Economica

*Economica* (2013) **80**, 670–697
doi:10.1111/ecca.12028

# Discounting Behaviour and the Magnitude Effect: Evidence from a Field Experiment in Denmark

By STEFFEN ANDERSEN†, GLENN W. HARRISON‡, MORTEN I. LAU†† and E. ELISABET RUTSTRÖM‡

†*Copenhagen Business School* ‡*Georgia State University, USA* ††*Copenhagen Business School and Durham University Business School*

Final version received 20 March 2013.

We evaluate the claim that individuals exhibit a magnitude effect in their discounting behaviour, where higher discount rates are inferred from choices made with lower principals, all else being equal. If the magnitude effect is quantitatively significant, it is not appropriate to use one discount rate that is independent of the scale of the project for cost–benefit analysis and capital budgeting. Using data from a field experiment in Denmark, we find statistically significant evidence of a magnitude effect that is *much* smaller than is claimed. This evidence surfaces only if one controls for unobserved individual heterogeneity in the population.

## INTRODUCTION

One anomaly about discounting behaviour is known as the 'magnitude effect': the finding that smaller amounts are discounted more than larger amounts. The magnitude effect has been presented as an explanation for declining discount rates with horizon, and would suggest that these differences should disappear if principals were larger. If magnitude effects are sufficiently large, as the literature is suggesting, then they may account for behaviour otherwise attributed to quasi-hyperbolic preferences. Moreover, if the magnitude effect is significant, then it is not appropriate to use one discount rate or discounting function, independent of the scale of the project, for the purposes of cost–benefit analysis and capital budgeting. We present new evidence on the magnitude effect from a methodology that is based on real rewards, transparent elicitation procedures that are incentive compatible, joint estimation of risk and time preferences, and allowing for unobservable subject heterogeneity when inferring discount rates.

One theory that predicts magnitude effects hypothesizes that individuals might be using a fixed monetary premium to decide whether to choose delayed payments rather than earlier payments, as well as some premium for delay that varies with horizon (Benhabib *et al.* 2010). In this case a subject might want to receive a minimum of $10 before delaying receipt; as the principal increases from $100 to $1000, say, the discount rate at which the subject switches from the sooner to the later payment decreases, since $10 is proportionately less as the principal increases. This theory is important because it suggests that evidence for sharply declining discount rates with horizon would disappear as the principal gets larger and larger. A direct test of the magnitude effect presents itself in this theory formulation: present subjects with two different principals and see if there is a difference in behaviour.[1]

We conduct artefactual field experiments with a representative sample of adult Danes to see if there is anything here that should be of concern to economists. Harrison and List (2004) define artefactual field experiments as experiments that use the type of stimuli that one would normally find in a laboratory environment but with a subject pool drawn from the field. Such experiments represent the first step from the laboratory to the field, and generally afford the greatest controls in terms of internal validity. For the purpose of testing if some behavioural anomaly from the laboratory applies in the field, which is the case here, artefactual field experiments are particularly valuable since one is changing only the subject pool. In our case the sample is representative of the adult Danish population, and is not specially drawn from some field setting that might bring with it some unique characteristics (e.g. a church or a convention of actuaries).

We find statistically significant evidence for the magnitude effect in our field sample of adult Danes. However, this evidence arises only after one carefully controls for unobserved individual heterogeneity, and could easily be missed if one assumed homogeneous preferences. We consider time horizons between 2 weeks and 12 months, and report a mean *annual* discount rate of 11.3% for the low principal of 1500 Danish kroner and 7.9% for the high principal of 3000 kroner. The most striking finding is that the evidence for the magnitude effect is at levels that are much lower than previously reported. For example, using a convenience sample of college students, Kirby (1997) looks at time horizons between 1 and 29 days, and reports *daily* discount rates of 5.6% and 3.6% for principals of \$10 and \$20, respectively.[2] Using daily compounding, the implied annual discount rates are 43,384,054,919% and 40,393,329%, which are huge by any standard. These discount rates imply a large magnitude effect where the former is over 1000 times larger than the latter. Halevy (2012) considers principals of \$10 and \$100, and a time horizon of one week, and he finds *weekly* discount rates of 5.9% and 3.7%, respectively. Using weekly compounding, the implied discount rates are 1871% and 561%. While the magnitude effect here is significantly smaller than in Kirby (1997), it is still large, with the former discount rate more than three times the latter. None of the other studies with real monetary rewards report estimated discount rates. Our review of studies using hypothetical rewards shows varying quantitative magnitude effects, but most are many times larger than what we find here. We report a magnitude effect of only 3.4 percentage points when comparing means of discount rate distributions, which is tiny in comparison to the previous literature. When we, more appropriately, compare medians, the effect almost goes away. Hence one could argue that although the magnitude effect is statistically significant in our study, it is not economically significant over a time horizon of one year.

