American Psychological Association

Decision

ISSN: 2325-9965

2020, Vol. 7, No. 3, 212–224
http://dx.doi.org/10.1037/dec0000121

BRIEF REPORT

Hierarchies Improve Individual Assessment of Temporal Discounting Behavior

M. Fiona Molloy
The Ohio State University

Ricardo J. Romeu
Indiana University

Peter D. Kvam
University of Florida

Peter R. Finn and Jerome Busemeyer
Indiana University

Brandon M. Turner
The Ohio State University

Delay discounting behavior has proven useful in assessing impulsivity across a wide range of populations. As such, accurate estimation of the shape of each individual's temporal discounting profile is paramount when drawing conclusions about how impulsivity relates to clinical and health outcomes such as gambling, addiction, and obesity. Here, we identify an estimation problem with current methods of assessing temporal discounting behavior and propose a simple solution. First, through a simulation study, we identify types of temporal discounting profiles that cannot reliably be estimated. Second, we show how imposing constraints through hierarchical modeling ameliorates these recovery problems. Finally, we apply our solution to a large data set from a temporal discounting task and illustrate the importance of reliable estimation within patient populations. We conclude with a brief discussion on how hierarchical Bayesian methods can aid in model estimation, compensate for small samples, and improve predictions of externalizing psychopathology.

Keywords: delay discounting, hyperbolic discounting, hierarchical Bayesian modeling, intertemporal choice

Supplemental materials: http://dx.doi.org/10.1037/dec0000121.supp

Delay discounting is a psychological phenomenon where the subjective valuation of a reward is lower than the objective value, in proportion to the amount of time a person must wait to obtain the reward. Often, this phenomenon is measured using monetary intertemporal choice tasks. These tasks are characterized by asking subjects to decide between a smaller-sooner option, for example, \$10 now, or a larger-later option, for example, \$20 in 2 weeks. A tendency to choose the smaller-sooner option can be used to measure impulsive choice. Numerous explanations of how observers trade reward for delay have been proposed, including

This article was published Online First March 26, 2020.

M. Fiona Molloy, Department of Psychology, The Ohio State University; Ricardo J. Romeu, Department of Psychological and Brain Sciences, Indiana University; Peter D. Kvam, Department of Psychology, University of Florida; Peter R. Finn and Jerome Busemeyer, Department of Psychological and Brain Sciences, Indiana University; Brandon M. Turner, Department of Psychology, The Ohio State University.

This work was funded by National Institute on Alcohol Abuse and Alcoholism R01AA13650 to Peter Finn, a CAREER Award from the National Science Foundation to Brandon Turner, and a National Institute of Drug Abuse Grant T32DA024628 to Ricardo Romeu.

Correspondence concerning this article should be addressed to Brandon M. Turner, Department of Psychology, The Ohio State University, 200C Lazenby Hall, 1827 Neil Avenue, Columbus, OH 43210. E-mail: turner.826@gmail.com

exponential discounting models (Becker & Murphy, 1988; Lancaster, 1963) as well as attention-based models (Cheng & González-Vallejo, 2016; Dai & Busemeyer, 2014; Scholten & Read, 2010; Turner et al., 2019), but the most common description is a hyperbolic function (Mazur, 1987). The hyperbolic function allows for the estimation of the discounting rate (k), which is unique for individuals based on their delay discounting behavior/preferences. Larger k values mean the larger-later option is more heavily discounted, signifying a preference for the more impulsive smaller-sooner option. In numerous applications, k has been used to generalize meaningful connections about how discounting behavior is related to substance use disorders and addiction (Bailey, Gerst, & Finn, 2018; Bickel & Marsch, 2001; Bickel, Odum, & Madden, 1999; Kirby, Petry, & Bickel, 1999) and other risk-taking and health behaviors (Daugherty & Brase, 2010).

Despite the importance of accurately estimating the temporal discounting rate, few analyses have investigated the recoverability of the parameters of common temporal discounting models. Parameters of the same model may vary in their recoverability. For example, in linear regression, the slope and intercept of a line are more difficult to estimate when the residual noise is high. Here, we first present a simulation study to expose weaknesses of standard estimation procedures in the hyperbolic discounting model. We found several types of common experimental designs where temporal discounting curves cannot be reliably estimated. We present a simple hierarchical solution that solves the issues identified in our simulation study and provide code to facilitate our recommended approach. After illustrating our solution with simulated data, we apply it to data from an intertemporal choice experiment (Finn, Gunn, & Gerst, 2015). Here, we show that the more constrained hierarchical estimates were more highly correlated with externalizing psychopathology than nonhierarchical estimates. We conclude with a general discussion emphasizing the importance of accurate assessment of temporal discounting behavior.

Model Specification

To assess our ability to reliably model temporal discounting behavior, we made a choice about the particular functional form. Although our results generalize to other models of temporal discounting behavior (see online supplemental material), we focus on the hyperbolic function due to its prevalence and intuitiveness. In our own applications, we noticed that there were occasionally reliability issues during the model-fitting process for some data structures. Unreliable (i.e., inaccurate, imprecise, or both inaccurate and imprecise) estimates are a major concern when the goal is to generalize a specific behavior (e.g., temporal discounting) to important societal problems (e.g., addiction). It is therefore critical to identify where and when the shape of the hyperbolic discounting model cannot be reliably estimated. In this section, we specify the general form of the model, then discuss its construction, both nonhierarchically and hierarchically, in a Bayesian framework.

