer would execute the payment directly and wait for the money to arrive on the student's account. The student would then be asked to verify if the money had arrived and sign a receipt. In case of a future payment, the lecturer signed a certificate on university letterhead. The certificate contained the amount to be paid and the date on which it would be paid, and it was signed by the lecturer. The certificate also contained the address and telephone number of the lecturer, and students were encouraged to contact the lecturer in case they changed bank accounts or they had any doubts about the payment. All time delays were chosen in such a way as to fall within the same academic year, to keep the costs of approaching the lecturer low, and to further reassure subjects of the future payment guarantee.

3.2 Results

I start from an examination of the nonparametric evidence for present-bias and strongly decreasing impatience, shown in figure 3. The figure plots decumulative choice proportions for the larger-later reward as the sooner-smaller reward increases from €33 through €49. The choice proportions for the 6-week delay from 6 weeks, shown in panel A, are shifted to the north-east of the choice patterns for the 6 week delay from the present, thus indicating present-bias ($p = 0.011$, two-sided Wilcoxon signed-rank test on individual-level choice proportions for the later option). Results for the 12 week delay from the present versus a 12-week delay from 12 weeks are very similar, and again indicate an increase in patience following the introduction of the up-front delay ($p \ll 0.001$; figure in online appendix F).

Panel B in figure 3 examines the evidence for strongly decreasing impatience by showing discounting patterns for 6-week delays from up-front delays of 6, 12, and 18 weeks (i.e., comparing time delays BC , CD, and DE). No systematic differences in discounting for the different upfront delays are apparent. Statistically, there is no difference in discounting between a 6-week delay from 6 weeks and a 6-week delay from 12 weeks ($p = 0.393$). There is also no difference when comparing a delay of 6 weeks from 12 weeks to a 6-week delay from 18 weeks ($p = 0.182$), or when comparing a 6-week delay from 6 weeks to a 6-week delay from 18 weeks

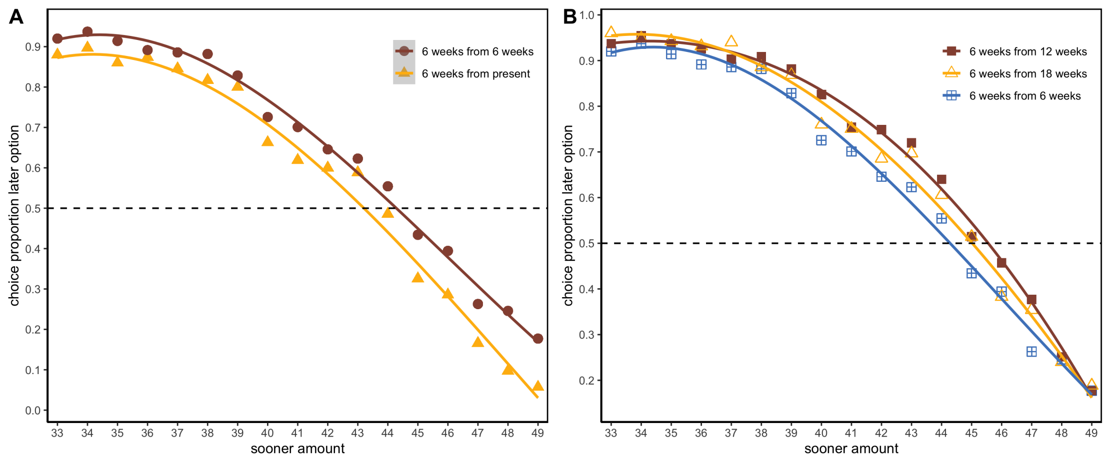


Figure 3: Non-parametric illustration of present-bias

Panel A compares a 6-week delay from the present to a 6-week delay from an up-front delay of 6 weeks using decumulative choice proportions for the larger-later reward as the sooner reward increases. The right panel shows the decumulative choice proportions for the 6-week delays with upfront delays of 6, 12, and 18 weeks. Panel A indicates clear evidence for present-bias, with the choice proportions of the later option systematically lower when the sooner option occurs in the present. Panel B indicates no systematic differences, providing no support for strongly decreasing impatience. The nonparametric choice proportions are fit with a third degree polynomial.

($p = 0.076$). Although this last comparison is marginally significant, it goes in the opposite direction of the previous comparison, resulting in an overall null result.

To test delay-dependence, I identify a probabilistic non-parametric discount factor from the sooner amount at which a subject starts choosing the smaller-sooner reward 50% of the time, divided by the later amount. Figure 4 shows two examples of delay-dependence. Panel A compares a 12-week delay from the present (AC) to its two underlying 6-week delays (AB and BC). Delay-dependence predicts $\delta_{0,12} > \delta_{0,6} \times \delta_{6,12}$, where δ is the discount factor and subscripts indicate time delays in weeks. The great majority of data points falls above the 45° line, indicating delay-dependence. Importantly, this also holds true when we exclude calculated discount factors $\delta_{0,6} \times \delta_{6,12}$ smaller than 0.64, which could otherwise bias the findings due to censoring effects ($p \ll 0.001$).

Panel B shows the patterns for the longest 24-week delay (AE) against the product of the discount factors for the 2 underlying 12-week periods (AC and CE). Delay-dependence is now revealed by behavioural patterns indicating that the discount factor over the whole period is larger than the product of the discount factors for the different underlying 12-week periods, $\delta_{0,24} > \delta_{0,12} \times \delta_{12,24}$. This effect is very pronounced, with virtually all points falling to the north-west of the 45

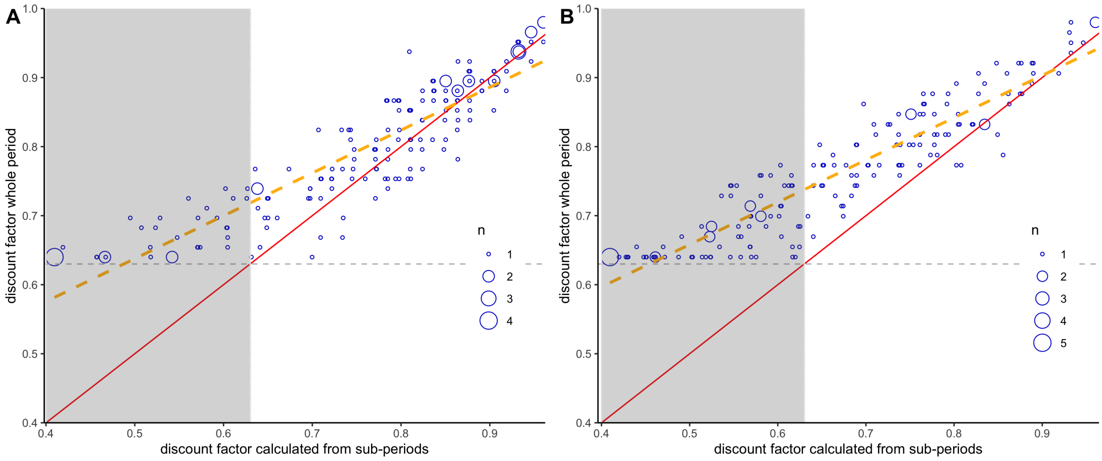


Figure 4: Non-parametric illustration of subadditivity

The comparisons are obtained by comparing a discount factor over the whole period with the product of the discount factors of the subperiods. The pattern in panel A obtains from a comparison of the discount factor $\delta_{0,12}$ with the product of the two discount factors $\delta_{0,6} \times \delta_{6,12}$, where the subscripts indicate the extremes of the time delays. The pattern in panel B obtains from a comparison of the discount factor $\delta_{0,24}$ to the product of the two underlying 12-week discount factors, $\delta_{0,12} \times \delta_{12,24}$. Dashed lines indicate correlations.

degree line. Once again, this effect is highly significant even after accounting for censoring and excluding individuals for whom the calculated discount factor from the shorter delays is lower than 0.64 ($p \ll 0.001$). Delay-dependence in the remaining comparisons are similar (see online appendix F).

I next estimate the noisy coding model, and test its predictive performance against different standard models augmented by a normally distributed, additive error term, as most commonly used in the literature. I use cross validation to test predictive performance, so that models with more parameters get penalized compared to simple models with fewer parameters. The point here is not to see whether the NCT model outperforms every other model, combined with any possible stochastic choice model, which would yield a staggering variety of combinations (see Regenwetter, Cavagnaro, Popova, Guo, Zwilling, Lim and Stevens, 2018, for a review). The point is rather to see whether the specific predictions of the model and the interaction between the parameters may give it an edge in terms of predictive performance when compared to other, 'typical' implementations.

Table 1 shows the test results, together with the estimated parameters. It also shows the discount functions corresponding to the different models—see online appendix D for a more detailed discussion and a literature review. The noisy cod-

Table 1: Mean parameter estimates and predictive performance test

model	discount par.	distortion par.	noise	ELPD diff.
noisy coding of time eq. 7	0.028 (0.001)	0.657 (0.018)	1.010 (0.013)	0 –
subadditive discounting $D(\tau_s, \tau_\ell) = {}^1/_{1+\zeta(\tau_\ell - \tau_s)^\gamma}$	0.046 (0.002)	0.596 (0.016)	0.124 (0.001)	−73.3 (32.2)
quasi-hyperbolic discounting $D(\tau) = \beta \times exp(-\rho\tau)$	0.012 (0.000)	0.950 (0.003)	0.152 (0.002)	−288.5 (36.1)
constant sensitivity discounting $D(\tau) = exp(-\widehat{\alpha}\tau^{\widehat{\beta}})$	0.030 (0.001)	0.763 (0.014)	0.152 (0.002)	−289.2 (36.7)
hyperbolic discounting $D(\tau) = ({}^1/_{1+\zeta\tau})^{\frac{\xi}{\zeta}}$	0.021 0.001	0.049 (0.005)	0.152 (0.002)	−299.2 (37.3)
as-soon-as-possible discounting $D(\tau_s, \tau_\ell) = {}^1/_{1+\zeta(\tau_\ell - \tau_s)}$	0.016 (0.000)	– –	0.146 (0.002)	−301.8 (36.6)
proportional discounting $D(\tau) = {}^1/_{1+\zeta\tau}$	– –	0.017 (0.000)	0.152 (0.002)	−330.2 (37.5)
exponential discounting $D(\tau) = exp(-\rho\tau)$	0.014 (0.000)	– –	0.154 (0.002)	−404.2 (38.5)

The parameters and standard errors (in parentheses) have been obtained in aggregate-level Bayesian estimations using an additive, normally distributed error term. The ELPD difference refers to the *expected log pointwise predictive density difference*, and measures the predictive performance of a model relative to the best-performing model measured by leave-one-out cross-validation (Vehtari, Gelman and Gabry, 2017). The ELPD difference constitutes a Bayesian generalization of the deviance information criterion. Results are virtually identical if I base the model tests on the Watanabe-Aikake Information criterion (WAIC) instead. Negative values indicate lower performance. The functional forms underlying the various models, and a review of their use in the literature, are include in online appendix D. The subjective parameter added to objective time delays is estimated at $\kappa^\beta = 2.71$ (standard error: 0.14).

ing model easily outperforms the exponential model. The NCT model also clearly outperforms the hyperbolic-class models (including proportional and hyperbolic discounting, as well as quasi-hyperbolic discounting, and the constant sensitivity model). The function coming closest in performance is Read's (2001) subadditive discount function. As-soon-as-possible discounting (Kable and Glimcher, 2010), which models delay-dependence for delays of different lengths from any upfront delay (see appendix D for details), performs less well.

