

# Time for Tea

## Discounting for Money and Consumption without the Utility Confound\*

Mohammed Abdellaoui<sup>1</sup>, Emmanuel Kemel<sup>1</sup>, Amma Panin<sup>2</sup>, and Ferdinand M. Vieider<sup>3</sup>

<sup>1</sup>*HEC Paris and CNRS, France*

<sup>2</sup>*University of Oxford, UK*

<sup>3</sup>*University of Reading, UK*

March 12, 2018

### Abstract

We conduct a field experiment in India comparing inter-temporal discount rates for money and the consumption of tea. We apply a novel method that allows us to directly access the discount function under discounted expected utility, while avoiding the utility confound that has bedeviled most previous studies. We find significantly more discounting for tea than for money. We find no difference in hyperbolicity. Analyzing the data in a Bayesian framework, we can indeed accept the null hypothesis of no hyperbolicity for either tea or money.

## 1 Motivation

Economists have long been interested in the extent to which economic actors may discount future consumption, and whether their discount functions may be constant, as originally postulated by [Samuelson \(1937\)](#). This interest has created an extensive database of experimental studies (see [Frederick, Loewenstein and O'Donoghue, 2002](#), for an overview). Those studies do, however, suffer from two main limitations. For one, they have mostly used monetary outcomes instead of consumption goods. Since theories on inter-temporal discounting are formulated based on consumption and not money, the use of monetary outcomes may have systematically distorted inferences if discounting for

---

\*The present study was financed by the German Science Foundation (DFG) under project VI 692/1-1 on “Risk preferences and economic behavior: Experimental evidence from the field”. We are grateful to Mary Philomena of SACRED for her support in the execution of this project. We would also like to thank Richie Ahuja, Director for Asia at the Environmental Defence Fund, for establishing the contact and making this project possible.

money and consumption are not the same. And second, much of this literature has used tradeoffs between different monetary amounts at different time periods and assumed linear utility to identify discounting. If utility is truly nonlinear, this may lead to systematic bias in the estimated discount functions (Chapman, 1996; Frederick et al., 2002).

We elicit time preferences for money and consumption of tea in a field experiment in India. Other than Augenblick, Niederle and Sprenger (2015), who compared decisions for money to decisions about a real effort task, we thus use two outcomes that fall both into the gain domain. Tea is interesting in this setting because it constitutes a luxury good on which poor households in India tend to over-spend (Banerjee and Duflo, 2011). We use a novel task that allows us to avoid the utility confound while eliciting a non-parametric discount function under discounted expected utility (*DEU*).<sup>1</sup> The task consists in offering a tradeoff between obtaining an outcome  $x$ , be it money or a consumption good, for sure at a later time  $\ell$ , or obtaining the same outcome  $x$  at an sooner time  $s$  with a probability  $p$ . By eliciting the probability that makes the decision maker indifferent between the two options, which we refer to as the *risk equivalent* of the time delay, we obtain an equation  $pD(s)u(x) = D(\ell)u(x)$ , where  $D$  is the discount function and  $u$  designates utility. After simplifying and rearranging, we thus obtain

$$p = \frac{D(\ell)}{D(s)}. \quad (1)$$

By setting  $s = 0$  we obtain  $D(s) = 1$ , so that we can now directly access the discount factor for a delay  $\ell$  from the present. By additionally using tasks with  $s > 0$ , we are able to assess whether this discount function is constant, or whether we observe hyperbolic behavior (Laibson, 1997; Rohde, 2010). By using outcomes constituted either by money or by amounts of tea with comparable monetary value, we can directly compare the discount functions for money and consumption.

We analyze the data in a Bayesian statistical framework. Such a framework has several advantages over a frequentist framework in our setup. For one, preference parameters are estimated as random quantities. This addresses concerns about the systematic distortion of preference parameters in structural estimations that have recently been raised in the literature (Apesteguia and Ballester, 2016; Bhatia and Loomes, 2017). These

---

<sup>1</sup>An alternative task to directly access the discount function under discounted utility has recently been proposed by Attema, Bleichrodt, Gao, Huang and Wakker (2016). The simplicity of our task, however, seems to make it better suited to investigations in field settings and with low-education subject populations.

contributions propose the use of random preference models as a solution to the concerns highlighted, and in our Bayesian setup preferences are random quantities by definition. Furthermore, it allows us to model the distribution of preferences parameters across our subject population in a hierarchical setup, thus allowing for individual heterogeneity, while avoiding the over-fitting of noisy individual observations typical of individual-level analysis (McElreath, 2016). Finally, Bayesian analysis makes it possible to accept the null hypothesis if the data support it (Kruschke, 2010; 2013)—a definite advantage in multi-parameter models, where some parameters may well be behaviourally irrelevant.