To begin, the hyperbolic discounting function (Mazur, 1987) is

$$V_{LL} = \frac{r_{LL}}{1 + kt}, \tag{1}$$

where V_{LL} is the subjective value of the delayed (larger-later) option, r_{LL} is the nondiscounted value of the delayed option, k is the estimated discounting rate, and t is the delay, in days, of the larger-later option. The likelihood of choosing the larger-later option is defined as

$$P_{LL} = \frac{1}{1 + e^{-m(V_{LL} - V_{SS})}}, \tag{2}$$

where P_{LL} is the likelihood of choosing the delayed option, m is the estimated sensitivity to changes in the discounted value (Dai & Busemeyer, 2014; Dai, Gunn, Gerst, Busemeyer, & Finn, 2016; Scherbaum, Haber, Morley, Underhill, & Moustafa, 2018; Wulff & van den Bos, 2018), V_{LL} is the subjective value of the larger-later option from Equation 1, and V_{SS} is the subjective value of the smaller-sooner option. For our purposes, the smaller-sooner option is always immediate ($t = 0$ days) and thus we can rewrite Equation 2 as

$$P_{LL} = \frac{1}{1 + e^{-m(V_{LL} - r_{SS})}}, \tag{3}$$

where r_{SS} is the amount in dollars of the immediate option.

Nonhierarchical Implementation

In the nonhierarchical version of the model, we estimate a k and m pair for every individual separately. Because we are interested in exploring the parameter space, we want to have relatively uninformed priors. The priors for k and m were uniformly distributed from 0 to 10:

$$k \sim \mathcal{U}(0, 10),$$
$$m \sim \mathcal{U}(0, 10).$$

These priors are considered uninformative because of their distribution and range. First, the range is constrained to be positive (as k and m can only be positive), but the spread of the distribution is much larger than typical k and m values. Additionally, a uniform distribution assumes the parameter could be any value within this range with equal probability.

Hierarchical Extension

As we will show below (and see online supplemental materials), the primary reason for the unreliable estimation is the degenerative shape of the likelihood function at the individual level. One solution to correct the shape is to construct a hierarchical model where information can be shared across individuals. While we are certainly not the first to propose hierarchical Bayesian modeling to estimate parameters of delay discounting (Chávez, Villalobos, Baroja, & Bouzas, 2017; Vincent, 2016), we aim to emphasize the importance of performing hierarchical analyses, which are made significantly more convenient within a Bayesian framework. Constructing a hierarchy across subjects allows us to use information from other subjects to compensate for missing information or insufficient data. To facilitate construction of the hierarchical model, we now specify truncated normal priors for both k and m:

$$k_s \sim \mathcal{TN}(\mu_k, \sigma_k, 0, \infty)$$
$$m_s \sim \mathcal{TN}(\mu_m, \sigma_m, 0, \infty),$$

where s indexes the subject, and $\mathcal{TN}(a, b, c, d)$ denotes a truncated normal distribution with mean a, precision b, lower bound c, and upper bound d. The range of the truncated normal constrains the estimates to be only positive, while still remaining largely uninformative. Hence, μ_k and μ_m are the means, and σ_k and σ_m are the precision terms, for k and m, respectively. We then assumed uniform priors from 0 to 10 for μ_k and μ_m:

$$\mu_k \sim \mathcal{U}(0, 10)$$
$$\mu_m \sim \mathcal{U}(0, 10),$$

and exponential priors with rate equal to one for both σ_k and σ_m:

$$\sigma_k \sim Exp(1)$$
$$\sigma_m \sim Exp(1).$$

The priors for μ_k and μ_m are analogous to the nonhierarchical priors for k and m. Example code for fitting this hierarchical model in Just Another Gibbs Sampler (JAGS) can be found here: https://github.com/MbCN-lab/hierarchical-discounting/. For generalizability, we create population-level parameters for only one group (e.g., one μ_k and one μ_m parameter). However, this framework can be extended to multiple groups based on the research question or design of the experiment. For example, there could be a hyperparameter for a control group and a separate hyperparameter for a treatment group.

Simulation Study

When applying the model to experimental data, we have no assurances on the accuracy of our estimates, as the true values of the model parameters are unknown. Hence, to properly assess statistical issues such as accuracy and precision, we must investigate the model in an environment where the true parameter values are known. Our first goal is to uncover areas of the parameter space that are problematic in terms of recoverability. To do this, we ran two different simulation studies. In the first simulation study, we defined a grid over the k and m space, generated data for each value in the grid, and recovered the model parameters. We repeated this procedure while varying the number of trials to explore the relationship between recovery and experimental design. Our hypothesis was that the reliability of the model param-

eters would increase systematically with additional trials. In the second simulation study, we compared the accuracy of estimation between hierarchical and nonhierarchical models. The hierarchical model incorporates more information, so we expected that, especially for fewer trials, the hierarchical model would correct the degenerative shape of the likelihood function.

Areas of Unreliable Estimates

Method. To explore the effect of the number of trials on parameter recovery, we generated data consisting of four different trial sizes: 30 trials, 50 trials, 70 trials, and 150 trials. The trial sizes, delays, and larger-later and smaller-sooner options were inspired by the experimental design used within Finn et al. (2015). Here, we will discuss the results from the two extremes of 30 and 150 trials, although every trial size was fit. The larger-later option was always \$50, whereas the smaller-sooner option was sampled from values from \$2.50 to \$47.50, in increments of \$2.50. The delays were uniformly sampled from the following set: 7, 14, 30, 90, and 365 days. We defined a 100-by-100 grid of values within the joint (k, m) parameter space. The sequence for k was from 0.01 to 1, increasing in steps of 0.01. The sequence for m was from 0.05 to 5, increasing in steps of 0.05. For each pair of k and m across these ranges, we calculated the V_{LL} and simulated a choice. For each trial, the V_{LL} was determined by $50/(1 + k * t)$, and then the P_{LL} was calculated by $1/(1 + exp(-m * (V_{LL} - V_{SS})))$. The simulated choice was generated by randomly sampling from a Bernoulli distribution with P_{LL} as the probability of success. Preferential choices have been found to be probabilistic, which this random sampling accounts for (Rieskamp, 2008).