We review the existing literature on magnitude effects in Section I, and present the experimental design in Section II. In Section III we specify a structural econometric model to infer discount rates, and in Section IV we report our findings. Section V concludes.

## I. Literature Review

Table 1 contains a summary tabulation of procedures and findings in the literature on magnitude effects. We carefully review the most important contributions here, and every other paper in Appendix A (available from the authors on request). The magnitude effect is a robust finding in the studies that are listed in Table 1, and appears in studies that rely on hypothetical payoffs as well as real monetary rewards. Previous studies that rely on monetary rewards consider the income range between \$10 and \$85, and time horizons between 1 day and 6 months, whereas studies based on hypothetical questions consider higher stakes (up to \$100,000) and longer horizons (up to 20 years). We concentrate our review on studies with real monetary rewards, but also discuss the earliest papers on

TABLE 1
REVIEW OF EXPERIMENTAL LITERATURE ON THE MAGNITUDE EFFECT

| Study | Sample (size) | Elicitation method | Horizon(s) | Principal amounts | Discount rates[a] | Hypothetical or real? |
|---|---|---|---|---|---|---|
| Thaler (1981) | Students ($N \approx 60$) | Open-ended Fill-in-the-blank | 3, 6, 12, 36, 60 & 120 months | \$15, \$75, \$250, \$1200, \$3000 | (120%, 139%), 98%, (34%, 69%, 120%), 69%, 29% for 12 months | Hypothetical |
| Ainslie and Haendel (1983) | Patients ($N$ = 18, 66 choices) | Choice | 3 days | \$2 up to \$10 | Not reported by magnitude | Real |
| Loewenstein (1987) | Students ($N$ = 30) | Open-ended Fill-in-the-blank | 3 hours, 1 & 3 days, 1 & 10 years | \$4 loss and \$1000 loss | 8772%, 3395%, 115%, 62%, 11% for \$4, 8772%, 1469%, 25%, 9%. 4% for \$1000 | Hypothetical |
| Benzion, Rapoport and Yagil (1989) | Economics and finance students ($N$ = 204) | Open-ended Fill-in-the-blank | 6 months, 1, 2 & 4 years | \$40, \$200, \$1000 and \$5000 | 46%, 32%, 29%, 19% for 6 months; 18%, 15%, 15%, 11% for 4 years (averaged over scenarios) | Hypothetical |
| Holcomb and Nelson (1992) | Business students ($N$ = 101) | Binary choice | 1, 7 & 14 days | \$5 and \$17 | Impossible to determine, but fraction choosing to delay imply magnitude effect | Real |

Table 1
Continued

| Study | Sample (size) | Elicitation method | Horizon(s) | Principal amounts | Discount rates[a] | Hypothetical or real? |
|---|---|---|---|---|---|---|
| Raineri and Rachlin (1993) | Psychology course requirement ($N$ = 120) | Titration | 1, 6 & 12 months (also 1, 5, 10, 25 & 50 years) | \$100, \$10,000 and \$1 million | 127%, 71%, 54% for \$100; 12%, 32%, 29% for \$10,000; 12%, 8%, 11% for \$1 million | Hypothetical |
| Shelley (1993) | Accounting students ($N$ = 74) | Open-ended Fill-in-the-blank | 6 months, 1, 2 & 4 years | \$40, \$200, \$1000 and \$5000 | 24%, 15%, 14%, 14% for \$40, \$200, \$1000 and \$5000, respectively, in the neutral frame; no ma gnitude effect for the payment or expedite frames | Hypothetical |
| Green, ristoe and Myerson (1994) | Students ($N$ = 24) | Binary choice with titration on delays | 1 week, 1 & 3 months | \$20 and \$50, \$100 and \$500, \$500 and \$1250 | 100%, 71% and 30% | Hypothetical |
| Green, Fry and Myerson (1994) | Young and old adults ($N$ = 12 each group) | Titration | 1 week, 1 & 6 months, 1, 3, 5, 10 & 25 years | \$1000 and \$10,000 | Young adults: 63%, 54%, 46%, 32%, 25% for \$1000; 40%, 34%, 29%, 19%, 15% for \$10,000 | Hypothetical |