4 Experiment II: Delays in days versus weeks

4.1 Experiment

Experiment I provided a proof of concept for the predictions of the NCT model. In particular, I showed that i) the empirical data line up with the stylized patterns predicted by the NCT model; and ii) the model outperforms a large set of standard discount functions used in the literature in predictive fit. Experiment II aims to directly test the approximate comparison explanation of time discounting underpinning the NCT model. Subjects were randomly allocated to identical time delays described either in *weeks* or in *days*. That is, a 12-week delay was presented in the days treatment as a delay of *84 days*. If the observed patterns are due to either preferences or motives such as self-control issues or difficulties imagining future utilities or needs, the representation of time delays in either days or weeks should make no difference. We also would not expect any differences based on the inherent uncertainty of the future, which remains identical across treatments. The same holds true for the model of Gabaix and Laibson (2017), given that in the latter time is perceived *objectively*.[12]

The NCT model, on the other hand, predicts systematic differences when identical delays are expressed either in days or in weeks due to the difference in numerical magnitudes. In particular, expressing time delays in days rather than weeks is expected to increase coding noise by making quick comparisons more difficult. Since the prior variance is learned from the noisy signals about the outside world, one should furthermore expect such larger coding noise to translate into an increase in the prior variance.[13] Given that both coding noise and the prior

[12]Error models do not make any predictions about systematic effects either. It is of course conceivable that the magnitude of the error could depend on the scale on which time delays are expressed. Note, however, that models such as proposed by Lu and Saito (2018) and He et al. (2019) would then predict differences in *decreasing impatience*—a prediction that is quite distinct from those emerging from the NCT model.

[13]Assume that subjects learn the stimuli of the environment from noisy signals. If the noisiness of the signals will be impacted, this ought to result in an increase of the variance, given that the shape parameter of the conjugate Normal-Gamma prior will be augmented with two parts—one given by the difference in posterior inference from the prior mean, and one given by the residual noise. See Vieider (2023) for a formalization of such a learning model.

variance are predicted to increase, the effect on discriminability will be ambiguous, since it depends on the relative increase in coding noise and variance. One should, however, expect an increase in the impatience parameter α when delays are expressed in days rather than weeks. This is predicted to happen because increased coding noise will directly impact learning of the prior mean, with noisy signals of large delays leading to an upward distortion of the mean. An interesting question further concerns the effects on present-bias. If the value of κ is expressed in the units of the time delays, then we would expect less present-bias in the *days* condition. If the value of κ is driven by inattention, as discussed above, then we should not expect any systematic effects.[14]

I ran the experiment on Prolific UK, with a sample size of 300. Subjects were paid a fixed fee for their time, and made hypothetical choices between options like the ones in experiment I. There is no evidence that using hypothetical instead of real choices impacts observed discounting behaviour (Cohen et al., 2020). The stimuli closely resemble those used in experiment I, except for i) the already-mentioned treatment variation between weeks and days, which was implemented between subjects; ii) that all time delays were doubled (i.e., the shortest delay was 12 weeks, while the longest delay was 48 weeks); and iii) the list of sooner amounts was extended down to £25, to have more room to detect impatience.[15]

4.2 Results

Figure 5 compares the nonparametric discount functions across the two treatment conditions, with panel A showing the functions based on delays from the present, and panel B showing the functions based on the four 12-week delays. In both cases, the discount function in the *days* conditions stays clearly below the discount

[14] A pre-analysis plan for the experiment discussing these hypotheses was registered at the Open Science Foundation under number osf.io/3cjy7 prior to running the experiment. The predictions and analysis below closely track this pre-analysis plan, unless stated otherwise.

[15] In order to avoid that the experiment got too long, each subject was only shown 158 randomly selected choices from the total to keep the experiment of identical length as experiment I. The exact choices being shown were randomly extracted at the individual level, so that the results reported below are unaffected by this. Simulations showed that such a design could yield superior results compared to the choices used in experiment I in hierarchical estimations in the presence of very impatient individuals.

function in the *weeks* condition (the difference for the function based on the two 24-week delays, which is not shown, is similar). This difference is indeed sizeable.

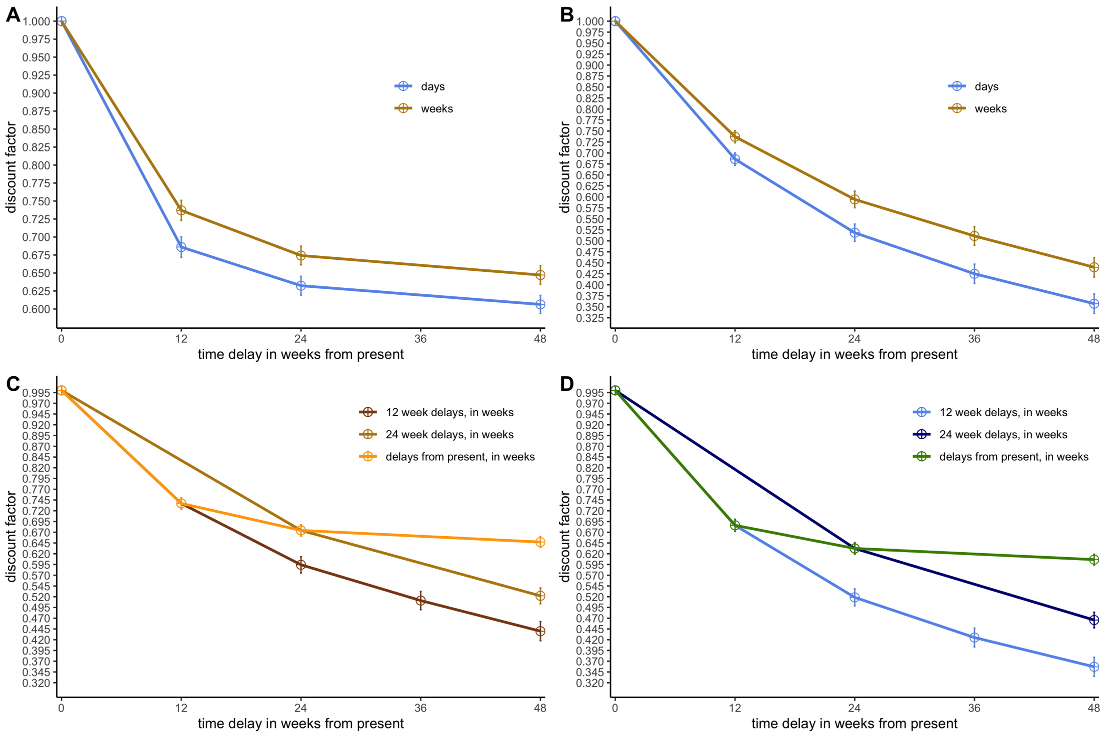


Figure 5: Discount funcions for days versus weeks

Nonparametric discount functions when time delays are expressed in weeks versus days. Panel A shows the functions based on delays of different lengths from the present. Panel B shows the functions calculated based on the four underlying 12-week delays from different points. Panel C shows discount functions obtained from different time delays in the *weeks* condition, whereas panel D shows the same type of functions for the *days* condition.

Panel C further shows the comparison of different discount functions for the *weeks* condition, and panel D shows the same comparison for the *days* condition. The stylized patterns predicted by the NCT model are the same across conditions, and correspond closely to those documented in experiment I. In particular, there is again clear evidence for present-bias in both the *days* and *weeks* condition, substantial delay-dependence in both conditions, and no evidence for strongly decreasing impatience in either condition (see online appendix H for additional graphs). The differences between the different functions obtained based on different delays, however, appears more pronounced in the *days* condition.

I next proceed to estimating the NCT model parameters. Figure 6 compares the estimates obtained from a Bayesian random-parameter model by treatment

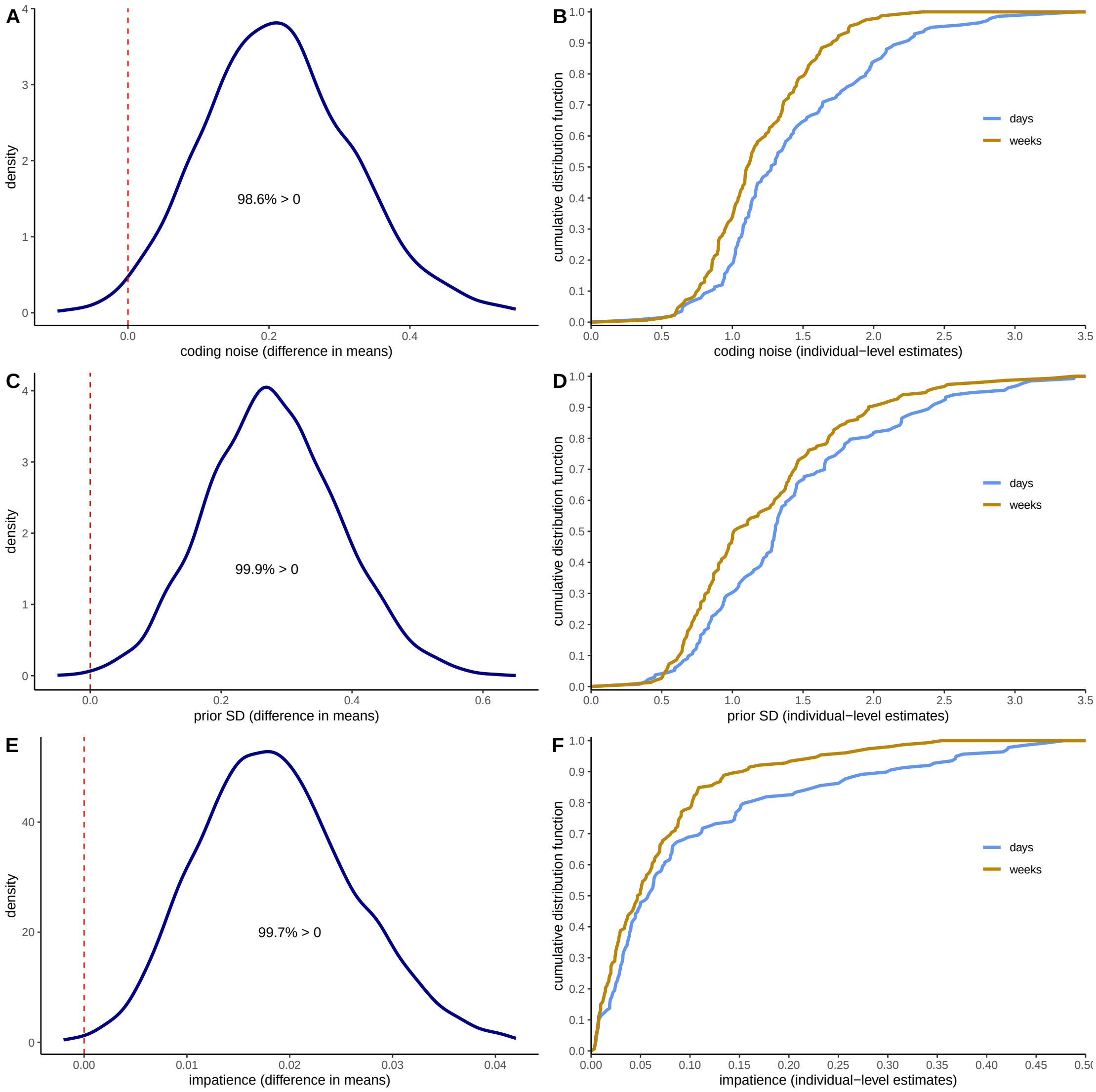


Figure 6: Density plots of NCT parameters ν, σ, and α, days versus weeks

NCT model parameters ν, σ, and α for *days* versus *weeks*. Panel A shows the difference in mean estimates of the coding parameters, as $\nu_{days} - \nu_{weeks}$, and indicates a significantly larger mean in the *days* condition. Panel B shows the cumulative distribution of individual-level parameter estimates of the coding noise. Panel C shows the difference in means of the prior SD, $\sigma_{days} - \sigma_{weeks}$, which is again significantly larger in the *days* condition. Panel D shows the cumulative distribution of individual-level parameters. Panel E shows the difference in the posterior draws describing the means of the impatience parameter α, $\alpha_{days} - \alpha_{weeks}$, which again supports the initial hypothesis. Panel F shows the cumulative distribution of individual-level impatience parameters.

condition (see online appendix E for a description of the econometrics).[16] Panel A shows the difference in the posterior draws for the mean coding noise parameter, ν. 98.6% of the posterior probability mass indicates that the average coding noise is larger in the *days* condition compared to the *weeks* condition. The individual-level

[16]The hyperpriors were chosen to be mildly regularizing, and changing the priors does not affect the results in any way. Importantly, I used hyperpriors that are *identical across treatment conditions*, so that they could not possibly bias the results.

estimates, shown in panel B, paint a similar picture. The cumulative distribution of coding noise parameters for *days* is clearly shifted to the right relative to the distribution for *weeks*, a difference that is significant according to a Wilcoxon ranksum test on the individual-level parameter estimates ($p < 0.001$).