Once the data were generated, we fit the nonhierarchical hyperbolic model specified above using JAGS (Plummer, 2003). We fit the model five times for the five data sets generated for every pair of k and m values across all four trial sizes. Each model was fit using three chains, where each chain was initialized for 3,000 adaptations, with a burn-in of 4,000 iterations and 6,000 samples. Hence, posterior samples consisted of 18,000 samples. Chains were visually assessed for convergence. Code for simulating data and fitting the nonhierarchical model to these simulated data can also be found at https://github.com/MbCN-lab/hierarchical-discounting/.

Results. Figure 1a shows the accuracy and precision of nonhierarchical k and m estimates over the grid of parameter values. Precision and accuracy are important considerations in evaluating parameter recovery. Ideally, for accuracy, the mean of the posterior should be close to the true value of the parameter. We quantified accuracy as the root-mean-square error (RMSE). RMSE is defined as

$$\sqrt{(k - \hat{k})^2 + (m - \hat{m})^2},$$

where k is the true k value, $\hat{k}$ is the estimated k value, m is the true m value, and $\hat{m}$ is the estimated m value. The RMSE is presented in the left column of Figure 1a. Comparing differences in spread between the prior and posterior measures precision and gives us some insight into how well the data constrain the estimate. This spread is compared by dividing the standard deviation of the posterior by the standard deviation of the prior. If this ratio is 1, the data do not provide much information about the estimate, as the spread of the posterior is the same as the spread of the prior. However, if the ratio is small, the data allow for precise estimates of the parameters. The right column of Figure 1a shows the plots of the standard deviation ratios.

Each row in Figure 1a corresponds to the number of generated trials (30 or 150) used to generate the data. As the number of trials increases, estimates become simultaneously more accurate and precise across the parameter space. Furthermore, across all trial sizes, small k and m pairs (i.e., $k < 0.2$ and $m < 1$), are most successfully recovered compared to the rest of the sample space. In some cases, for large values of m, the accuracy is high (low RMSE), but the precision is much lower (high SD ratio). This pattern results from the selection of the priors. Because the prior was set to be uniformly distributed from 0 to 10, even if the posterior is not constrained at all by the data, the mean will still be 5. Thus, for values of m closer to 5, we see relatively smaller RMSEs but not necessarily smaller SD ratios. Overall, we found large differences in recoverability across the parameter space, especially when the experimental design con-

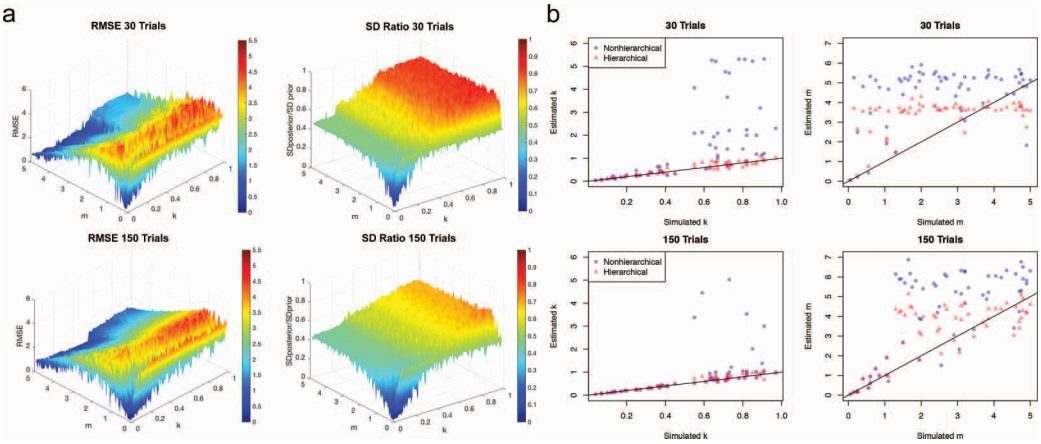

Figure 1. Simulation study results. (a) The problematic areas of recovery obtained from the nonhierarchical simulation study. The first and second rows show the results of parameter recovery of k (x-axis) and m (y-axis) for 30 and 150 generated trials, respectively. The left column shows how close the estimated k and m values were to the true k and m values, quantified by the root-mean-square error (RMSE; z-axis). The right column compares the prior and posteriors of the estimates, quantified by dividing the standard deviation of the posterior by the standard deviation of the prior. Red-orange colors denote a *SD* ratio of 1, that is, the posterior resembled the prior, whereas blue-green colors denote a more constrained posterior. (b) Nonhierarchical and hierarchical parameter recovery. The first and second rows (again for 30 and 150 trials, respectively) compare hierarchical (blue circles) and nonhierarchical (red triangles) estimates for k (left panels) and m (right panels). The x-axis shows the simulated or "true" values used to generate the data, and the y-axis shows the estimated parameter values using both hierarchical and nonhierarchical methods. The black line signifies where the simulated and estimated values are equivalent. See the online article for the color version of this figure.

sisted of fewer trials. Table 1 summarizes these results across trial sizes.

Hierarchical Versus Nonhierarchical Recovery

Method. In the second simulation study, we explored the differences between hierarchical and nonhierarchical estimation. First, 60 different k and m pairs were randomly generated within the same range as above (between 0.01 and 1 for k and between 0.05 and 5 for m). Data were generated for these pairs in the same way described above. There were 240 different data sets fit to both models. The nonhierarchical and hierarchical models were fit using the same model-fitting procedure described above.

Table 1
Summary of Nonhierarchical Root-Mean-Square Error (RMSE) and SD Ratios Across 30, 50, 70, and 150 Trials

Variable	Trials			
	30	50	70	150
M RMSE	2.384	2.355	1.971	1.636
Mdn $SD_{posterior}/SD_{prior}$	0.649	0.537	0.485	0.382

Results. The hierarchical model provided consistently more accurate estimates than the nonhierarchical model. Figure 1b shows the results of the simulated data fit to the nonhierarchical and hierarchical models. As before, each row corresponds to the number of trials generated, where the left column shows the k parameter and the right column shows the m parameter. In both models, each subject has a k and m estimate, so to directly compare the two models, we show k_s and m_s for the hierarchical model, not the population estimates μ_k and μ_m. Simulated values are on the x-axis, and estimated values are on the y-axis. Blue circles represent the means of the posterior estimates for the nonhierarchical model, red triangles represent the means of the posterior estimates for the

hierarchical model, and the black line shows where the simulated and estimated values are equal.