The present paper proceeds as follows. Section 2 describes the experiment and introduces our analytical methods. Section 3 presents the results. Section 4 fits our findings into the wider literature, and concludes the paper.

## 2 Experiment and Theory

*Stimuli and treatments.* We administered 6 choice lists per subject with each recurring twice—once for money, and once for tea—for a total of 12 choice lists. All choice lists offered a tradeoff between an amount  $x$  with a given probability in a sooner period, denoted  $s$ , and the same amount for sure in a later period, denoted  $\ell$ . The amount consisted of 400 Indian Rupees in the money condition, and of 2 kg of tea in the consumption condition (with a price of 400 Rs on the local market). 400 Rs corresponded to a daily wage of a male agricultural laborer at the time of the experiment. In order to elicit indifference, the probability of winning in the sooner period was varied in steps of 0.05 between 0 and 1. All choice problems were represented physically, laying out the monetary sums/amounts of tea to be won, and using colored balls to represent probabilities. The choice lists differed in terms of the time delays, containing the delay pairs  $(s, \ell) = \{(0, 2), (2, 4), (4, 6), (0, 3), (3, 6), (0, 6)\}$ . The treatment was administered within-subject, and at the end one choice was randomly selected for real pay—the standard procedure in this type of task.

*Participants and context.* We conducted the experiment in Ramanagara district, Karnataka state, India. The district has a size of approximately 50 by 100 km, and at the time of the experiment was home to slightly more than 2 million people. We first randomly sampled 9 villages from the rural villages in the district (excluding the larger

villages located close to the Bangalore-Mysore highway). In those villages, all households were invited to participate for ethical reasons, yielding a total sample of 946 individuals. Household members were chosen using a Kish grid. The experiments were administered in individual interviews by 20 enumerators. The enumerators were extensively trained on the tasks by the authors, and had the opportunity to accumulate experience in pilot sessions conducted in different villages. They were supervised in the field by a postdoc speaking the local language, who monitored randomly selected experimental sessions to ensure that the same procedures were followed by all the enumerators.

*Procedures and protocol.* Once a village had been selected, we contacted the village head to announce and explain the experiments. We then started by constructing a household roster based on lists available at the village level, and using identifying information from IDs and ration cards to obtain unique identifiers for each household. The initial roster contained a questionnaire, and subjects were paid for their participation. We used this occasion to announce the subsequent experiment, and to obtain written consent for the use of data from each participant. The payment was administered with a three day delay, using the procedure that was later to be used in the experiment (see below). This was meant to build trust in the fact that the future payments would truly be administered. Decisions in the choice tasks were recorded on laptop computers, although enumerators used physical devices to carefully explain the choices and outcomes, including envelopes representing cash and samples of tea in the appropriate weights. The order of questions was randomised within a given treatment, and the order of treatments was also randomised. At the end of the experiment, subjects randomly drew a number between 1 and 12 to select the payoff relevant question. After this, subjects randomly drew a number corresponding to one of the rows within the choice list. Their decision on this row was recorded for payment. The whole procedure from instructions to signing the payment certificates took the median subject just under an hour.

*Future payments.* Trust in future payoffs is crucial when conducting time experiments. We took several measures to ensure such trust. For one, the ‘immediate’ payoff was always administered with a three day delay, to avoid introducing differences in transaction costs. At the end of the experiment, the respondents received a certificate signed by the enumerator listing the amount to be paid, and the date at which the payment was due. The certificate contained the logo, the name, and the address of the office we maintained

in the district (the WZB India Field Office). It also contained the address and logo of a local NGO, SACRED, that was familiar to the participants, since they maintained operations supporting farming in the villages. The address and telephone numbers of both organizations was reported on the certificate, and participants were explicitly encouraged to get in touch in case they had any questions about the payments.

*Theoretical model.* We describe choices over prospects defined over streams of at most two outcomes. We are interested in modeling tradeoffs of the type  $(x_s, p; 0_s) \sim x_\ell$  obtained by varying the probability  $p$  to obtain indifference, where an agent is indifferent between the risky prospect,  $(x_s, p; 0)$ , that gives  $x$  with probability  $p$  or else 0 at time the sooner timer  $s$ , and the temporal prospect  $(x_\ell, 1; 0)$ , which gives the *same* outcome  $x$  for sure at time the later time  $\ell$ , which we simply designate by  $x_\ell$ . Assuming DEU, we can model such a tradeoff as  $pD(s)u(x) = D(\ell)u(x)$ , which simplifies to

$$p = \frac{D(\ell)}{D(s)}. \quad (2)$$

For the special case of  $s = 0$ , we thus can directly access the discount function, with  $D(\ell) = p$ , without any interference from utility, which drops out of the equation. Using two different time delays, one can easily identify a discount function, including potential non-constant elements of discounting.