A similar effect occurs for the standard deviation of the prior, σ, which is larger for *days* than *weeks* both based on the difference in posterior draws of the mean (panel C; 99.9% of the probability mass of the difference is positive), and according to a ranksum test of individual-level parameters (panel D; $p = 0.001$). These results thus coincide with the predicted effects. Finally, panel E shows the difference in posterior draws of the mean impatience parameter, α. Once again, the draws for *days* significantly exceed those for *weeks*, with a probability mass of 99.7% in favour of the hypothesized effect. Panel F shows the difference in individual-level parameters. The cumulative distribution for the *days* condition stays consistently below the one for the *weeks* condition, an effect that is again significant according to a ranksum test on the estimated parameters ($p = 0.013$). This difference is indeed what drives the difference in nonparametric discount functions shown in figure 5.

As a next step, we can examine the model parameters β and κ, for which the predictions were more ambivalent. Figure 7 indicates the effects for discriminability β and for κ. Discriminability increases somewhat. However, only 76.7% of the probability mass of the difference in posterior draws indicates a value that is larger for *days* than for *weeks* (panel A), thus falling short of conventional levels used to indicate statistical significance. A similar picture emerges from a test on the individual-level estimates ($p = 0.374$; panel B). We also observe no differences for κ in the aggregate test shown in panel C. A similar conclusion follows from a test on the individual-level parameters shown in panel D ($p = 0.102$).[17] It thus appears that κ is not expressed on the scale of the delays displayed in the experiment, suggesting instead that κ may be driven by variations in attention.

[17] One may argue that the variable of import in this case is not κ, but rather κ^{β}. Given that β does also not change significantly between treatments, however, tests based on this derived variable yield similar results.

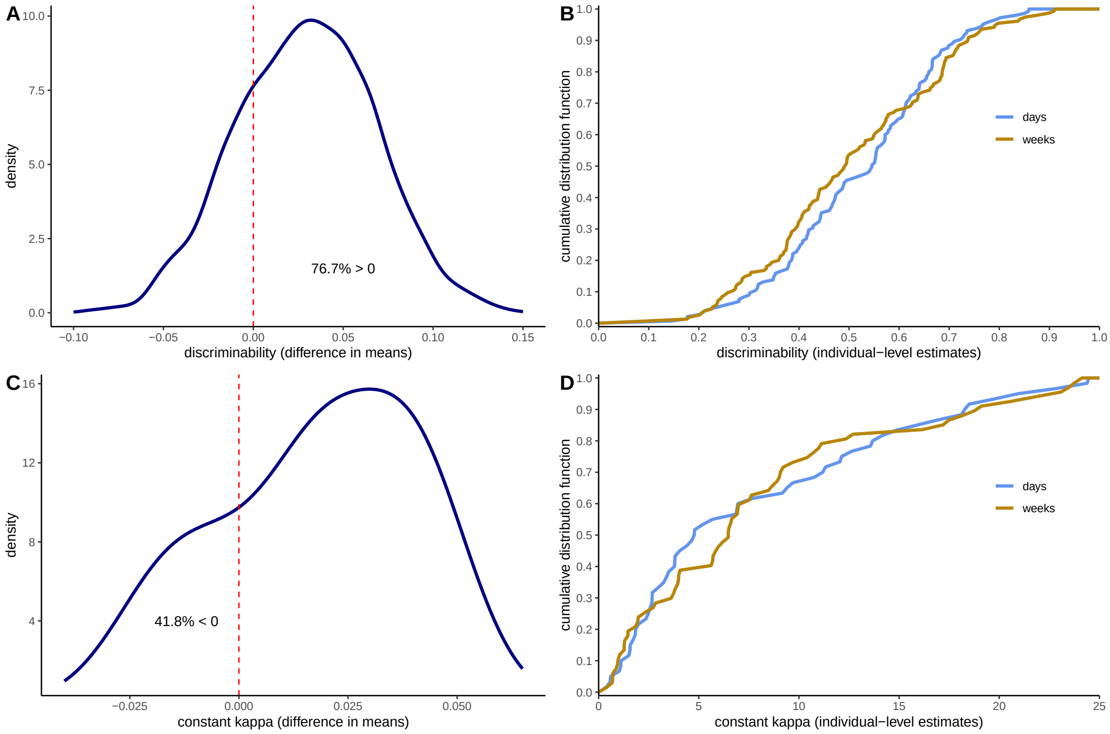


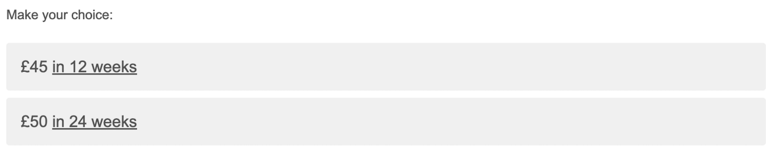
Figure 7: Density plots of NCT parameters β and κ, days versus weeks

NCT model parameters β and κ for *days* versus *weeks*. Panel A plots the difference in posterior draws of the discriminability parameters between conditions,$\beta_{days} - \beta_{weeks}$, and shows 76.7% of the probability mass in favour of the parameter being larger for *days*. Panel B plots the cumulative distributions of individual-level discriminability parameters. Panel C plots the difference of the posterior draws for the mean for κ, $\kappa_{days} - \kappa_{weeks}$. Only 41.8% of the probability mass indicates a parameter that is smaller in the *days* condition. Panel D draws the cumulative distribution of the individual-level κ parameters by treatment condition.

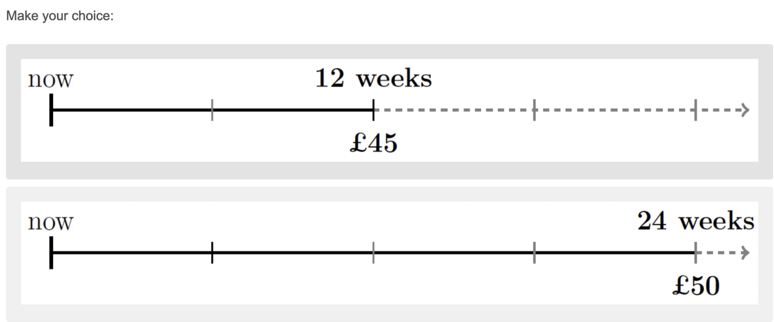
5 Experiment III: Visual Aid for Time Delays

5.1 Experiment

Experiment II showed that making quick comparisons more difficult by expressing time delays in days rather than weeks impacted coding noise, the prior variance, and impatience. Present-bias, however, was unaffected by the manipulation of the unit in which the time delay was expressed. Experiment III aims to directly test the *attentional* interpretation I proposed of the model. In particular, I randomize whether subjects see time delays in weeks described by means of textual statements like in experiments I and II, or whether they are given a visual aid representing time delays by arrows with length proportional to the delays themselves. This manipulation is meant to specifically focusing attention onto the time dimension.[18] The visual aid is shown in figure 8. In the control or *textual* condition, time delays are described verbally, as done previously. In the *visual* condition, time delays are depicted by means of arrows which are proportional in length to the time delay between now and the moment the reward is paid out. I ran the experiment with 301 subjects on Prolific using the stimuli of experiment I.[19] Choices were hypothetical like in experiment II. I excluded 3 subjects from the analysis who constitute extreme outliers caused by largely random response patterns. Keeping these subjects does not substantially affect the conclusions drawn below.



(a) Textual condition



(b) Visual condition

Figure 8: Choise situation for textual and visual condition

Panel 8(a) shows a choice situation in the *textual condition*. Panel 8(b) shows the equivalent choice situation in the *visual* condition. Subjects could indicate their choice by clicking on their preferred option, and subsequently moving on the the following screen.

A key prediction is that the visual aid for time will decrease coding noise, delivering an inverse manipulation to the one shown in experiment II. Importantly, however, by focusing attention on the time dimension, the visual aid manipulation

[18] Ebert and Prelec (2007) also propose a manipulation using a visual aid. Their exclusive focus on constant sensitivity discounting in delays from the present, however, means that they cannot disentangle the different motives we are interested in here. In particular, they cannot disentangle strongly decreasing impatience from delay-dependence and present-bias.

[19] Using the stimuli of experiment I instead of experiment II allowed me to have the sooner amounts vary between 33 and 49, and to present all binary tradeooffs to all subjects, since these ranges are sufficient to measure the discoutning for the shorter time delays.

is predicted to also decrease present-bias. This prediction arises from the observation that the visual aid does not only simplify the approximate appraisal and comparison of time delays, but that it also impacts attention paid to the time delays themselves. If the attentional interpretation of κ given above is warranted, we should thus expect to see a decrease in present-bias. Predictions on other parameters are less crisp. In particular, changes in discriminability β will depend on the extent to which the prior variance σ decreases following the predicted decrease in coding noise ν. Any decrease in σ may well be weaker than the increase in experiment II due to floor effects, but this is largely an empirical question.

5.2 Results

I start from examining the most distinctive prediction arising from the increased attention to the time delays hypothesized in the visual condition—the decrease in present-bias. Figure 9 shows decumulative choice proportions of the larger-later reward as the smaller-sooner reward increases. Panel A shows the choice proportions in the textual condition, comparing a 6-week delay from the present to a 6-week delay from 6 weeks. Panel C shows the equivalent choice proportions in the textual treatment for a 12-week delay from the present, and compares it to choice proportions for a 12-week delay from 12 weeks. In both cases, there is clear evidence for present-bias, much like in experiments I and II ($p < 0.001$ in both cases; Wilcoxon signed-rank test on individual choice proportions). Panels B and D show the equivalent figures for the visual treatment. The choice proportions between the delays from the present, and the delays from the upfront delay are now indistinguishable ($p > 0.45$ in both cases). One can furthermore see that the difference is closed from the side of the immediate delays, i.e. we clearly witness *a reduction in present-bias*, rather than an overall increase in impatience.

I next proceed to estimating the NCT model parameters. Figure 10 shows a comparison of the parameters ν, σ, and β across treatment conditions. Panel A shows the difference in the posterior draws of the mean for the coding parameter ν, $\nu_v - \nu_t$, where v stands for 'visual' and t for 'textual'. 98.6% of the draws indicate a

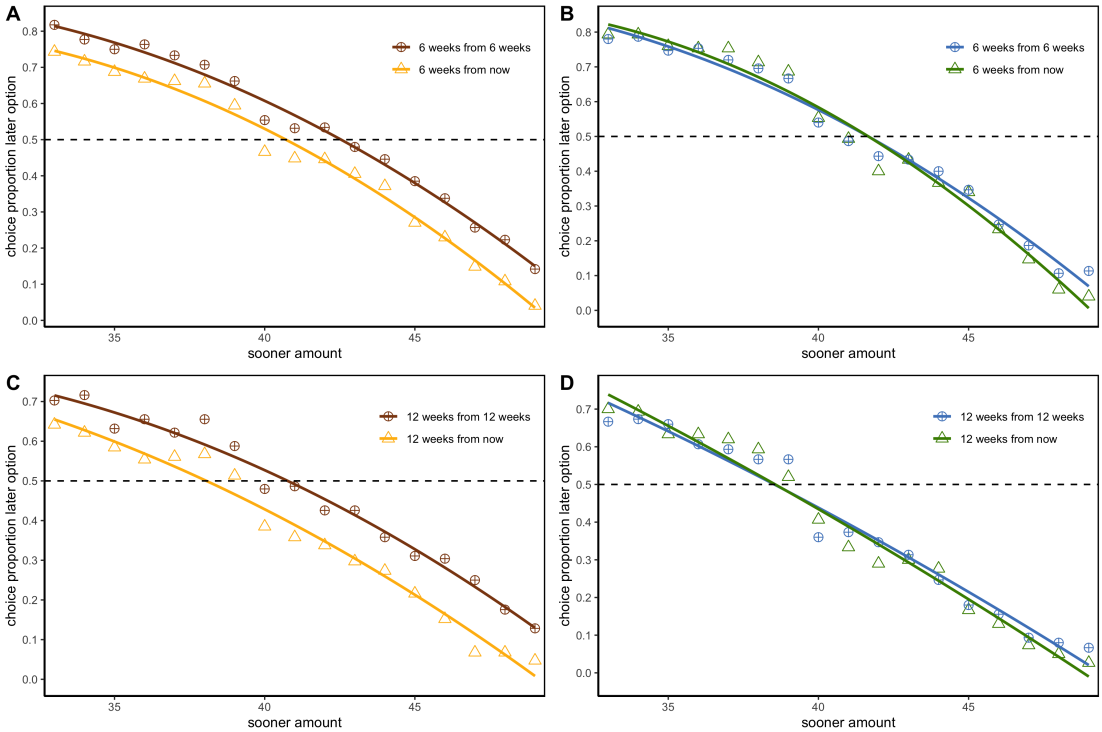


Figure 9: Decumulative choice porportions for larger-later option, textual vs visual