TABLE 1
CONTINUED

| Study | Sample (size) | Elicitation method | Horizon(s) | Principal amounts | Discount rates[a] | Hypothetical or real? |
|---|---|---|---|---|---|---|
| Chapman and Elstein (1995) | Psychology students ($N = 70$) | Open-ended Fill-in-the-blank | 6 months, 1, 2 & 4 years | \$200, \$1000, \$5000 and \$25,000 | For 6 months (\$200, 400%), (\$1000, 210%), (\$5000 and \$25,000, 75%); for 1 year (\$200, 200%), (\$1000, 125%) | Hypothetical |
| | Psychology students ($N = 34$) | | 1, 3, 6 & 12 years | \$500, \$1000, \$2000 and \$4000 | For 1 year (\$500, 150%), (\$1000, 125%), (\$2000 and \$4000, 50%) | Hypothetical |
| Kirby and Maraković (1995) | Students ($N = 21$) | Open-ended Sealed-bid auction | 3–29 days | \$14.50 and \$28.50 | Not reported by magnitude | Real |
| | Students ($N = 18$) | | 3–29 days | \$14.50 and \$28.50 | Not reported by magnitude | Hypothetical |
| Chapman (1996) | Students ($N = 40$) | Fill-in-the-blank | 1, 3, 6 & 12 years | \$500, \$1000, \$2000 and \$4000 | For 1 year 115%, 115%, 75%, 75%, respectively; for 3 years 60%, 50%, 40%, 40% respectively | Hypothetical |
| | Students ($N = 77$) | Fill-in-the-blank | 1 & 9 years | Gains and losses of \$500, \$1500 and \$4500 | For 1 year gain of principal 130%, 55%, 55% respectively, and for a loss of principal 60%, 25%, 25% | Hypothetical |

TABLE 1
CONTINUED

| Study | Sample (size) | Elicitation method | Horizon(s) | Principal amounts | Discount rates[a] | Hypothetical or real? |
|---|---|---|---|---|---|---|
| Kirby and Maraković (1996) | Students ($N$ = 621) | Choice | 10–70 days | \$16 up to \$83 | Not reported by magnitude | 'Almost hypothetical' |
| Kirby (1997) | Students and others ($N$ = 24, 28, 20) | Bidding | 1–29 days | \$10 or \$20 in the future | *Daily* rates of 5.6% (3.6%), 2.4% (1.4%), 3.5% (2.4%) for each sample and the \$10 (\$20) magnitude, respectively | Real |
| Green, Myerson and McFadden (1997) | Students ($N$ = 20) | Binary choice with titration on delays | 3 & 6 months, 1, 3, 5, 10 & 20 years | \$100, \$2000, \$25,000 (and \$100,000 in the future) | \$100 principal: 75%, 50%, 39%, 17%, 18%, 10% and 8% by horizon; \$2000 principal: 21%, 19%, 12%, 10%, 12%, 10% and 7%; \$25,000 principal: 12%, 10%, 9%, 8%, 8%, 5% and 6% | Hypothetical |