*Analysis.* In addition to the nonparametric analysis, we fit discount functions to the data using a Bayesian statistical framework. We estimate a model  $p_{in} \sim \mathcal{N}(D(\ell)/D(s), \sigma^2)$ , where  $p_{in}$  is the risk equivalent elicited in choice list  $i$  from respondent  $n$ . We allow for heteroscedasticity in the error term by setting  $\sigma = \tau e^{\tau t(\ell-s)}$ . This explicitly takes account of subadditivity—the finding that annualized discount rates decrease in the time delay used to obtain them (Read, 2001)—and fits our data significantly better than a homoscedastic specification (WAIC of 3180.1 against 3265.9 for the homoscedastic version,  $p < 0.001$ ). The discount function  $D$  is thereby defined as follows:

$$D(t) = \begin{cases} 1 & \text{if } t = 0 \\ \beta e^{-rt} & \text{if } t > 0, \end{cases} \quad (3)$$

where  $r$  is the yearly discount rate, and  $0 < \beta \leq 1$  drops out of the equation when both  $s > 0$  and  $\ell > 0$ . Values of  $\beta$  smaller than 1 indicate quasi-hyperbolic behavior, whereby

future payoffs are discounted more heavily when compared to an immediate payoff than when compared to other future payoffs. We decompose the parameters to allow for treatment heterogeneity by letting  $r = r_0 + \alpha \times Tea + r_n$  and  $\beta = \beta_0 + \gamma \times Tea + \beta_n$ . The parameters  $\alpha$  and  $\gamma$  thus capture the aggregate treatment effect for  $r$  and  $\beta$  respectively, i.e. the degree to which the discount rate and quasi-hyperbolicity for tea differ from the parameters estimated for money;  $r_0$  and  $\beta_0$  capture the aggregate parameter estimates for money; and  $r_n$  and  $\beta_n$  are individual-level intercepts, which allow us to capture individual heterogeneity and to cluster observations. We also allow for heterogeneity in treatment effects across individuals by implementing a random slope model (see e.g. [Snijders and Bosker, 2012](#)). We thus let  $\alpha = \alpha_0 + \alpha_n$  and  $\gamma = \gamma_0 + \gamma_n$ , where  $\alpha_0$  and  $\gamma_0$  capture the average treatment effect in the sample, while  $\alpha_n$  and  $\gamma_n$  capture individual deviations from those aggregate estimates. We model the four individual-level parameters as following a multivariate normal distribution with mean 0 and a full variance-covariance matrix  $\Sigma$ . This hierarchical setup allows us to make full use of all the information at our disposal, while estimating correlations between the different parameters. Individual-level estimates are informed by endogenously-estimated priors provided by the aggregate parameter estimates, and explicitly modeling the covariance structure means that observations for one condition serve to inform estimations for the other as well ([McElreath, 2016](#)). For the aggregate parameters, we used diffuse, mildly regularizing priors, which only contain information about the likely range of the parameters, but do not restrict the estimates in any way—further details are provided in the Online Appendix. All estimations were run in Stan and launched from R using Rstan ([Stan Development Team, 2017](#)). We used 8000 iterations after warmup to obtain our results. Convergence of results was verified by visual inspection, and by examination of Gelman’s  $\hat{R}$  ([Gelman, Carlin, Stern, Dunson, Vehtari and Rubin, 2014](#)).

## 3 Results

### 3.1 Nonparametric analysis

We start by analysing the data non-parametrically. Table 1 shows summary statistics for the different REs and tests whether they are the same across treatments using matched sample signed-rank tests. All discount factors for tea can be seen to be lower than those for money. The difference is, however, only significant in two out of six cases at the

5% level, and in two more cases at the 10% level using nonparametric tests. Clearly, these tests are not independent from each other, and the more interesting question is whether the overall discount functions differ from each other. A first crude indication is given by the mean of the discount factors, shown in the last row of the table, which is indeed significantly smaller for tea than for money. A more appropriate test can be carried out by taking into account all observations, but clustering the errors at the subject level. Tests meeting these requirements indeed indicate significantly stronger discounting for tea than for money ( $t(935) = 4.09, p < 0.001$ , cluster-adjusted t-test;  $z = 26.27, p < 0.001$ , Somers' D with clustered errors).