Across trials, in both the hierarchical and nonhierarchical models, the RMSE and standard deviation ratios were smaller than k than for m (although this may be attributed to the smaller range of true k values). Importantly, for k and m in both hierarchical and nonhierarchical models, as trial size increases, estimated k and m values converge to the true values. However, hierarchical estimates are consistently closer to the true values than nonhierarchical estimates. This disparity is exacerbated as the true values for k and m increase. The recovery of both the hierarchical and nonhierarchical models is affected by trial size, but misestimates common in a smaller number of trials are still closer to the true values for hierarchical estimates than for nonhierarchical estimates. Yet, even in the context of 150 trials (where the nonhierarchical model provided reasonable estimates for most of the simulated k values), there are still multiple cases where the hierarchical model recovers the true value but the nonhierarchical model overestimates the true value. This suggests that hierarchical analyses are still preferred even with a large number of trials.

We also examined the joint posteriors for the nonhierarchically and hierarchically estimated k and m pairs shown in Figure 1b. The nonhierarchical model displayed a clustering pattern at larger m values. This clustering pattern forces the k values to become smaller, suggesting the flat shape of the joint likelihood function may be causing this misestimation. However, once a hierarchy is introduced, the increased information from the group-level parameters allows the individual-level estimates to be "pulled in" by introducing a central tendency to the likelihood when little information is present for that individual. This alleviates the overestimation problem by imposing a low conditional likelihood Pr(individual – group) for high values of k or m, thus drawing them closer to the other (lower) individual-level estimates.

Real Data

In the previous section, we described a problem in the estimation procedure of hyperbolic discounting and proposed using a hierarchical framework and/or larger trial sizes to alleviate this issue. In this section, we apply this solution of hierarchical Bayesian modeling to real experimental data of 622 subjects from Finn et al. (2015). We compared the recovery and constraint of nonhierarchical and hierarchical models. Additionally, the data from this experiment involved a variable number of trials per subject, which allowed us to explore recoverability of nonhierarchical and hierarchical models as a function of trial size.

Method

Experimental paradigm. The data we investigated are part of a larger study by Finn et al. (2015) on the relationships between intertemporal choice, working memory capacity, and externalizing psychopathology. Here, we will give a brief overview of the delay discounting task design, but refer to the original article for more detail. A total of 622 subjects completed the delay discounting task on a computer. They were asked to choose between an immediate monetary option and a delayed monetary option. The amount of the immediate option ranged from \$2.50 to \$47.50 in increments of \$2.50. The amount of the delayed option was always \$50. Note that the delayed amount was always larger than the immediate amount. The delayed option was delayed by 1 week, 2 weeks, 3 months, 6 months, or 1 year. These delays were then converted to days in the model-fitting procedure used here and by Finn et al. (2015). Each delayed option was presented in one of six randomly presented blocks. The block of a set delay consisted of both ascending and descending trials. For ascending trials, the smaller-sooner option started at \$2.50 and increased to a highest possible value of \$47.50 in increments of \$2.50. As soon as a subject chose the smaller-sooner option, the sequence of ascending trials ended. In descending trials, the smaller-sooner options started at \$47.50 and went down to \$2.50 in increments of \$2.50 resulting in a stair-step titration procedure originally aimed at experimentally identifying the indifference point (rather than the likelihood function used in a model-based approach). As soon as the subject chose the larger-later option, the descending trials stopped. The order of ascending and descending trials was randomized. Because of this design, there is not a fixed trial size across individuals. The number of trials for this task

ranged from 16 to 149 trials, with a mean of 115 trials. This feature gives us the opportunity to explore the effect of trial size in estimating k and m, with all other aspects of the experimental design equal.

Model-fitting procedure. To estimate the model parameters for the nonhierarchical model, we initialized three chains for 1,500 iterations with a burn-in of 2,000 iterations and sampled for 3,000 iterations, resulting in 9,000 samples of the joint posterior distribution. To estimate the model parameters for the hierarchical model, we initialized three chains for 1,500 iterations with a burn-in of 4,000 iterations and sampled for 5,000 iterations, resulting in 15,000 samples of the joint posterior distribution. Chains were visually assessed for convergence.

Results

Nonhierarchical recovery. The nonhierarchical model accurately predicted the probability of choosing a larger-later option, yet the estimates of k and m were not as expected. The nonhierarchical model very closely predicted the observed P_{LL}s, although there were a few cases of over- and underestimations, particularly at lower P_{LL}s. However, an ability to recover the choice proportions in the data does not guarantee that we can recover the generating parameters. This is partly because some choice pairs are much more informative than others for different parts of the parameter range. For example, a choice of \$5 now versus \$50 in 1 week is more diagnostic at extremely large values of k, while a choice of \$47.50 now versus \$50 in 1 year will be more diagnostic at extremely small values of k. Therefore, the right or wrong mixture of choices could result in particularly good or bad estimation of the model parameters even if it is able to accurately generate the observed choice proportions.

The left panel of Figure 2a shows the log-transformed parameter estimates for the nonhierarchical model fit to each subject. Each point represents a subject's mean of the posterior for the log-transformed k (x-axis) and log-transformed m (y-axis) estimates. The points also display the range of trials that subjects completed, where the cyan circles denote fewer than 50 trials, green squares denote 50–69 trials, red pluses denote 70–89 trials, and blue crosses denote more than 90 trials. The nonhierarchical model had stark differences in parameter recovery across trial sizes in the simulation study, so it is not surprising that this pattern also exists in the real data. This is most noticeable with fewer trials (denoted by cyan circles in Figure 2a), where subjects who completed fewer than 50 trials had the lowest log-transformed m estimates and the highest log-transformed k estimates.