Decumulative choice proportions of the larger-later reward as the smaller-sooner reward increases from £33 through £49. Curves fit to the choice proportions are based on second degree polynomials. Panel A compares the choice proportions for a 6-week delay from the present, to those for a 6-week delay from 6 weeks. The curve shifts to the northeast as the upfront delay is introduced, indicating present-bias. Panel B shows the equivalent comparison for the visual condition. No shift is now apparent, and the two curves are virtually indistinguishable. Panel C shows a similar effect as in panel A for a 12-week delay from the present, versus a 12-week delay from 12 weeks in the textual condition. Panel D shows the same comparison for the visual treatment, and again fails to replicate the distinct present-bias pattern in panel C.

lower coding noise in the visual condition. The individual-level parameters, shown in panel B, paint a similar picture. Coding noise parameters in the visual condition are clearly shifted to the left of those in the textual treatment ($p < 0.001$, signed rank test). Panels C and D show the equivalent results for the standard deviation of the prior, σ. The effect goes in the direction of a smaller standard deviation in the visual condition, but fails to reach significance both in the aggregate (79.8% of the probability), and for the individual-level parameters ($p = 0.713$).

Panel E shows the difference in posterior draws of the means for the discriminability parameter β. 90.3% of the probability mass indicates that discriminability is larger in the visual condition compared to the textual treatment. Panel F shows the empirical cumulative distribution function of the individual-level parameters. The distribution for the visual condition is clearly to the right of the one

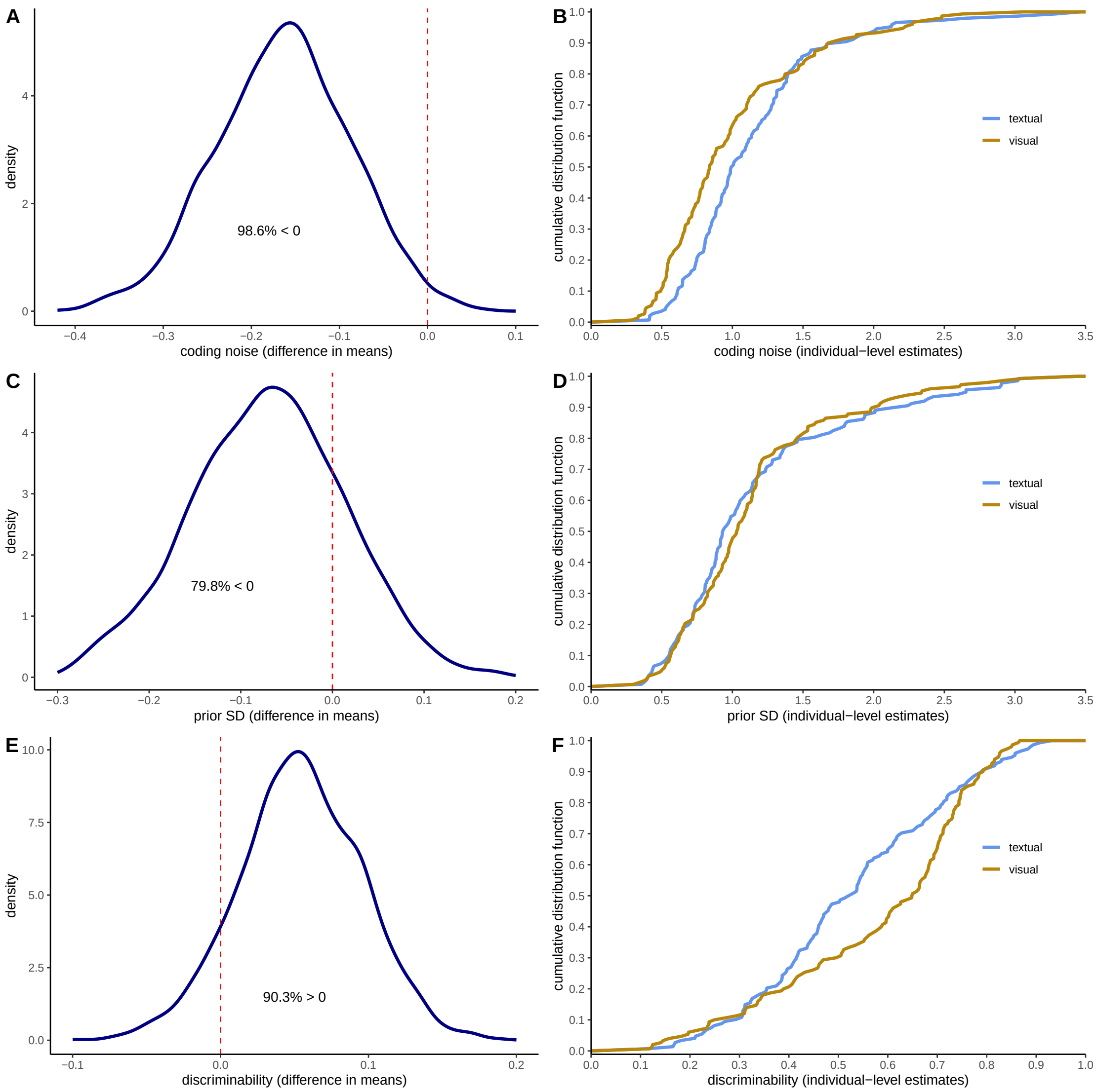


Figure 10: Comparison of model parameters ν, σ, and β

Panel A shows differences in posterior draws of the means for coding noise ν, and panel B shows difference in individual-level parameters estimates in a hierarchical model. Both figures show clear evidence that coding noise is smaller in the visual condition compared to the textual condition. Panel C and D indicate no clear difference in the prior SD σ. Panels E and F show evidence that discriminability β is larger in the visual treatment compared to the textual treatment. The less pronounced effect on the average mean occurs because of outliers.

for the textual treatment. This effect is indeed highly significant ($p = 0.007$).[20] Overall, we thus find clear evidence indicating that coding noise is reduced, while discriminability is increased in the visual condition. These results are consistent with the increase in attention to the time dimension posited by the NCT model.

Figure 11 shows the equivalent distributions for impatience α, and for the computational parameter κ. As expected, there are no differences in impatience

[20] The difference in the strngth of the evidence between the aggregate and individual-level tests is explained by a few outliers, which disproportionately influence the aggregate esimates.

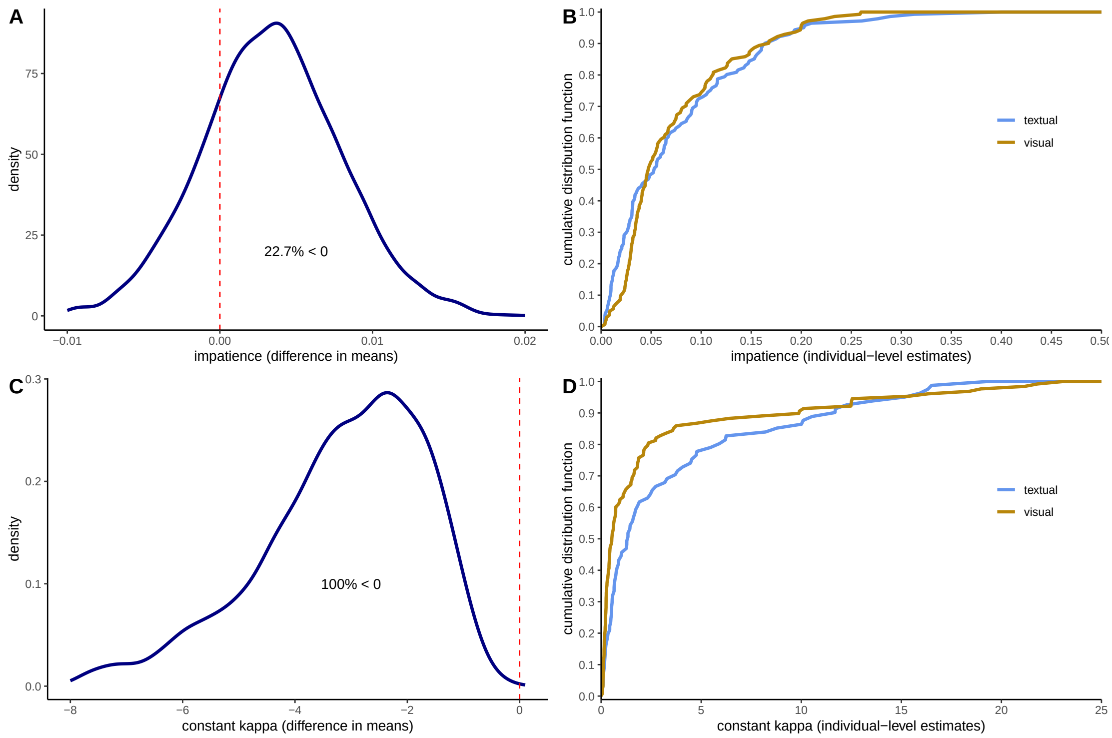


Figure 11: Comparison of model parameters α and κ

Panel A shows the difference in posterior draws of the mean for the impatience parameter α, and Panel B shows the empirical cumulative distribution functions of the individual-level estimates. There is no difference between treatments, as predicted. Panel C shows the difference in posterior draws of the means for κ, and Panel D shows the differences in individual-level parameters. The differences are pronounced, showing a clear reduction in present-bias in the visual condition compared to the textual condition.

α at either the aggregate or the individual level ($p = 0.375$). We do, however, find clear differences in κ. The aggregate-level evidence in panel C indicates that 100% of the difference in posterior draws of the means are negative, thus indicating that the average parameter is significantly smaller in the visual condition. A similar picture emerges from the individual-level estimates, where the parameters estimated from the textual condition clearly exceed those estimated from the visual condition ($p < 0.001$). There is thus clear evidence that providing a visual aid focuses attention on the time dimension, thereby reducing present-bias.

6 Discussion and Conclusion

The noisy coding of time model suggests that choices between rewards obtaining at different time periods may be subject to systematic manipulation. This was showcased in experiment II, where expressing identical time delays in days rather than weeks systematically increased impatience in a way predicted by the model. Experiment III added to that evidence by showcasing how the provision of visual aids for the lengths of the time delays drastically reduced present-bias. Further evidence comes from a series of experiments executed by Ebert and Prelec (2007). Using delays of different lengths from the present, the latter showed that time pressure decreased sensitivity to time delays, whereas the provision of visual clues summarizing the time delays increased sensitivity to the same time delays. Seen through the lens of the noisy coding of time model, these results further illustrate how the model parameters can be systematically manipulated.

The model predictions and experimental findings described above are consistent with a number of stylized facts in the reward discounting literature. For instance, there is a highly consistent body of evidence showing that discount functions show a hyperbolic shape when measured using time delays of different length from the present (Thaler, 1981; Kirby and Maraković, 1995; Ebert and Prelec, 2007; Zauberman et al., 2009). It has long been known that discounting increases as time delays are subdivided into smaller intervals, an effect that has been referred to as *subadditive* discounting (Read, 2001). This effect has indeed been shown to be strong and pervasive, even though it is often ignored in investigations of discounting (Dohmen et al., 2017). Kable and Glimcher (2010) have documented similar levels of delay-dependence independently of the upfront delay. Read, Frederick, Orsel and Rahman (2005) have documented a 'date/delay effect', whereby discounting weakens substantially when delayed payoffs are associated with calendar dates rather than with time delays. This effect is highly consistent with the model I present here, since in the absence of any delays to be noisily encoded, we would expect the effect of noisy delay perception to disappear.