**Table 1:** Risk equivalents for money and tea

$(s, \ell)$	money		tea		signed rank test	
	mean	sd	mean	sd	z stat.	p-value
(0, 3)	0.722	0.249	0.709	0.259	1.353	p=0.176
(0, 6)	0.679	0.273	0.649	0.273	2.404	p=0.016
(3, 6)	0.708	0.266	0.682	0.279	2.747	p=0.006
(0, 2)	0.708	0.265	0.697	0.271	1.127	p=0.260
(2, 4)	0.729	0.251	0.715	0.253	1.671	p=0.095
(4, 6)	0.721	0.253	0.699	0.273	1.869	p=0.062
mean	0.710	0.202	0.691	0.210	3.518	<0.001

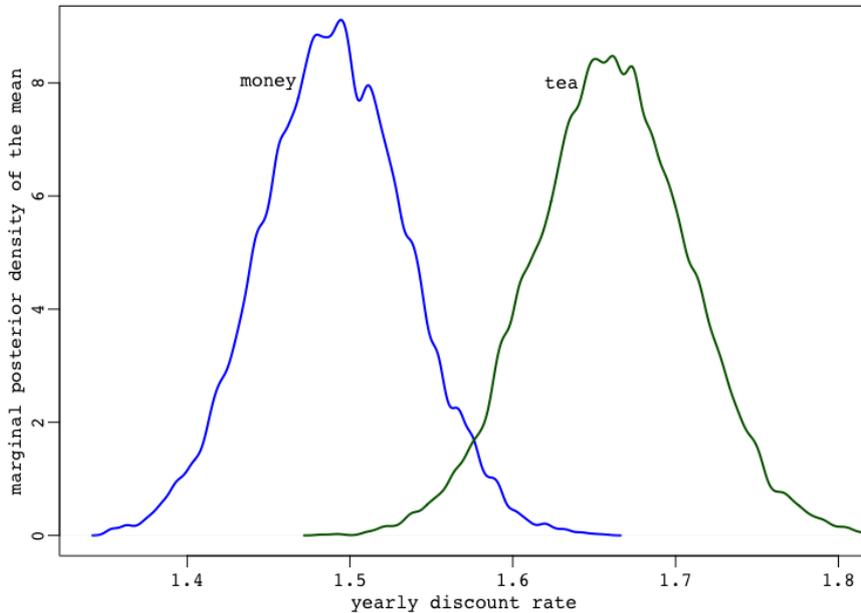
The statistical tests discussed above do not tell us much about the economic significance of the difference. Calculating the discount rate for a six months delay from the present, and extrapolating to an annual discount rate using quarterly compounding gives us a yearly discount rate for money of 117%. High discount rates are routinely found in developing countries (e.g., [Ashraf, Karlan and Yin, 2006](#), found 50% *monthly* discount rates in the Philippines), and may partially be due to the short measuring horizon ([Pender, 1996](#); [Read, 2001](#)). The equivalent discount rate for tea is 137% for the six months delay, and thus 20 percentage points higher than for money. We thus conclude that the difference is not only statistically significant, but also economically so.

Finally, we can take a look at whether discounting is constant or rather follows a quasi-hyperbolic or hyperbolic pattern. Constant discounting predicts that discounting depends only on the time delay, but not on the starting period. Designating discount rates for a given delay as  $\delta_{s\ell}$ , we would thus expect that  $\delta_{02} = \delta_{24} = \delta_{46}$ , and that  $\delta_{03} = \delta_{36}$ . In principle, we would also expect that  $\delta_{03} = \delta_{06}/\delta_{36}$ , although the latter equality may be contaminated by discounting depending on the length of the time delay

itself (Read, 2001). Testing these relations using matched sample tests, we find some differences to be significant. These differences, however, at times go in the expected direction of less discounting in choices involving up-front delays, and at times in the opposite direction. Structural estimations will tell us more about the aggregate effects.