Because we do not know the true values of the k and m values for real data, we cannot directly test the accuracy of the estimate, using a measure such as RMSE. However, because we used a grid in the simulation study, we can get an idea of the robustness of the estimates. Some of these estimated values fall on areas, found through our simulation study, that are difficult to recover (evidenced by high RMSEs). Higher RMSEs also signify that the estimates are biased. These problematic points were all with small m and larger k values, and they were present even with additional trials. However, the majority of the estimates that were within the range of our simulation study fell within areas with small m and k values (less than 1 and 0.2, respectively), which had small RMSEs in our simulation study.

Hierarchical recovery. A subject's probability of choosing the larger-later option was also accurately predicted in the hierarchical model. The hierarchical model tended to overestimate smaller P_{LL}s, which may be a result of the hierarchical model's tendency to pull outliers toward the population mean. However, this is only a minor difference, and overall the model correctly predicts P_{LL}. The right panel of Figure 2a shows the log-transformed k and m estimates for the hierarchical model, organized by trial number. Similar to the nonhierarchical model, there was a relationship between trial size and parameter estimates. For example, the estimates for 50 and fewer trials had smaller log-transformed m and larger log-transformed k estimates.

Comparing the hierarchical and nonhierarchical parameter estimates, the range of estimates was more constrained for the hierarchical estimates than for the nonhierarchical estimates. This pattern was also observed in the simulation study and demonstrates the property of shrinkage inherent to hierarchical models. Figure 2b directly compares an individual's mean hierarchical and nonhierarchical estimates. While the

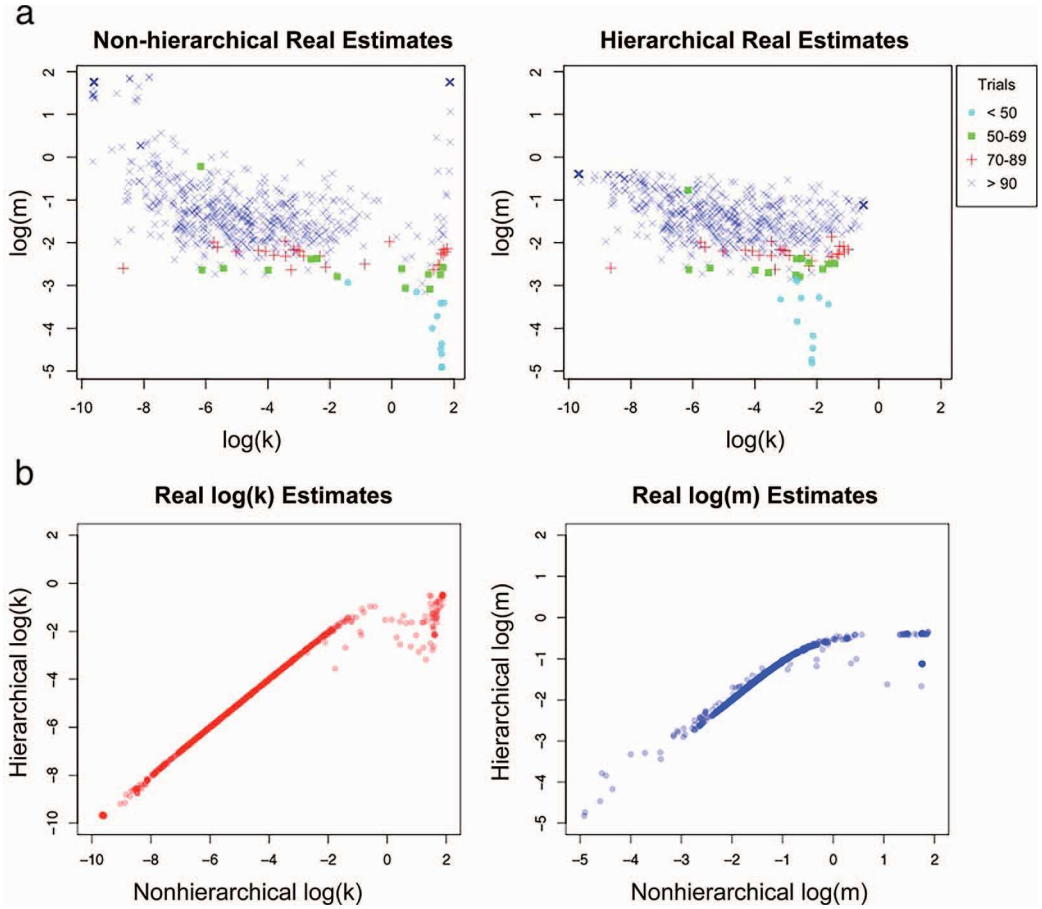

Figure 2. Real data results. Row a shows the k and m estimates of experimental data (Finn et al., 2015). Each point represents a subject's k (x-axis) and m (y-axis) estimates in a nonhierarchical (left) or hierarchical (right) framework. A point's color and shape signify the number of trials that subject completed, where the cyan circles are fewer than 50 trials, green squares are 50–69 trials, red pluses are 70–89 trials, and blue crosses are more than 90 trials. Row b shows the consistency between the nonhierarchical (x-axis) and hierarchical (y-axis) estimates for log-transformed k (left) and m (right). Each point represents mean estimates for a single subject. See the online article for the color version of this figure.

extreme values show the highest discrepancy, the estimates are still highly correlated. The correlations between the nonhierarchical and hierarchical log(k) estimates and log(m) estimates for the real data are 0.96 and 0.87, respectively.