The underlying working mechanism of the noisy coding of time model relies on logarithmic transformations and Bayesian updating. It is important to emphasize that there is no notion here that humans consciously execute such complex mathematical operations. The logarithmic coding of numbers has been shown to be physiologically hardwired in the brain (Dehaene, 2003), possibly because of its optimality for adaptation (Howard and Shankar, 2018). The Bayesian combination of likelihood and prior, furthermore, takes the form of a simple linear combination of two variables. Indeed, the success of the Bayesian Brain model in neuroscience derives in no small part from the observation that it can easily be implemented in biologically realistic neural network structures.

The noisy coding of time model I have presented makes stochastic predictions on choice behaviour in tradeoffs between smaller-sooner and larger-later amounts based on the noisy perception of *time delays*. A related literature has studied the effects of randomness in choices and/or preferences on inter-temporal discounting (Lu and Saito, 2018; He et al., 2019). These papers have highlighted that particular patterns of randomness in responses or preferences may result in hyperbolic as-if discounting. The predictions emerging from these two classes of models are quite distinct. While the noisy coding model predicts delay-dependence and present-bias based on different mechanisms, these error models predict decreasing impatience by postulating randomness in choices or preferences. The predictions of these models are thus similar to those of the hyperbolic discounting literature. Note furthermore that the underlying mechanism is also different. While typical error models either model utility as being implemented with noise or add random fluctuation to preference parameters, the NCT model sees time itself as being noisily perceived.

None of the experiments I presented showed any evidence for strongly decreasing impatience—a finding that is in line with much of the recent literature (see appendix D for a review). Nevertheless, an unresolved question concerns why strongly decreasing impatience is at least *sometimes* observed in experiments using monetary payoffs. A recent literature has attributed seemingly hyperbolic patterns to the inherent uncertainty of the future (Halevy, 2008; Chakraborty et al., 2020; Epper and Fehr-Duda, 2023). If a decision maker has doubts about the delivery of future payouts, and if this decision maker exhibits the certainty effect, then they will exhibit seemingly hyperbolic patterns of discounting. I have been particularly careful to eliminate future uncertainty by having experiment I conducted by a trusted person (the lecturer) and in an environment where interactions are repeated and ongoing, or by using a hypothetical setup in experiment II and III. That being said, future uncertainty could be easily incorporated into the model by merging it with the model of Vieider (2021), where I use a formally identical setup to model likelihood-distortions under uncertainty.

Decision making under risk and over time have been depicted as closely related (Prelec and Loewenstein, 1991). A particular relation between the two approaches is between probability distortions (probabilistic insensitivity) and time distortions (time insensitivity). In the noisy coding of time model I presented, time insensitivity can be traced back to noise in the encoding of time delays. In a twin paper on decision making under uncertainty (Vieider, 2021), I show that likelihood-sensitivity can be traced back to noise in the coding of likelihood ratios in binary wagers using a formally equivalent setup. This is consistent with the results of Epper, Fehr-Duda and Bruhin (2011), who documented a correlation between the two phenomena. Joint investigations of decision making under risk and over time in a noisy coding setup thus hold the promise of further illuminating the common underlying drivers of choice behaviour.

References

Attema, Arthur E, Han Bleichrodt, Kirsten IM Rohde, and Peter P Wakker (2010) 'Time-tradeoff sequences for analyzing discounting and time inconsistency.' *Management Science* 56(11), 2015–2030

Baillon, Aurélien, Han Bleichrodt, and Vitalie Spinu (2020) 'Searching for the reference point.' *Management Science* 66(1), 93–112

Bettinger, Eric, and Robert Slonim (2007) 'Patience among children.' *Journal of Public Economics* 91(1-2), 343–363

Bleichrodt, Han, Kirsten I. M. Rohde, and Peter P. Wakker (2009) 'Non-hyperbolic time inconsistency.' *Games and Economic Behavior* 66(1), 27–38

Bleichrodt, Han, Yu Gao, and Kirsten IM Rohde (2016) 'A measurement of decreasing impatience for health and money.' *Journal of Risk and Uncertainty* 52(3), 213–231

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper (2010) 'Risk and Rationality: Uncovering Heterogeneity in Probability Distortion.' *Econometrica* 78(4), 1375–1412

Carpenter, Bob, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell (2017) 'Stan: A probabilistic programming language.' *Journal of Statistical Software* 76(1), 1–32

Cavagnaro, Daniel R, Gabriel J Aranovich, Samuel M McClure, Mark A Pitt, and Jay I Myung (2016) 'On the functional form of temporal discounting: An optimized adaptive test.' *Journal of Risk and Uncertainty* 52(3), 233–254

Chakraborty, Anujit, Yoram Halevy, and Kota Saito (2020) 'The relation between behavior under risk and over time.' *American Economic Review: Insights* 2(1), 1–16

Cohen, Jonathan, Keith Marzilli Ericson, David Laibson, and John Myles White (2020) 'Measuring time preferences.' *Journal of Economic Literature* 58(2), 299–347

Conte, Anna, John D. Hey, and Peter G. Moffatt (2011) 'Mixture models of choice under risk.' *Journal of Econometrics* 162(1), 79–88

Cubitt, Robin P, and Daniel Read (2007) 'Can intertemporal choice experiments elicit time preferences for consumption?' *Experimental Economics* 10(4), 369–389

Dayan, Peter, and Laurence F Abbott (2001) *Theoretical neuroscience: computational and mathematical modeling of neural systems* (Computational Neuroscience Series)

Dehaene, Stanislas (2003) 'The neural basis of the weber–fechner law: a logarithmic mental number line.' *Trends in cognitive sciences* 7(4), 145–147

Dehaene, Stanislas, and Jean-Pierre Changeux (1993) 'Development of elementary numerical abilities: A neuronal model.' *Journal of cognitive neuroscience*

5(4), 390–407

Doepke, Matthias, and Fabrizio Zilibotti (2014) 'Culture, Entrepreneurship, and Growth.' In 'Handbook of Economic Growth,' vol. 2

_ (2017) 'Parenting with style: Altruism and paternalism in intergenerational preference transmission.' *Econometrica* 85(5), 1331–1371

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2017) 'The robustness and pervasiveness of sub-additivity in intertemporal choice.' In 'Working Paper'

Ebert, Jane E. J., and Drazen Prelec (2007) 'The Fragility of Time: Time-Insensitivity and Valuation of the Near and Far Future.' *Management Science* 53(9), 1423–1438

Enke, Benjamin, and Thomas Graeber (2019) 'Cognitive uncertainty.' Technical Report, National Bureau of Economic Research

Enke, Benjamin, Thomas Graeber, and Ryan Oprea (2023) 'Complexity and time.' Technical Report, Mimeo

Epper, Thomas, and Helga Fehr-Duda (2018) 'Risk in time: The intertwined nature of risk taking and time discounting.' *Working Paper*

_ (2023) 'Risk in time: The intertwined nature of risk taking and time discounting.' *Journal of the European Economic Association, forthcoming*

Epper, Thomas, Helga Fehr-Duda, and Adrian Bruhin (2011) 'Viewing the future through a warped lens: Why uncertainty generates hyperbolic discounting.' *Journal of Risk and Uncertainty* 43(3), 169–203

Frederick, Shane, George Loewenstein, and Ted O'Donoghue (2002) 'Time Discounting and Time Preference: A Critical Review.' *Journal of Economic Literature* 40(2), 351–401

Gabaix, Xavier, and David Laibson (2017) 'Myopia and discounting.' Technical Report, National bureau of economic research

Gelman, Andrew, Jessica Hwang, and Aki Vehtari (2014) 'Understanding predictive information criteria for bayesian models.' *Statistics and computing* 24(6), 997–1016

Gold, Joshua I, and Michael N Shadlen (2001) 'Neural computations that underlie decisions about sensory stimuli.' *Trends in cognitive sciences* 5(1), 10–16

Gul, Faruk, and Wolfgang Pesendorfer (2001) 'Temptation and self-control.' *Econometrica* 69(6), 1403–1435

Halevy, Yoram (2008) 'Strotz Meets Allais: Diminishing Impatience and the Certainty Effect.' *The American Economic Review* 98(3), 1145–1162

He, Lisheng, Russell Golman, and Sudeep Bhatia (2019) 'Variable time preference.' *Cognitive Psychology* 111, 53–79

Hey, John D., and Chris Orme (1994) 'Investigating Generalizations of Expected Utility Theory Using Experimental Data.' *Econometrica* 62(6), 1291–1326

Howard, Marc W, and Karthik H Shankar (2018) 'Neural scaling laws for an uncertain world.' *Psychological review* 125(1), 47

Imai, Taisuke, Tom A Rutter, and Colin F Camerer (2021) 'Meta-analysis of present-bias estimation using convex time budgets.' *The Economic Journal* 131(636), 1788–1814

Kable, Joseph W, and Paul W Glimcher (2010) 'An 'as soon as possible' effect in human intertemporal decision making: behavioral evidence and neural mechanisms.' *Journal of neurophysiology* 103(5), 2513–2531

Khaw, Mel Win, Ziang Li, and Michael Woodford (2021) 'Cognitive imprecision and small-stakes risk aversion.' *The Review of Economic Studies* 88(4), 1979–2013

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

Kirby, Kris N., and Nino N. Maraković (1995) 'Modeling Myopic Decisions: Evidence for Hyperbolic Delay-Discounting within Subjects and Amounts.' *Organizational Behavior and Human Decision Processes* 64(1), 22–30

Laibson, David (1997) 'Golden Eggs and Hyperbolic Discounting.' *The Quarterly Journal of Economics* 112(2), 443–478

Lim, Seung-Lark, John P O'Doherty, and Antonio Rangel (2011) 'The decision value computations in the vmpfc and striatum use a relative value code that is guided by visual attention.' *Journal of Neuroscience* 31(37), 13214–13223

Loewenstein, George, and Drazen Prelec (1992) 'Anomalies in Intertemporal Choice: Evidence and an Interpretation.' *The Quarterly Journal of Economics* 107(2), 573–597

Lu, Jay, and Kota Saito (2018) 'Random intertemporal choice.' *Journal of Economic Theory* 177, 780–815

Mazur, James E (1987) 'An adjusting procedure for studying delayed reinforcement.' *Quantitative analyses of behavior* 5, 55–73

McElreath, Richard (2016) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Academic Press)

Mischel, Walter, and Ebbe B Ebbesen (1970) 'Attention in delay of gratification.' *Journal of Personality and Social Psychology* 16(2), 329

Mischel, Walter, Yuichi Shoda, and Monica L Rodriguez (1989) 'Delay of gratification in children.' *Science* 244(4907), 933–938

Natenzon, Paulo (2019) 'Random choice and learning.' *Journal of Political Economy* 127(1), 419–457

Netzer, Nick (2009) 'Evolution of time preferences and attitudes toward risk.' *American Economic Review* 99(3), 937–55

Netzer, Nick, Arthur Robson, Jakub Steiner, and Pavel Kocourek (2021) 'Endogenous risk attitudes.' Working Paper

Nieder, Andreas, and Earl K Miller (2003) 'Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex.' *Neuron* 37(1), 149–157

Oprea, Ryan (2022) 'Simplicity equivalents.' *Working Paper*

Phelps, E. S., and R. A. Pollak (1968) 'On Second-Best National Saving and Game-Equilibrium Growth.' *The Review of Economic Studies* 35(2), 185–199

Prelec, Drazen (2004) 'Decreasing Impatience: A Criterion for Non-stationary Time Preference and "Hyperbolic" Discounting.' *Scandinavian Journal of Economics* 106(3), 511–532

Prelec, Drazen, and George Loewenstein (1991) 'Decision Making over Time and

Under Uncertainty: A Common Approach.' *Management Science* 37(7), 770–786

Read, Daniel (2001) 'Is Time-Discounting Hyperbolic or Subadditive?' *Journal of Risk and Uncertainty* 23(1), 5–32

Read, Daniel, and Peter HMP Roelofsma (2003) 'Subadditive versus hyperbolic discounting: A comparison of choice and matching.' *Organizational behavior and human decision processes* 91(2), 140–153

Read, Daniel, Shane Frederick, Burcu Orsel, and Juwaria Rahman (2005) 'Four score and seven years from now: The date/delay effect in temporal discounting.' *Management Science* 51(9), 1326–1335

Regenwetter, Michel, Daniel R Cavagnaro, Anna Popova, Ying Guo, Chris Zwilling, Shiau Hong Lim, and Jeffrey R Stevens (2018) 'Heterogeneity and parsimony in intertemporal choice.' *Decision* 5(2), 63

Robson, Arthur J (2001) 'The biological basis of economic behavior.' *Journal of Economic Literature* 39(1), 11–33

Rohde, Kirsten IM (2019) 'Measuring decreasing and increasing impatience.' *Management Science* 65(4), 1700–1716

Samuelson, Paul A. (1937) 'A Note on Measurement of Utility.' *The Review of Economic Studies* 4(2), 155–161

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

Stan Development Team (2017) 'RStan: the R interface to Stan.' R package version 2.17.2

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

Vehtari, Aki, Andrew Gelman, and Jonah Gabry (2017) 'Practical bayesian model evaluation using leave-one-out cross-validation and waic.' *Statistics and computing* 27(5), 1413–1432

Vieider, Ferdinand M. (2021) 'Decisions under uncertainty as bayesian inference.'