### 3.2 DEU with quasi-hyperbolic discounting

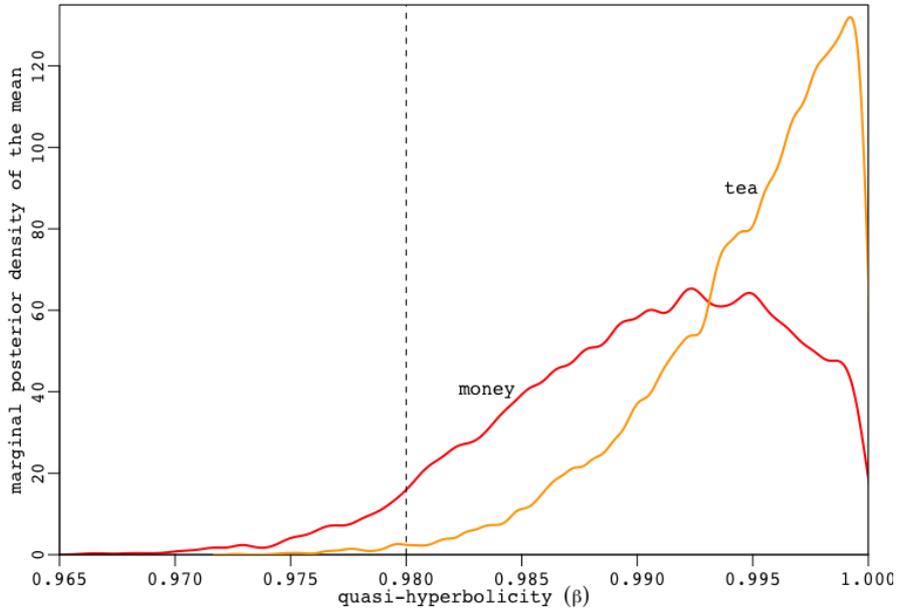
We now estimate a structural model of our data. Such an estimation will allow us to summarize the information contained in the discount factors in a small number of parameters. The absence of the utility confound carries over to these estimations, so that we only fit a discount function to the nonparametric observations.



**Figure 1:** Yearly discount rate: money vs. tea

Figure 1 shows the marginal posterior densities of the mean of the yearly discount rates  $r$  for the two conditions. These can be thought of as probability distributions expressing the uncertainty surrounding the mean parameter estimates. The yearly discount rate for tea is clearly larger than the discount rate for money, with means of 166% and 149% respectively. To see whether the two are significantly different from a statistical point of view, we need to take a look at the difference of the posterior densities, or equivalently, at the mean regression parameter on the tea treatment  $\alpha$ . The latter has a 95% highest density interval (*HDI*) ranging from 0.105 to 0.234, indicating that 95% of its

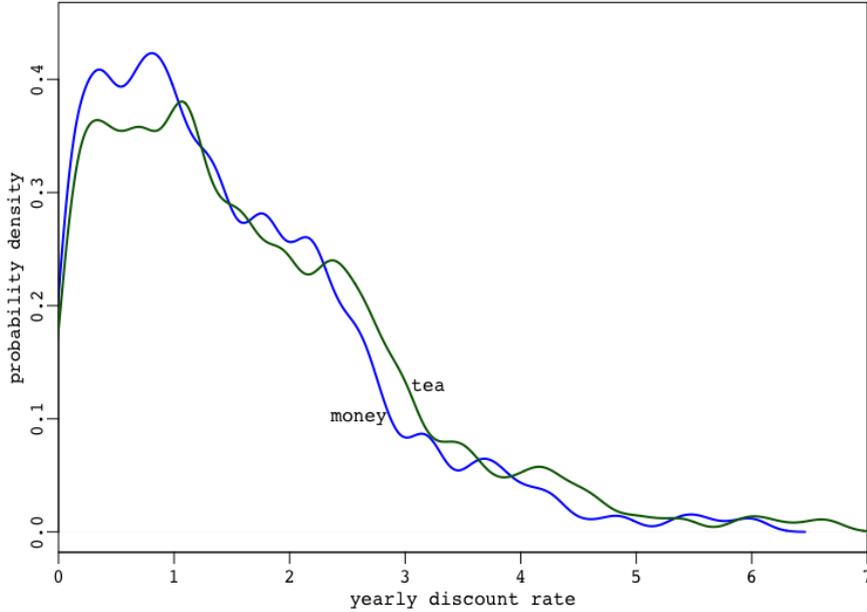
probability mass falls within that interval. None of its probability mass falls below zero, so that we can conclude with confidence that the discount rate for tea is significantly larger than the discount rate for money. With a mean difference of 17 percentage points, the difference is further significant not just statistically, but also economically.



**Figure 2:** Quasi-hyperbolicity parameter: money vs. tea

Figure 2 shows the marginal posterior densities of the means for the quasi-hyperbolicity parameter  $\beta$  by treatment. Two things stand out. First, the two hyperbolicity parameters are not significantly different from each other (the parameter capturing their difference has a 95% HDI ranging from  $-0.01$  to  $0.02$ ). Second, there is little hyperbolicity in the data in general. Indeed, almost all the probability mass falls on values larger than 0.98. We can thus conclude that, at least from an economic point of view, quasi-hyperbolicity is negligible in our data. Statistically, our Bayesian setup allows us to accept the null for a given parameter if there is enough support for the null hypothesis (Kruschke, 2010; 2013). In order to test this, we establish a region of practical equivalence—a region that may be practically equivalent to the limiting value of 1—ranging from 0.98 to 1. For money, we find 95% of the probability mass to fall into that region. For tea, the equivalent figure is 99%. We thus accept the null hypothesis that quasi-hyperbolicity does not occur in our data for both money and tea.