Nonhierarchical and Hierarchical Constraint

When estimating latent parameters from real experimental data, we will never know the true parameter values. Therefore, we are not able to use measures of deviation between true and estimated values to evaluate the accuracy of the estimates as we did in the simulation studies. However, we are still able to compare the posteriors of the estimates to the priors. If the data are uninformative, they will not provide sufficient constraint and the posterior will subsequently resemble the prior.

Some of the nonhierarchical estimates for k and m had posterior means that are outside the typical range we would observe in most sets of data and subjects. For example, individual-level

estimates of k values are usually small, but the nonhierarchical model often had estimates much larger than 1, clustering at 5. Our results from the simulation study suggest that these large values may be the result of uncertainty, but when these individual estimates, Pr(data – individual), are uncertain, they can be constrained in the hierarchical model by the marginal Pr(individual – group) and prior Pr(group). To illustrate this phenomenon, Figure 3 shows the nonhierarchical (left) and hierarchical (right) joint k and m posteriors for one representative subject, Subject 42, who completed 31 trials total. The nonhierarchical posterior for k in Subject 42 is underconstrained and resembles the prior. This constraint problem was more prevalent in subjects with fewer trials, such as Subject 42, as both k and m were more reliably estimated in general with a larger number of trials. However, it is worth noting this nonhierarchical constraint problem also existed in some subjects who completed more trials. For all underconstrained subjects, the additional data provided by the hierarchical structure provided significantly more constraint on the posterior.

Accurate parameter estimation also has an impact on our understanding of the relationship between delay discounting and other variables of interest, such as externalizing psychopathology (EXT). We compared the correlations between the estimated parameters and the measures of EXT calculated in Finn et al. (2015) and found that the parameters estimated in a hierarchical framework had stronger correlations to EXT. The correlation between EXT and k was slightly higher for the hierarchical estimates ($\rho = 0.29$) than for the nonhierarchical estimates ($\rho = 0.24$). The discrepancy was larger for the correlations with m, where correlations between EXT and m were much more negative for the hierarchical estimates ($\rho = -0.20$) than for the nonhierarchical estimates ($\rho = -0.064$). These differences in correlation exemplify the effect that an estimation procedure can have on theoretical conclusions.

Discussion

Accurately assessing individual temporal discounting curves should be an obligation when generalizing to societal problems. Yet, we identified clear problems with the recoverability of many representative forms of temporal discounting profiles and found that these problems are exacerbated when using an experimental design with fewer trials. However, these problems need not be prohibitive for investigating societal impacts of impulsive behavior. Instead, we have provided a simple solution using hierarchical Bayesian estimation to effectively

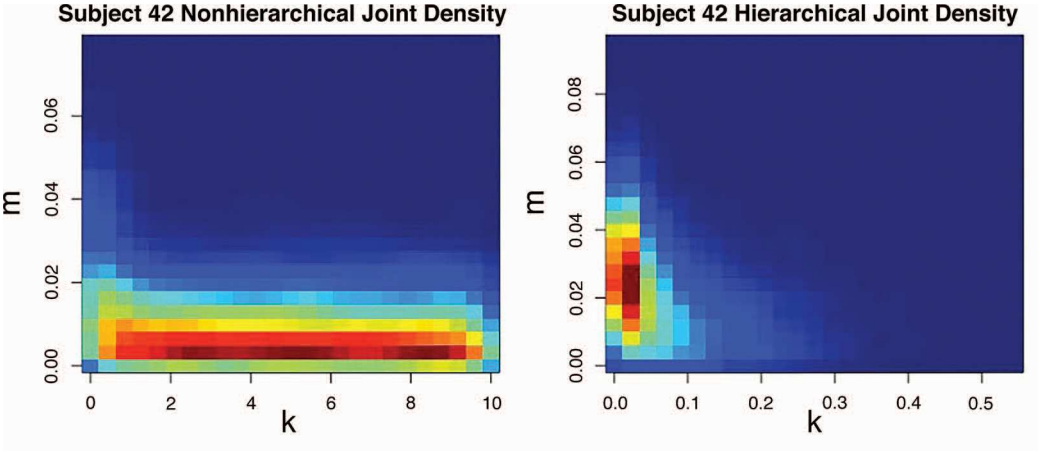

Figure 3. Hierarchical constraint. Shows the nonhierarchical and hierarchical joint posteriors. The marginal k (x-axis) and m (y-axis) posterior distributions for Subject 42 are plotted as joint kernel density plots for the nonhierarchical (left) and hierarchical (right) estimates. See the online article for the color version of this figure.

"pool" information across subjects. In using this approach, we have shown that the previously identified problematic temporal discounting profiles can be corrected.

Our findings have implications within model selection, design of intertemporal choice tasks, and subject exclusion criteria. First, accurate recovery is essential in modeling. While we only explored a simple two-parameter model, these results can be generalized to alternative-wise models. In the Appendix, we present analytical results that demonstrate why these misestimation problems occur with alternative-wise models more generally. In fact, as the Appendix shows, not even adding a utility parameter—which forms the most general intertemporal choice model—can alleviate this misestimation problem. We note more informative priors would help with constraining the model, even in the nonhierarchical case, to lead to more accurate estimation. For example, the uniform prior on k consists of a larger range of k values than those that are typically seen. However, while informative priors aid in estimation, they require stronger assumptions about these unknown parameters. Therefore, we would still recommend hierarchical models as they do not make these assumptions, instead providing a data-driven approach.