Working Paper WP 21/1022 of Faculty of Economics and Business Administration, Ghent University, Belgium 21/1022, Ghent University, Faculty of Economics and Business Administration

— (2023) 'A dynamic model of predictive stimulus coding.' *Mimeo*

von Gaudecker, Hans-Martin, Arthur van Soest, and Erik Wengström (2011) 'Heterogeneity in Risky Choice Behaviour in a Broad Population.' *American Economic Review* 101(2), 664–694

Whalen, John, Charles R Gallistel, and Rochel Gelman (1999) 'Nonverbal counting in humans: The psychophysics of number representation.' *Psychological science* 10(2), 130–137

Zauberman, Gal, B Kyu Kim, Selin A Malkoc, and James R Bettman (2009) 'Discounting time and time discounting: Subjective time perception and intertemporal preferences.' *Journal of Marketing Research* 46(4), 543–556

ONLINE APPENDIX

A Model derivation

To obtain actionable choice quantities, we need to combine the likelihood and prior by Bayesian updating. Since r is a scalar, the posterior probability density function of the mental time delay on a logarithmic scale will be:

$$p\left[ln(t)|r\right] = \mathcal{N}\left(\frac{\sigma^2}{\sigma^2+\nu^2} \times r + \frac{\nu^2}{\sigma^2+\nu^2} \times \mu \ , \ \frac{\nu^2\sigma^2}{\nu^2+\sigma^2}\right), \tag{9}$$

where p indicates the probability density, and we can define $\beta \triangleq \frac{\sigma^2}{\sigma^2+\nu^2}$, so that $(1-\beta) \triangleq \frac{\nu^2}{\sigma^2+\nu^2}$. The parameter β thus gives us the weight attributed to the likelihood relative to the prior mean, μ. It depends on the relative uncertainty associated to the mental signal versus the mental prior. This gives us the expectation of the posterior defined over the logarithm of t:

$$E[ln(t)|r] = \beta r + (1-\beta)\mu.$$

Defining $\alpha \triangleq e^{(1-\beta)\mu}$ yields the posterior expectation of t used in the main text. Plugging this expression back into the mental choice rule in equation 2 and solving for r yields the threshold equation 6.

We then exploit the known distributional properties of the mental signal r. Taking the z-score of the likelihood by subtracting the mean and dividing by the standard deviation, we obtain $z = \frac{r-ln(t)}{\nu}$, or equivalently, $z = \frac{\beta r - \beta ln(t)}{\beta\nu}$. We can then compare this z-score to an equivalent z-score obtained from the threshold equation, $z_t = \frac{\beta r - \left[ln(-ln(\frac{y}{x})) - ln(\alpha)\right]}{\beta\nu}$. Subtracting the former from the latter gives us the probabilistic choice rule in equation 7. Given that the difference of two standard normal variables will itself follow a standard normal distribution, we can interpret the derived equation directly as a Probit choice rule.

B Choice rule on orignal scale

In my preferred specification, I have assumed that the choice rule is transformed from its original scale by taking the log twice. This does in no way affect the conclusions. If we work on the original scale, the optimal choice rule is $exp(-t)x > y$. We now derive the posterior expectation of the mental time delay t on the logarithmic scale just like above. To obtain the posterior expectation of the time delay t itself, we exploit the properties of the log-normal distribution, which has a mean $\mu + \frac{1}{2}\sigma_p^2$, where $\sigma_p^2 \triangleq \frac{\nu^2\sigma^2}{\nu^2+\sigma^2}$ is the posterior variance. We thus obtain

$$E[t|r] = exp\left(\beta r + (1-\beta)\mu + \frac{1}{2}\sigma_p^2\right),$$

or alternatively, $E[t|r] = exp\left(\beta r + (1-\beta)(\mu + \frac{1}{2}\sigma^2)\right)$. Substituting this mental quantity for t in the choice rule and rearranging, we get $exp\left(-exp(\beta r + ln(\alpha))\right) > \frac{y}{x}$. Taking the logarithm of both sides, multiplying by -1, taking the logarithm again, and rearranging we once again obtain the threshold equation $r < \beta^{-1}\left[ln\left(-ln\left(\frac{y}{x}\right)\right) - ln(\alpha)\right]$, shown in the main text.

The only difference from the choice rule presented in the main text is now in the definition of the impatience parameter, which takes the form $\tilde{\alpha} \triangleq exp((1-\beta)(\mu + \frac{1}{2}\sigma^2))$ instead of $\alpha \triangleq exp((1-\beta)\mu)$. This adds to the intuition of why an increase in impatience was Note, however, that the two models deliver predictions that are empirically indistinguishable. Indeed, the two models fit to the same data show a difference in predictive fit that is exactly equal to 0. The reason is very simply that μ is a 'free parameter'. It is free in the sense that it does not enter into the definition of any other parameter, and can thus flexibly adjust following the difference in the definition of $\tilde{\alpha}$ and α.

C Generalization to include outcome distortions

I have assumed outcomes to be perceived without noise. If outcomes are also subject to noisy coding, the two dimensions will need to be derived jointly to accurately represent the decision-making process. Starting from the once-logged choice rule $\tau_\ell - \tau_s < -ln\left(\frac{y}{x}\right)$, let $z \triangleq -ln(\frac{y}{x})$. One can then assume that z is subject to noisy coding. Assuming that the log of the ratio of the smaller to larger payment is encoded in a way similar to time delays, we obtain the following likelihood and prior: $r_z \sim \mathcal{N}(ln(z), \nu^2)$, $ln(z) \sim \mathcal{N}(\mu_o, \sigma_o^2)$. Using the usual procedure and after defining $\gamma \triangleq \frac{\sigma_o^2}{\sigma_o^2+\nu^2}$, we obtain the posterior expectation $E[ln(z)|r_z] = \gamma r_z + (1-\gamma)\mu_o$. Substituting this expression together with the one for the time delay into the choice rule, we obtain the threshold equation $\beta r - \gamma r_z < -ln(\alpha)$, which assumes that the sooner and later amounts share a common prior (see Khaw et al., 2021, for an analogous assumption for decisions under risk). Distributing the signals jointly, obtaining the z-score, and comparing it to the threshold z-score, we obtain the following probabilistic choice rule:

$$Pr[(x,\ell) \succ (y,0)] = \Phi\left(\frac{\gamma \times ln\left[-ln\left(\frac{y}{x}\right)\right] - [ln(\alpha) + \beta \times ln(t)]}{\nu\sqrt{\beta^2+\gamma^2}}\right). \tag{10}$$

Notice that the addition of outcome transformations now affects mostly The model parameters α and β. That is, one could rewrite the numerator of the expression in parentheses as $ln\left[-ln\left(\frac{y}{x}\right)\right] - \gamma^{-1}\left[ln(\alpha) + \beta \times ln(t)\right]$ (as long as one rewrites the denominator as $\gamma^{-1}\left(\nu\sqrt{\beta^2+\gamma^2}\right)$). Once can then see that the addition of a parameter $\gamma \leq 1$ would simply serve to reinforce the conclusions already reached in the main text.

D Fit with the literature

In this section, I briefly review some of the most closely related discount functions used in the literature. While an exhaustive review of discounting models and functions is beyond the scope of this article, this will allow me to compare the key features of the NCT model with those of some formally related functions.

Samuelson (1937) famously introduced a model in which utilities of future outcomes are discounted by an exponentially decreasing discount function:

$$D(\tau) = exp(-\rho\tau), \tag{11}$$

where D is the discount function and ρ is the subjective discount rate. The model thus entails constant impatience and no delay-dependence. It is also deterministic—a feature that it shares will all discount functions discussed here in sequence.

While the model above remains the normative benchmark, many descriptive deviations from that benchmark have been observed over the years (Thaler, 1981; Loewenstein and Prelec, 1992; Read, 2001). Many types of discount functions have been proposed to account for such 'anomalies'. A popular form in economics is the quasi-hyperbolic discount function (Phelps and Pollak, 1968; Laibson, 1997):

$$D(\tau) = \beta \times exp(-\rho\tau), \tag{12}$$

where $\beta \triangleq 1$ for immediate payments, and $\beta < 1$ thus captures present-biased behaviour when one of the payouts is immediate. Once an up-front delay is included, β drops out and discounting is exponential. Recent meta-analyses provide clear evidence for present-bias in experiments using monetary rewards, although present-bias may be even stronger when time discounting is measured for consumption (Imai et al., 2021).

A parallel literature has discussed hyperbolic discounting. Hyperbolic functions are often motivated by the common difference effect, an anomaly whereby the discount rate for a given time delay decreases as a common delay is added to both outcomes (Loewenstein and Prelec, 1992). Empirically, however, the function has most often been fit using delays of different length from the present (Thaler, 1981; Kirby, 1997; Ebert and Prelec, 2007; Zauberman et al., 2009), thus potentially confounding strongly decreasing impatience and delay-dependence (Read, 2001). In addition to the constant sensitivity function proposed and axiomatized by Ebert and Prelec (2007) already discussed above, Mazur (1987) introduced an early one-parameter function capturing 'proportional discounting':

$$D(\tau) = \frac{1}{1 + \zeta\tau}, \tag{13}$$

where ζ is the subjective discounting parameter. Discounting is by definition proportional to the time delay. Sozou (1998) provides biological foundations for this function based on hazard rates, i.e. probabilities of future rewards not materializing. Just like all the other functions from the general hyperbolic class, the function cannot account for delay-dependence at all. Even when fitting purely data obtained with different delays from the present, however, the functional fit is typically poor. Figure 12 shows the fit to the data obtained with delays AB, AC, and AE in experiment I, and compares it to an exponential benchmark. Although the function fits better than the exponential function, the improvement in fit is small, and overall, fit remains poor.

Another popular function is the hyperbolic function proposed in the seminal paper of Loewenstein and Prelec (1992). The function takes the following form:

$$D(\tau) = \left(\frac{1}{1 + \zeta\tau}\right)^{\frac{\xi}{\zeta}}, \tag{14}$$

where $\zeta \geq 0$ measures the degree of decreasing impatience. The proportional discount function seen above emerges as a special case when $\xi = \zeta$. Again, decreasing impatience, present-bias, and delay dependence cannot be separately quantified.