We have so far focused on aggregate parameters only. Indeed, such a focus remains

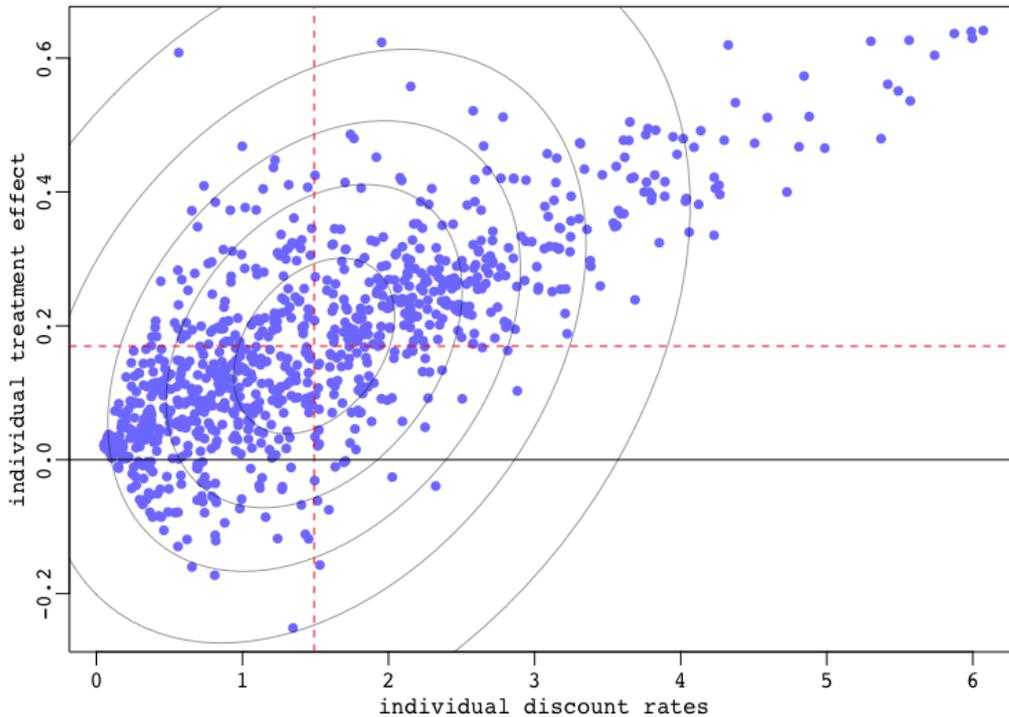


**Figure 3:** Distribution of individual-level discount parameters: money vs. tea

the standard in the literature. Our hierarchical Bayesian approach, however, allows us to obtain individual-level estimates of the parameters that are informed by the aggregate distribution, thus avoiding the danger of over-fitting noisy individual data. To the extent that observations lie very far from the mean, and to the extent that they are characterized by high levels of noise, such observations will be ‘shrunk’ towards the mean by the hierarchical priors. Figure 3 shows the distribution of the individual-level discount rates. The distribution can be seen to be highly skewed. While the mean individual parameters for money and tea are 1.49 and 1.66 respectively, the medians are considerably lower at 1.26 and 1.41. For money, 41% of all subject-level parameters fall below a discount rate of 100%, and 21% fall below 50%. For tea, the equivalent numbers are 35% and 18%. The two distributions are significantly different at any level of significance using either a paired t-test or a Wilcoxon signed rank test.

Effects for money and tea are also highly correlated. Figure 4 shows a scatter plot of the individual discount rates for money against the individual treatment effects, i.e. differences in behavior when tea is at stake instead of money. The elliptical contour lines represent the variance-covariance matrix of the two parameters<sup>2</sup>, while the dashed hor-

<sup>2</sup>In our Bayesian setup, these can again be interpreted in a probabilistic way. The horizontal distance between the contour lines can thus be interpreted as the variance in discount rates for money. Seen from the center outwards, the contour lines correspond to 10%, 30%, 50%, 70%, and 90% of the variance. The vertical distance has a similar interpretation for the regression coefficient  $\alpha$  capturing the treatment



**Figure 4:** Correlation plot of random intercepts (discount rates) and random slopes (treatment effects)

horizontal and vertical lines indicate the mean estimates. Individuals with higher discount rates also tend to have larger treatment effects for tea ( $\rho = 0.328$ , 95% HDI of the correlation coefficient  $[0.070, 0.598]$ ). This shows that monetary experiments may result in the estimation of distortions in discounting that are not only large, but also systematic. Indeed, beyond a level effect, this correlation highlights a distributional effect whereby we observe increased heterogeneity for tea as compared to money.