Additionally, misestimation can affect model selection criteria. For example, Ericson, White, Laibson, and Cohen (2015) had issues with cross-validation using the hyperbolic discounting model and therefore favored a heuristic model of intertemporal choice. These results could be explained by a failure to recover some k and m values, resulting in difficulties predicting out-of-sample data. Second, delay discounting parameters are often estimated using only 27 trials, especially in clinical contexts (Kirby et al., 1999). Our results suggest additional trials lead to more accurate estimation, so if possible, having at least two to three times more trials makes a significant difference in constraint and accuracy of estimates. Few trials may be inadequate due to the fact that the hyperbolic curve parameterized with different k and m presents similar functional forms for some pairs of rewards and delays. Having more trials can help distinguish between similar values of k and m. However, if this is not feasible, the hierarchical framework can compensate for the lack of data and provide significantly better estimates. Lastly, hierarchical models allow for the inclusion of more subjects. Even subjects who choose exclusively larger-later or smaller-sooner options are theoretically interesting, especially when studying impulsivity. By pooling information across the entire group, a more diverse sample can be studied. Overall, hierarchical Bayesian modeling can address many limitations in studying delay discounting.

References

Bailey, A., Gerst, K., & Finn, P. (2018). Delay discounting of losses and rewards in alcohol use disorder: The effect of working memory load. *Psychology of Addictive Behaviors, 32,* 197–204. http://dx.doi.org/10.1037/adb0000341

Becker, G. S., & Murphy, K. M. (1988). A theory of rational addiction. *Journal of Political Economy, 96,* 675–700. http://dx.doi.org/10.1086/261558

Bickel, W. K., & Marsch, L. A. (2001). Toward a behavioral economic understanding of drug dependence: Delay discounting processes. *Addiction, 96,* 73–86. http://dx.doi.org/10.1046/j.1360-0443.2001.961736.x

Bickel, W. K., Odum, A. L., & Madden, G. J. (1999). Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology, 146,* 447–454. http://dx.doi.org/10.1007/PL00005490

Chávez, M. E., Villalobos, E., Baroja, J. L., & Bouzas, A. (2017). Hierarchical Bayesian modeling of intertemporal choice. *Judgment and Decision Making, 12,* 19–28.

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

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General, 143,* 1489. http://dx.doi.org/10.1037/a0035976

Dai, J., Gunn, R. L., Gerst, K. R., Busemeyer, J. R., & Finn, P. R. (2016). A random utility model of delay discounting and its application to people with externalizing psychopathology. *Psychological Assessment, 28,* 1198. http://dx.doi.org/10.1037/pas0000248

Daugherty, J. R., & Brase, G. L. (2010). Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective. *Personality and Individual Differences, 48,* 202–207. http://dx.doi.org/10.1016/j.paid.2009.10.007

Ericson, K., White, J., Laibson, D., & Cohen, J. (2015). Money earlier or later? Simple heuristics

explain intertemporal choices better than delay discounting does. *Psychological Science, 26,* 826–833. http://dx.doi.org/10.1177/0956797615572232

Finn, P. R., Gunn, R. L., & Gerst, K. R. (2015). The effects of a working memory load on delay discounting in those with externalizing psychopathology. *Clinical Psychological Science: A Journal of the Association for Psychological Science, 3,* 202–214. http://dx.doi.org/10.1177/2167702614542279

Folland, G. B. (2013). *Real analysis: Modern techniques and their applications*. New York, NY: Wiley.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128,* 78–87. http://dx.doi.org/10.1037/0096-3445.128.1.78

Lancaster, K. (1963). An axiomatic theory of consumer time preference. *International Economic Review, 4,* 221–231. http://dx.doi.org/10.2307/2525488

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

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (pp. 1–10).

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 1446–1465. http://dx.doi.org/10.1037/a0013646

Ross, S. M. (1996). *Stochastic processes*. New York, NY: Wiley.

Scherbaum, S., Haber, P., Morley, K., Underhill, D., & Moustafa, A. A. (2018). Biased and less sensitive: A gamified approach to delay discounting in heroin addiction. *Journal of Clinical and Experimental Neuropsychology, 40,* 139–150. http://dx.doi.org/10.1080/13803395.2017.1324022

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

Turner, B. M., Rodriguez, C. A., Liu, Q., Molloy, M. F., Hoogendijk, M., & McClure, S. M. (2019). On the neural and mechanistic bases of self-control. *Cerebral Cortex, 29,* 732–750. http://dx.doi.org/10.1093/cercor/bhx355

Vincent, B. T. (2016). Hierarchical Bayesian estimation and hypothesis testing for delay discounting tasks. *Behavior Research Methods, 48,* 1608–1620. http://dx.doi.org/10.3758/s13428-015-0672-2

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

Appendix

The Limit Problem: Hyperbolic Case

To show the generalizability of our analyses, we demonstrate what happens to the hyperbolic model, and to alternative-wise models in general, as we take informative limits of the parameters in the model. The recovery issues of the hyperbolic model mainly reflect the fact that one parameter can "overshadow" another. By overshadow, we mean that one parameter can grow or shrink independently of another. This issue, which we call here the limit problem, leads to extreme nonidentifiability in the alternative-wise models.

Recall that the hyperbolic model is given by the following function:

$$H(x, t; k) = \frac{x^{\rho}}{1 + kt}, \tag{4}$$

where x is the objective amount, $t \geq 0$ is the time delay, $k \geq 0$ is our impulsivity parameter, and $\rho \geq 0$ is our utility parameter. As mentioned in the main text, we add a utility parameter here to make the analyses more general. The probability that a larger-later (LL) option will be chosen over a smaller-sooner (SS) option (which we shorten to $P(LL \gtrsim SS)$) is given by either a logistic function:

$$P_l(LL \gtrsim SS) = \frac{1}{1 + \exp(-mD)}, \tag{5}$$

or by a standard cumulative normal (CDF; Dai & Busemeyer, 2014):

$$P_N(LL \gtrsim SS) = \Phi\left(\frac{D}{\sigma}\right) = P(N \leq D/\sigma), \tag{6}$$

where $D = H(LL) - H(SS)$ is the difference in utility between the LL and SS options, σ indicates choice variability, and $N \sim \text{Normal}(0, 1)$. We interpret $D > 0$ to mean the LL option is preferred on average, and $D < 0$ means that the SS option is preferred on average. Note that by taking the appropriate limits, we can make the logistic model mimic the normal CDF model. Thus, for simplicity, we focus our limit arguments on the normal CDF. We emphasize, though, that model mimicry does not imply that the parameters that enact the mimicry are the same between models (e.g., there are more ways for the logistic model to converge to 1, say, than with the CDF model).