Bleichrodt, Rohde and Wakker (2009) proposed a more flexible double-exponential family of functional forms, that allow to overcome limitations of the above function in capturing very strong patterns of decreasing impatience, as well as increasing impatience. Their constant relative decreasing impatience function converges to

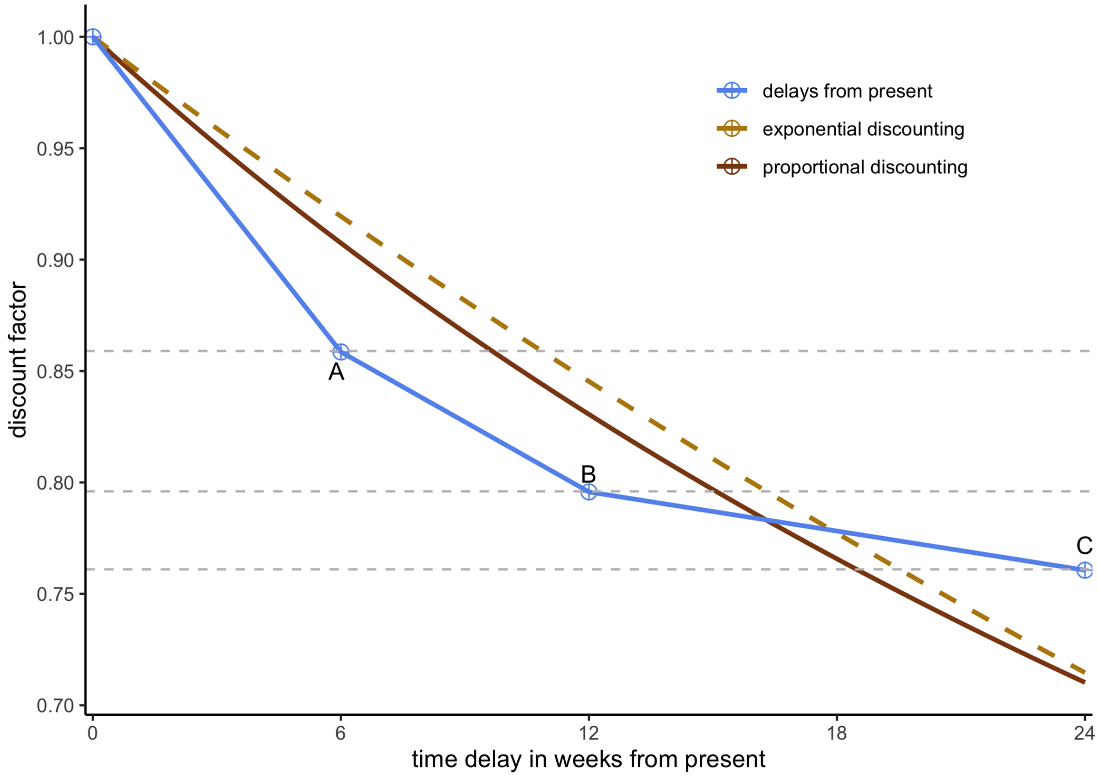


Figure 12: Fit of proportional discounting function to delays from the present

The figure shows the fit of a proportional discounting function to the aggregate level data from experiment I, obtained using delays of different length from the present. Though better than the fit of the exponential function, overall the fit remains poor.

the function of Ebert and Prelec (2007) when immediate payouts are included. Their constant absolute decreasing impatience function takes the following form:

$$D(\tau) = \begin{cases} e^{r \times e^{-c\tau} - 1} & \text{if } c > 0 \\ e^{-r\tau} & \text{if } c = 0 \\ e^{-r \times e^{-c\tau}} & \text{if } c < 0, \end{cases} \tag{15}$$

where the parameter c indicates the convexity of the discount function $D(.)$. For $c = 0$, the function converges to exponential discounting. Once again, the function is geared towards capturing decreasing impatience, but foresees no separate mechanisms for present-bias and delay-dependence.

Notwithstanding the large literature on generalized hyperbolic discount functions, the evidence on impatience decreasing for constant delays as the up-front delay increases is mixed at best. Read (2001) found no evidence for strongly decreasing impatience across 3 experiments. Developing and testing a novel index for decreasing impatience, Rohde (2019) found evidence for present-bias, but not for further decreases in impatience with larger up-front delays. He et al. (2019) concluded from two experiments that "decreasing impatience is not as robust as is widely held" (p. 63). Similar conclusions were reached by a number of other recent papers carefully controlling for delay-dependence (Attema, Bleichrodt, Rohde and Wakker, 2010; Cavagnaro, Aranovich, McClure, Pitt and Myung, 2016). On the other hand, Bleichrodt, Gao and Rohde (2016) report evidence for strongly decreasing impatience for both health and money. Ultimately, however, the global evidence remains difficult to assess due to the variety of measurement methods and stimulus types used in the literature. It is worth noting that the predictions of the NCT model apply to a binary choice setup, which may be difficult to map into other methodologies used in the literature. For instance, Read and Roelofsma (2003) explicitly tested direct choice against matching techniques, and found some evidence for strongly decreasing impatience in matching, but none in choice tasks.

While an in-depth review of the literature on decreasing impatience is beyond the scope of this article, there is no doubt that at least *some* papers have documented such patterns. A possible explanation is that these findings may be driven by an additional motive I have not modelled. Halevy (2008) proposed a model whereby choice patterns reminiscent of hyperbolic discounting may arise in trade-offs between timed, nominally sure amounts because of the inherent uncertainty of the future. If agents have some uncertainty about future payoffs, and if they exhibit nonlinear probability distortions, then this will result in present-biased behaviour. Epper and Fehr-Duda (2018) further showed how the uncertainty of the future could result in strongly decreasing impatience, as well as accounting for a number of other stylized facts that have been documented in the literature. The uncertainty in these models is captured by the subjective probability that a future payment will really take place and can be enjoyed. The uncertainty in the model I have presented concerns the mental representation of the time delay itself. That being said, the two types of uncertainty may well co-exist, and the model here presented could easily be generalized to include the inherent uncertainty of the future—a point to which I will return in the discussion.

Finally, Read (2001) proposed the following functional form to account for delay-dependent preferences between a sooner delay τ_s and a later delay τ_ℓ:

$$D(\tau_s, \tau_\ell) = \frac{1}{1 + \zeta(\tau_\ell - \tau_s)^\gamma}, \tag{16}$$

where ζ again captures discounting proportional to the time *delay*, and $0 < \gamma < 1$ captures delay-dependence. Kable and Glimcher (2010) propose a variant of this function imposing $\gamma \triangleq 1$, which captures what they refer to as 'as soon as possible (ASAP) discounting', so called because it captures seemingly hyperbolic patterns for increasing delays no matter what the up-front delay. The function is founded on neural data which failed to detect separate coding of an 'immediacy signal' in the brain—an observation that is consistent with the NCT model implementation of present-bias. The main differences with the NCT model are that these models cannot account for present-bias, and that the two parameters are nominally independent.

Several recent models have proposed setups in which hyperbolic discounting emerges due to randomness in decisions and in tastes. Lu and Saito (2018) proposed a random preference model in which the discount function of a decision maker is stochastic. He et al. (2019) combined an exponential discounting model with both stochastic choice under the form of a logit model, and stochastic tastes under the form of a random preference model. A central feature of these models is that they arrive at predictions of decreasing impatience even when the underlying discounting is exponential. These models are, however, geared towards decreasing impatience, and none of them predicts delay-dependence, making them rather distinct from the NCT model proposed here.

E Econometric approach

The noisy coding of time model is inherently stochastic, so that it can be directly implemented without any further need for separate assumptions about the error structure. I augment traditional models, which are deterministic in nature, by a normally distributed, additive error term, which constitutes the most common error model used in the literature. While some error models have been proposed that may result in non-stationary discounting patterns even when starting from an optimal setup of exponential discounting (Lu and Saito, 2018; He et al., 2019), these error models result in predictions that are quite distinct from those emerging from the model here presented, and which are rather closer to traditional setups.

I use a Bayesian random-parameter setup to obtain individual-level parameter estimates jointly with aggregate estimates, which serve as endogenously-estimated priors for the individual estimates. Compared to purely aggregate estimates, such a setup has the advantage of producing individual-level estimates of the parameters of interest; compared to individual-level estimates, such a model discounts noisy outliers, thus resulting in increased predictive performance (Conte, Hey and Moffatt, 2011; von Gaudecker, van Soest and Wengström, 2011; Baillon, Bleichrodt and Spinu, 2020). I thus obtain the posterior estimate $p_i(\boldsymbol{\theta}_n|z)$ given choice data z over the individual parameter vector $\boldsymbol{\theta}_n$ from

$$p_i(\boldsymbol{\theta}_n|z) \propto p(z|\boldsymbol{\theta}_n) \times p(\boldsymbol{\theta}_n), \tag{17}$$

where z takes the value 1 if the larger-later reward is chosen and 0 otherwise, and where the likleihood $p(z|\boldsymbol{\theta}_n)$ is defined as follows:

$$p(z|\boldsymbol{\theta}_n) = (Pr[(x, \tau_\ell) \succ (y, \tau_s)])^z \times (1 - Pr[(x, \tau_\ell) \succ (y, \tau_s)])^{1-z}, \tag{18}$$

with $Pr[(x, \tau_\ell) \succ (y, \tau_s)]$ taking the form of the choice probability in equation 7 for the NCT model. For the standard models, on the other hand, the choice probability will be defined by the functional form underlying the model, plus a normally distributed additive error term, assumed to take the form of 'white noise'

(Hey and Orme, 1994; Bruhin, Fehr-Duda and Epper, 2010).

Finally, the prior distribution for the individual-level parameters, $p(\boldsymbol{\theta}_n)$, takes the following form:

$$p(\boldsymbol{\theta}_n) = \mathcal{N}(\overline{\boldsymbol{\theta}}, \boldsymbol{\Sigma}), \tag{19}$$

where $\overline{\boldsymbol{\theta}}$ is a vector containing the aggregate parameter means, and $\boldsymbol{\Sigma}$ is a covariance matrix of the individual-level parameters. Both $\overline{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}$ are endogenously estimated from the aggregate data, and serve as priors for the individual-level estimates. The hyperpriors for the parameters in $\overline{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}$ are chosen to be mildly regularizing, thus helping the algorithm to converge, but being wide enough to accommodate any plausible parameter values that may emerge from the data. This follows best practices in Bayesian econometrics (McElreath, 2016), and the estimates reported below are not sensitive to changes in the hyperpriors used, given that the amount of data can easily overpower any prior at the aggregate level.

I maximize the logged sum over the choice-level observations i of the likelihood function described above using Bayesian simulations in Stan (Carpenter, Gelman, Hoffman, Lee, Goodrich, Betancourt, Brubaker, Guo, Li and Riddell, 2017), launched from RStan (Stan Development Team, 2017). I conduct comparisons between different models using leave-one-out cross-validation (Gelman, Hwang and Vehtari, 2014; Vehtari et al., 2017). The choice of cross-validation methods for model selection serves to avoid the overfitting of existing data, instead focusing on the *predictive* performance of the models. This ensures coherence with the random-parameter setup used, and constitutes a more adequate test of model fit in the present setting than alternative methods geared towards optimizing the fit to existing data.

F Additional results experiment I

Figure 13 shows the comparison of decumulative choice proportions for the delay AC compared to the delay CE, i.e. for 12 week delays where the first delay occurs from the present, and the second from an upfront delay of 12 weeks.