TABLE 1
CONTINUED

| Study | Sample (size) | Elicitation method | Horizon(s) | Principal amounts | Discount rates[a] | Hypothetical or real? |
|---|---|---|---|---|---|---|
| Chapman and Winquist (1998) | Students ($N$ = 50) | Open-ended Fill-in-the-blank | 3 months | Taxi ride: \$5, \$10, \$15, \$20. Haircut: \$7.50, \$15, \$30, \$60. Restaurant dinner: \$20, \$40, \$80, \$160. | Values inferred from eyeballing graphs; roughly 5000% at \$5, slightly over 2000% for \$20, just under 2000% at \$30, and slightly in excess of 1000% at \$160 | Hypothetical |
| Kirby, Petry and Bickel (1999) | Adults in control ($N$ = 60) | Choice | 7–186 days | (\$25 to \$35), (\$50 to \$60) and (\$75 to \$85) | Small but statistically significant magnitude effect: no values reported | Real |
| Du, Green and Myerson (2002) | Chinese, Japanese and American students ($N$ = 79) | Titration | 1, 3 & 9 months, 2, 5, 10 & 20 years | \$200 and \$10,000 | Values reported graphically, impossible to infer numerically | Hypothetical |
| Estle, Green, Myerson and Holt (2007) | Students ($N$ = 47) | Titration | 1 week, 1 & 6 months, 1 & 3 years | \$40 and \$100 | ANOVA analysis; magnitudes not reported | Hypothetical |

TABLE 1
CONTINUED

| Study | Sample (size) | Elicitation method | Horizon(s) | Principal amounts | Discount rates[a] | Hypothetical or real? |
|---|---|---|---|---|---|---|
| Benhabib, Bisin and Schotter (2010) | Students ($N$ = 27) | Open-ended Fill-in-the-blank | 3 days, 1 & 2 weeks, 1, 3 & 6 months | \$1, \$4, \$8, \$15, \$40, \$80 and many interim values | Not reported by magnitude: inspection of choices for 3 subjects (their Figure 1) broadly consistent with effect | Real |
| Scholten and Read (2010) | Adults on web ($N$ = 196) | Choice | 16, 19 & 22 months | \$25, \$50, \$100 and \$250 | Not reported by magnitude | Hypothetical |
| Halevy (2012) | Students ($N$ = 149) | Choice | 1 week | \$10 and \$100 | Weekly rate of 5.9% for \$10 and 3.7% for \$100 | Real |

[a]Discount rates are reported in order of stake magnitudes. Values in parentheses are alternative rates for the same magnitude.

magnitude effects that rely on hypothetical questions, and studies that allow for non-linear utility functions.

The early papers on magnitude effects use open-ended elicitation methods where subjects are asked to state the delayed amount that they would require to be willing to give up a certain amount today. For example, Thaler (1981) considered a $3 \times 3$ design, varying the initial amount and the time delay. For one group of subjects, the amounts were \$15, \$250 and \$3000, and the delays were 3 months, 1 year and 3 years; for another group, the amounts were \$75, \$250 and \$1200, and the delays were 6 months, 1 year and 5 years; and for the final group, the amounts were \$15, \$250 and \$3000, and the delays were 1 month, 1 year and 10 years.[3] All choices were hypothetical, and the sample consisted of US college students. Median responses are reported, with no tests of statistical significance. Table 1 tabulates these results, converted to the annual effective rate assuming daily compounding, and supports the claim of magnitude effects.[4] Loewenstein (1987) and Benzion *et al.* (1989) apply similar elicitation methods and find support for the magnitude effect between losses of \$1000 and \$4, and between gains of \$40 and \$5000. There is substantial variation in the time horizon across the treatments in these studies, and the magnitude effect is robust to variation in time delay, except for the 3-hour horizon in Loewenstein (1987).

Holcomb and Nelson (1992) employ procedures that anticipate the more recent application of modern experimental methods to the elicitation of discount rates, and indeed this work is heavily cited by Coller and Williams (1999) in their seminal study. Holcomb and Nelson (1992) used 101 business students in a series of 41 binary choice tasks, where the sooner and later amounts were presented as given, and the subject simply needed to pick one. Treatments included the use of front-end delays on the sooner option of 0 days, 1 day or 7 days, the use of a low principal of \$5 or a high principal of \$17, and the use of a low daily interest rate of 1.5% or a high daily interest rate of 3.0%. One of the 41 choices by each subject was selected for payment at the end of the session, so this appears to be the first study to examine the magnitude effect with real rewards and incentive-compatible elicitation procedures. Focusing on their results with no front-end day and the higher interest rate, which are representative in this case, they observe that 23%, 41% and 56% of the subjects chose to delay with the principal of \$5 and horizons of 1 day, 1 week and 2 weeks, respectively, whereas 53%, 79% and 87% chose to delay for the corresponding horizons when the principal was \$17. Precise discount rates cannot be inferred from this design without strong parametric assumptions, since we know only that the discount rate was in the interval [0%, 1.5%), [1.5%, 3.0%) or [3%, ∞%). But the fractions reported are enough to conclude that there is a magnitude effect at work in these data.[5]