We note that Φ is a continuous function, since it is integration against another continuous function (Folland, 2013). That is, a "passage of the limit" through the cumulative distribution function is permissible: $\lim_{x \to x_0} \Phi(x) = \Phi(x_0)$ (Folland, 2013; Ross, 1996).

The most relevant limit to our discussion concerns the case when $k \to \infty$. Denote the LL option as ($y, s delay) and the SS option as ($x, t delay). The limit is given by (holding σ fixed):

$$\begin{aligned} &\lim_{k \to +\infty} \Phi\left(\frac{D}{\sigma}\right) \\ &\quad = \lim_{k \to +\infty} \Phi\left(\frac{\dfrac{y^{\rho}}{1 + ks} - \dfrac{x^{\rho}}{1 + kt}}{\sigma}\right) \to \Phi\left(\frac{0 - 0}{\sigma}\right) \\ &\quad = 1/2. \end{aligned} \tag{7}$$

Equation 7 states that the hyperbolic model "converges" to a random chance model as we increase k, meaning that the predicted choice is completely random as impulsivity increases, as expected. Figure A1 shows what a "likelihood surface" (a G^2 surface) would look like as a function of σ and k. For more details on how this figure was generated, see the online supplemental material. For this G^2 surface, a larger value indicates a poorer fit to the data; the convergence to the random chance model is

(*Appendix continues*)

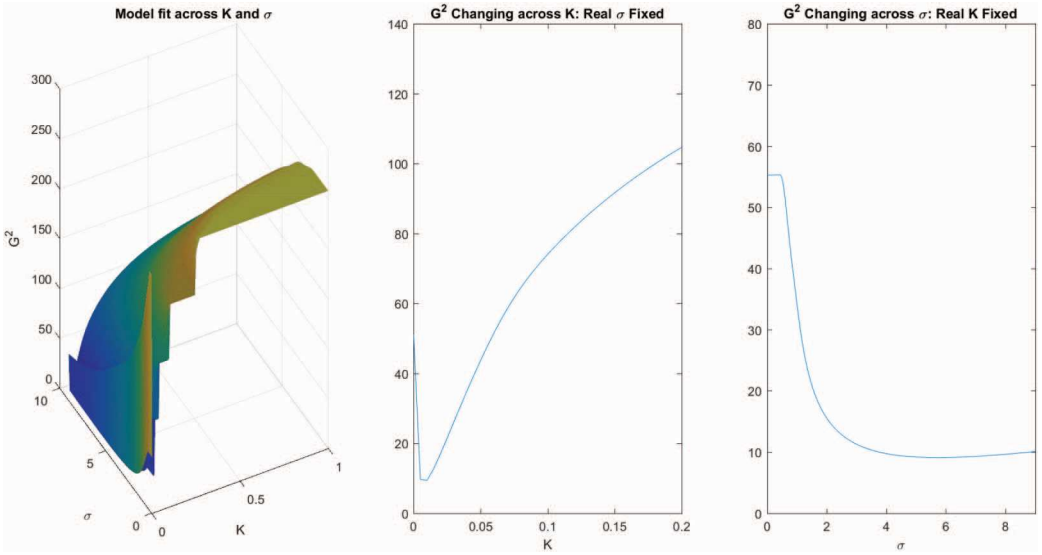

Figure A1. The G^2 surface for the (stochastic) hyperbolic model. The surface is based on simulated data generated from the median values of Dai et al. (2016). The middle graph shows the "slice" of the surface at the true σ value; the *k*-axis is zoomed in. The right-most panel shows the slice of the surface at the true *k* value. Note that while the fit becomes better for larger σ (lower G^2), the surface becomes flatter around the true value, leading to great deviations between the estimated value and the true value. See the online article for the color version of this figure.

noted by the progressively poorer fit as k increases. Note the steep climb toward an upper bound. The surface also flattens out, regardless of the $\sigma > 0$, for increasing k, so that larger parameter pairs will reproduce the same predicted data. This leads to the extreme parameter recovery issues.

The hierarchical approach ameliorates this issue by constraining our search (through a constrained prior) for k values to avoid getting stuck in the flat region, where the model approximates a random chance model.

The limit behavior of the hyperbolic model is not relegated only to the hyperbolic model. This generalizability arises from the convergence of the hyperbolic model to a random chance model. As we saw from before, with $k \to +\infty$, we have $D \to 0$. This is because the hyperbolic model is a decreasing function of *k:* With all other variables held constant, $H(x, t, k) \to 0$ as $k \to +\infty$. This simply reflects the interpretation of *k:* Greater impulsivity should lead to greater discounting of delayed rewards, with maximal discounting equivalent to ascribing zero subjective value to any delayed reward. This is the prediction for any alternative-wise discounting model. Therefore, these limit problems are not restricted to the hyperbolic model. This is because, as the difference in utility between the options goes to zero, $\sigma > 0$ does not change with k. After applying the limit theorem, the ratio will go to zero, making the probability (using either P_N or P_l) converge to 1/2. Importantly, this will happen for any alternative-wise model that requires subjective value to decrease with greater impulsivity (which is effectively all of them), and what this subjective value is becomes irrelevant for large k or σ. Hence, adding a utility parameter will not fix this convergence issue. In addition, simply changing the base model of the alternative-wise approach will not fix it either, as long as the choice variability and impulsivity parameters are independent of one another. Thus, the results of the main text apply to any alternative-wise model with independent parameters, not just the hyperbolic model.

Received March 18, 2019
Revision received January 31, 2020
Accepted February 24, 2020 ■