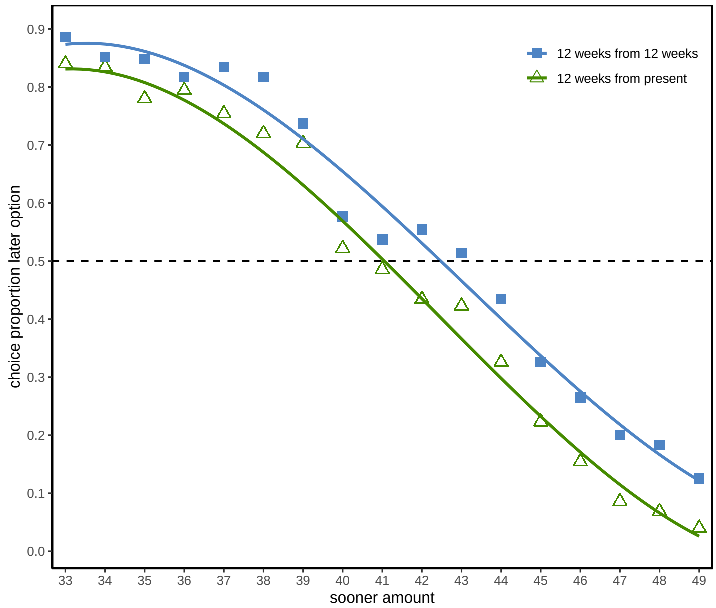
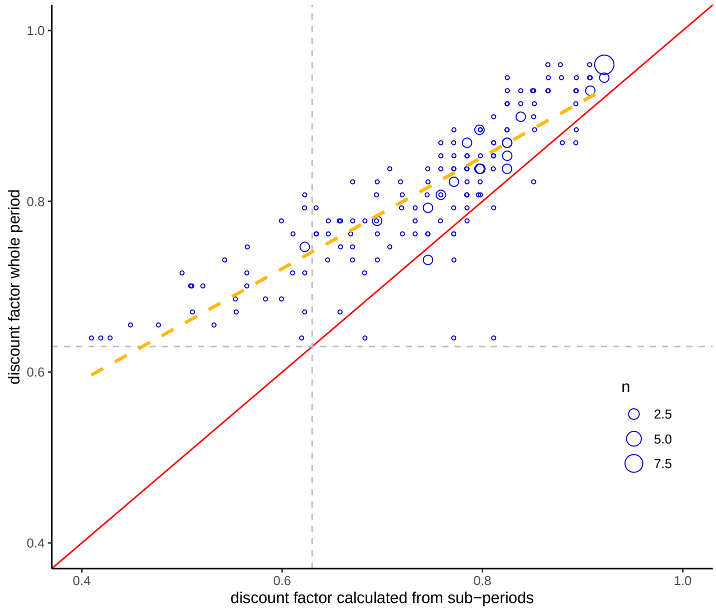
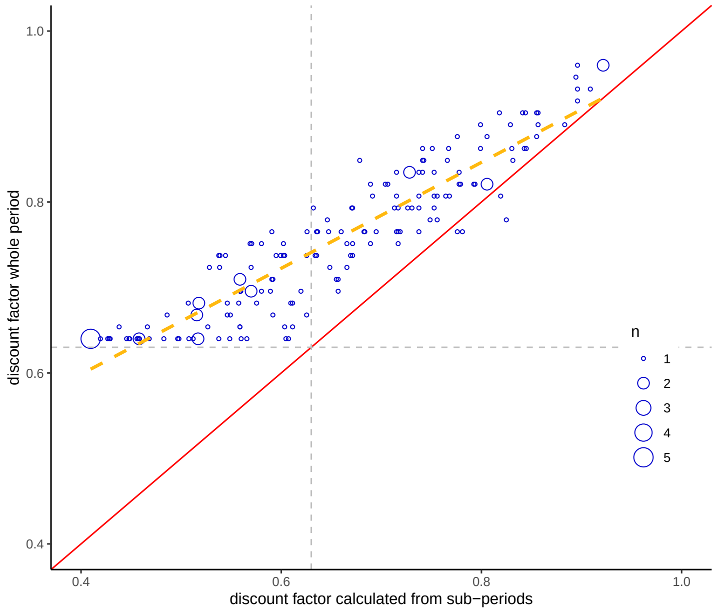


Figure 13: present-bias for 12 week delays—additonal comparisons



(a) subadditivity, 12 weeks from 12 weeks



(b) subadditivity, 24 weeks vs 12-week delays

Figure 14: Non-parametric illustration of subadditivity—additonal comparisons

The subadditivity comparisons are obtained by comparing a discount factor over the whole period with the product of the discount factors of the subperiods. The pattern in panel 14(a) thus obtains from a comparison of the discount factor $\delta_{12,24}$ with the product of the two discount factors $\delta_{12,18} \times \delta_{18,24}$, where the subscripted numbers indicate the extremes of the time delays. The pattern in panel 14(b) obtains from a comparison of the discount factor $\delta_{0,24}$ to the product of all underlying 12-week discount factors, $\delta_{0,12} \times \delta_{12,24}$. Dashed lines indicate correlations.

G Experimental Instructions

Thank you for taking part in this experiment. You will be asked to take some decisions involving time delays. On each screen, you will be asked to choose between an amount of money that is paid at **a sooner moment in time**, and an amount that is paid **at a later moment in time**. Time delays are always indicated in weeks from today. Here is an example of a choice task:

Make your choice:

€49 immediately ○ ○ €50 in 24 weeks

You will be presented repeatedly with such tasks, and you are asked to indicate your choice for each one of those tasks. The experiment will take approximately 20 minutes. At the end of this class, **10 participants will be randomly drawn to play one of their choice for real money**. Notice that both the amounts and the time delays involved may change from screen to screen. Please consider the information carefully and choose your preferred option.

At the end of the experiment, **the lecturer will announce the students selected to play for one of their choices**. If you have chosen the **immediate amount** in the randomly selected choice, we will transfer the corresponding amount to your bank account immediately while you are waiting. Since transfers amongst all major banks are immediate, you will be asked to check your bank account and **confirm the receipt of the money before leaving**.

If you have chosen a **delayed amount** in the randomly selected choice, then that amount will be paid to you on the indicated day. For instance, an amount indicated as obtaining **in 4 weeks** will be paid 4 weeks from today, i.e. on Monday November 15th. The lecturer will make a note of your bank account to organize the transfer. You will receive **a certificate signed by the lecturer to guarantee the transfer**. The certificate will indicate the amount to be paid and the date of transfer. It will also contain the contact details of the lecturer, for the case that you change bank account or have any questions concerning the transfer.

H Additional results experiment II

Here, I document the stylized patterns predicted by the NCT model. Figure 15 shows decumulative choice proportions for the periods AB versus BC for weeks (panel A) versus days (panel B). In both cases, there is clear evidence for present-bias. The same is true when comparing a delays between now and 24 weeks, to a delay between 24 weeks and 48 weeks. Once again, present-bias is string both in the weeks treatment (panel C) and in the days treatment (panel D).

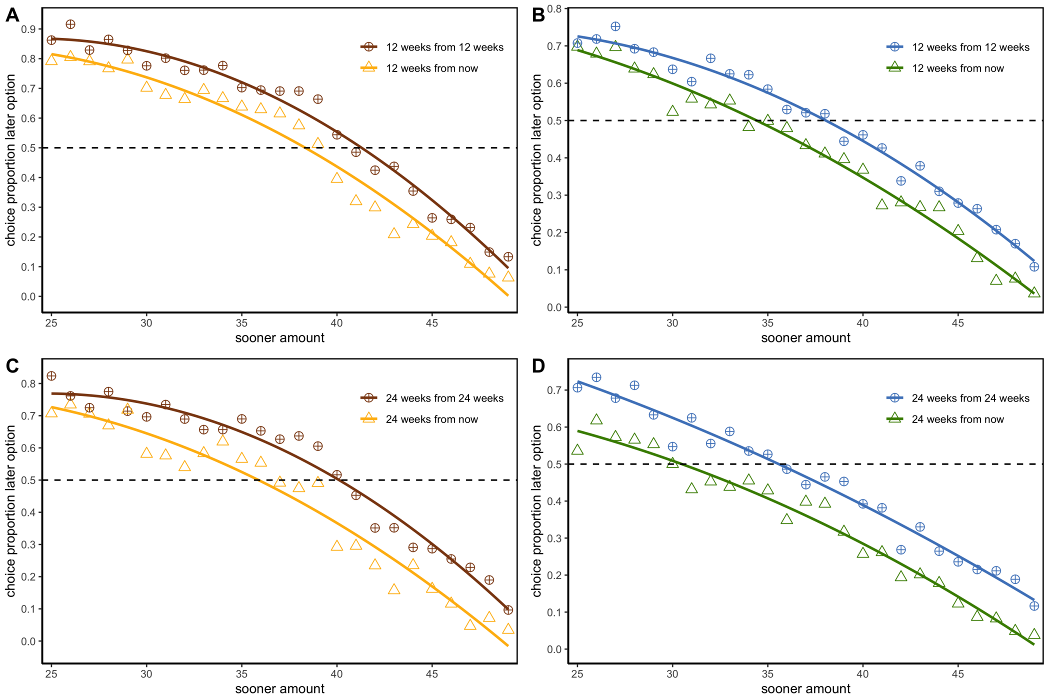


Figure 15: Decumulative choice proportions for larger-later option, present-bias

Figure 16 displays decumulative choice proportions for the larger-later reward as the upfront delay is scaled up. We see the same patterns as in experiment I. In particular, there is no consistent effect indicating decreasing impatience as the up-front delay is increased, and this is true both for weeks and the days frame, and both using the interval AC (panel A) and the interval AE (panel B).

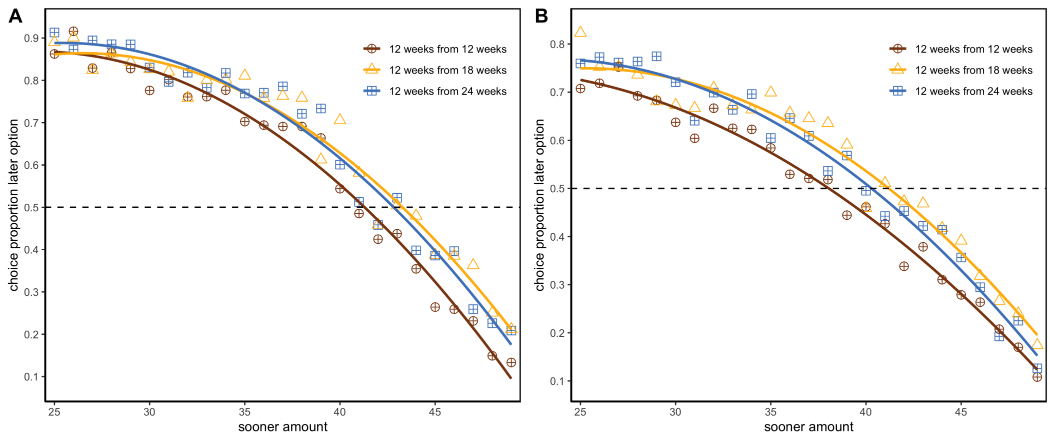


Figure 16: Decumulative choice proportions for larger-later option, common difference effect

Finally, figure 17 documents delay dependence. Once again, delay dependence is string in both the week frame (panel A) and days frame (panel B).

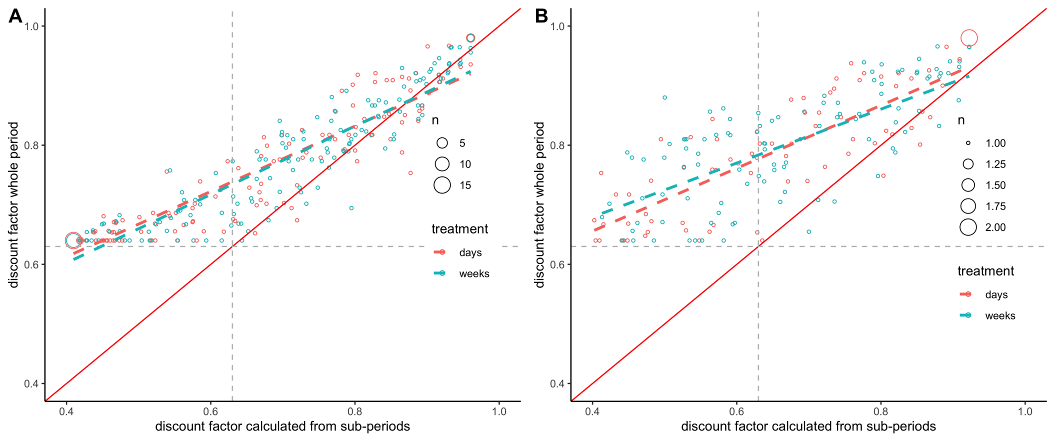


Figure 17: Scatter plot of discount factors over long time delays against factors calculated from shorter time delays