## 4 Conclusion

The experimental literature in economics has long investigated and measured time preferences using monetary outcomes. The original models on time preferences are, however, formulated in terms of consumption, and the use of money may be problematic inasmuch money can easily be borrowed and saved. Experimentally elicited discount rates may furthermore have been distorted by the assumption of linear utility (Chapman, 1996; Frederick et al., 2002). We devised a method that allows us to nonparametrically access

---

effect of using tea instead of money. Extensions along the diagonal capture the correlation structure of the data—see Online Appendix for a formal definition of the variance-covariance matrix.

the discount function under discounted expected utility. By using a fixed amount of either money or tea as outcomes, we could thus obtain two perfectly comparable discount functions for money and consumption that are unaffected by the utility confound. We found significantly more discounting for tea than for money, but no difference in hyperbolicity, which we found to play no role in our data. The result thus showcases that using money instead of consumption goods is not an innocent assumption, and may systematically distort inferences on discounting.

Our paper significantly adds to recent attempts to quantify the difference between discounting for money and consumption. [Reuben, Sapienza and Zingales \(2010\)](#) compared discounting for money to discounting for chocolate, finding higher discount rates for chocolate. Their conclusions are, however, predicated on the assumption of linear utility (i.e., no desire for inter-temporal smoothing of consumption). [Augenblick et al. \(2015\)](#) investigated the stationarity of time preferences for money and effort in a dynamic experiment where subjects could revise their choices, and found hyperbolic patterns for effort, but not for money. While they did control for utility curvature in their setup, they could do so only at the aggregate level, assuming that everybody has the same utility. The test presented furthermore combined a gain frame (monetary payoffs) with a loss frame (effort provision). This constitutes a confound, given that time preferences have long been known to differ for gains and losses ([Loewenstein and Prelec, 1992](#); [Frederick et al., 2002](#)). Other than in that paper, we compared two consumption goods (instead of one good and one effort provision task). [Ubfal \(2016\)](#) compared discounting for 19 different goods in a field experiment in Uganda, finding higher levels of discounting for some of them. He could, however, not directly account for utility curvature and used mostly hypothetical incentives. In his subsample using real incentives, only 3 effects were found to be significant at the 10% level—close to what we would expect to observe by chance given the large number of tests.

Our results show clearly that consumption goods may trigger higher levels of discounting than money, even once utility curvature is properly accounted for. One open question is the extent to which these results would generalize to other types of consumption goods. We chose tea because it constitutes a luxury good that is highly valued by our Indian households, and on which cash-strapped households in India tend to overspend relatively to other consumption goods. It does not seem unreasonable to assume that discounting may further increase for goods that are less storable than tea. Using

the risk-equivalents here presented, it is straightforward to test such hypotheses in future research.

## References

- Apestequia, Jose, and Miguel A Ballester (2016) ‘Monotone stochastic choice models: The case of risk and time preferences.’ *Journal of Political Economy*
- Ashraf, Nava, Dean Karlan, and Wesley Yin (2006) ‘Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines.’ *The Quarterly Journal of Economics* 121(2), 635–672
- Attema, Arthur E., Han Bleichrodt, Yu Gao, Zhenxing Huang, and Peter P. Wakker (2016) ‘Measuring Discounting without Measuring Utility.’ *American Economic Review* 106(6), 1476–1494
- Augenblick, Ned, Muriel Niederle, and Charles Sprenger (2015) ‘Working over Time: Dynamic Inconsistency in Real Effort Tasks.’ *The Quarterly Journal of Economics* p. qjv020
- Banerjee, Abhijit V., and Ester Duflo (2011) *Poor Economics. A Radical Rethink of the Way to Fight Global Poverty.* (New York: Public Affairs)
- Bhatia, Sudeep, and Graham Loomes (2017) ‘Noisy preferences in risky choice: a cautionary note.’ *Psychological Review*
- Chapman, G B (1996) ‘Temporal discounting and utility for health and money.’ *Journal of experimental psychology. Learning, memory, and cognition* 22(3), 771–791
- 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
- Gelman, Andrew (2006) ‘Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper).’ *Bayesian analysis* 1(3), 515–534
- Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin (2014) *Bayesian data analysis*, vol. 2 (CRC press Boca Raton, FL)
- Kruschke, John K (2010) ‘Bayesian data analysis.’ *Wiley Interdisciplinary Reviews: Cognitive Science* 1(5), 658–676
- (2013) ‘Bayesian estimation supersedes the t-test.’ *Journal of Experimental Psychology: General* 142(2), 573
- Laibson, David (1997) ‘Golden Eggs and Hyperbolic Discounting.’ *The Quarterly Journal of Economics* 112(2), 443–478
- Loewenstein, George, and Drazen Prelec (1992) ‘Anomalies in Intertemporal Choice:

- Evidence and an Interpretation.’ *The Quarterly Journal of Economics* 107(2), 573–597
- McElreath, Richard (2016) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (Academic Press)
- Pender, John L. (1996) ‘Discount rates and credit markets: Theory and evidence from rural india.’ *Journal of Development Economics* 50(2), 257–296
- Read, Daniel (2001) ‘Is Time-Discounting Hyperbolic or Subadditive?’ *Journal of Risk and Uncertainty* 23(1), 5–32
- Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2010) ‘Time discounting for primary and monetary rewards.’ *Economics Letters* 106(2), 125–127
- Rohde, Kirsten I. M. (2010) ‘The hyperbolic factor: A measure of time inconsistency.’ *Journal of Risk and Uncertainty* 41(2), 125–140
- Samuelson, Paul A. (1937) ‘A Note on Measurement of Utility.’ *The Review of Economic Studies* 4(2), 155–161
- Snijders, Tom A. B., and Roel J. Bosker (2012) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd ed. (London: Sage Publications)
- Stan Development Team (2017) ‘RStan: the R interface to Stan.’ R package version 2.17.2
- Ubfal, Diego (2016) ‘How general are time preferences? Eliciting good-specific discount rates.’ *Journal of Development Economics* 118, 150–170

## ONLINE APPENDIX

### A The Bayesian model

#### The variance-covariance matrix

The individual-level parameters (random intercepts and random slopes) are modelled as following multi-variate normal distribution:

$$\begin{pmatrix} r_n \\ \beta_n \\ \alpha_n \\ \gamma_n \end{pmatrix} \sim \mathcal{N} \left[ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_r^2 & 0 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & \sigma_\alpha^2 & 0 \\ 0 & 0 & 0 & \sigma_\gamma^2 \end{pmatrix} \mathbf{R} \begin{pmatrix} \sigma_r^2 & 0 & 0 & 0 \\ 0 & \sigma_\beta^2 & 0 & 0 \\ 0 & 0 & \sigma_\alpha^2 & 0 \\ 0 & 0 & 0 & \sigma_\gamma^2 \end{pmatrix} \right]$$

where  $\sigma_r^2$  is the variance of the discount rate, and similarly for the other variance parameters. The matrix  $\mathbf{R}$  is a  $4 \times 4$  matrix of correlation coefficients, which takes the following form:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho_{r\beta} & \rho_{r\alpha} & \rho_{r\gamma} \\ \rho_{r\beta} & 1 & \rho_{\beta\alpha} & \rho_{\beta\gamma} \\ \rho_{r\alpha} & \rho_{\beta\alpha} & 1 & \rho_{\alpha\gamma} \\ \rho_{r\gamma} & \rho_{\beta\gamma} & \rho_{\alpha\gamma} & 1 \end{pmatrix}$$

#### The hyperpriors

We use only mildly regularizing priors throughout. Such priors provide an indication of possible parameter ranges, while not restricting the exploration in any meaningful way. Given the large amount of data at our disposal, the choice of prior is easily overwhelmed by the information provided therein, so that our estimates are resistant to a wide variety of priors. In the estimations reported in the paper, we have adopted the following priors:

$$\begin{aligned} r_0 &\sim \mathcal{N}(1, 1) \\ \alpha_0 &\sim \mathcal{N}(0, 0.5) \\ \gamma_0 &\sim \mathcal{N}(0, 0.5) \end{aligned}$$

$$\tau \sim \text{HalfCauchy}(0, 0.5)$$

$$\tau_t \sim \text{HalfCauchy}(0, 10)$$

We follow diffuse normal priors for the estimation of our main parameters because they render the Bayesian algorithm more efficient. The half-cauchy priors for the distributional parameters follow best practices ([Gelman, 2006](#)). For the matrix of correlation coefficients, we use the LKJcorrelation prior proposed by [McElreath \(2016\)](#).