Chapman (1996) is a remarkable study because it is the first to try to estimate discount rates over the utility of time-dated final outcomes. It contains three experiments: the first two are replications and extensions of the basic design of Chapman and Elstein (1995), and the third attempts to correct for concave utility when inferring discount rates. We focus on the elicitation of discount rates over monetary outcomes; a major theme in the overall design is a comparison of discount rates defined over monetary and health outcomes.

Experiment 1 of Chapman (1996) poses 32 questions to 40 students fulfilling a course credit. These questions call for a fill-in-the-blank answer for the amount of money in the future equal to a fixed amount today. Horizons of 1, 3, 6 and 12 years were used, and principal amounts of \$500, \$1000, \$2000 and \$4000 were employed. Geometric means of log-transformed annual discount rates are used in the graphs, which provide the only basis for eyeballing those values. Figure 1 (Chapman 1996, p. 774) shows a clear magnitude effect, with discount rates for 1 year being roughly 115%, 115%, 75% and 75% for each principal amount, respectively; for a 3-year horizon, these rates drop to 60%, 50%, 40% and 40%, respectively; this decline and convergence pattern persists for the later horizons.

Experiment 2 of Chapman (1996) is similar, and extends the design to consider loss frames as well as gain frames. In this case the instrument contained 48 questions of the same fill-in-the-blank format, administered to 77 students again fulfilling a course credit. The horizons were only 1 year and 9 years, and the principal amounts were \$500, \$1500 and \$4500. The evidence suggests a clear magnitude effect for the 1-year horizon, and none for the 9-year horizon. For the shorter horizon, the discount rates, eyeballed from Figure 2A (Chapman 1996, p. 778), are 130% for a gain of \$500, 55% for gains of \$1500 or \$4500, 60% for a loss of \$500, and 25% for losses of \$1500 or \$4500. Thus there is evidence of a diminishing magnitude effect in these data.

Experiment 3 of Chapman (1996) is the one that seeks to correct for concave utility. The experimental design contains 38 questions that follow the style of Experiment 2, to elicit discount rates defined over monetary outcomes if one assumes linear utility. The horizons were 1, 3 and 9 years, the principal amounts were \$500, \$1000 and \$4500, and gain and loss frames were again considered. Eyeballing Figure 5A (Chapman 1996, p. 784), there is evidence of a magnitude effect for 1- and 3-year discount rates. This baseline, familiar from Experiments 1 and 2, is important in comparison to the utility-corrected discount rates.

Utility scales were elicited using the following procedure:

> In Part 2, participants answered 18 utility assessment questions in which they matched intervals to be of the same subjective magnitude (von Winterfeldt and Edwards (1986; p. 232)). For example, participants considered the interval from \$0 to \$500 and specified a value $x$ such that the interval from \$500 to \$$x$ was of the same subjective magnitude.Specifically, participants were asked to consider two intervals; for example,
>
> Interval A: You expected to win \$0 but instead will win \$500.
>
> Interval B: You expected to win \$500 but instead will win ___.
>
> They were asked to specify the amount that would have to appear in the blank for Interval B to make Interval B subjectively the same as Interval A. If a participant answered, say, \$1,200, the next question concerned an interval from \$500 to \$1,200 and a second from \$1,200 to $y$, in which the participant must specify $y$. Participants answered four such questions.
>
> If one arbitrarily defines \$0 to be 0 utiles and \$500 to be 1 utile, then participants' responses are associated with 2, 3, 4, and 5 utiles. To measure utility for monetary losses, I asked participants to match the interval from \$500 to \$0 to the interval from \$0 to $z$, in which $z$ was a negative amount. This answer corresponded to $-1$ utile. Four additional questions specified $-2$, $-3$, $-4$, and $-5$ utiles. (Chapman 1996, p. 782)

From these responses, power utility functions were estimated, one for each subject. Ignoring the reliability of the determination of these utility functions for the moment, the implied utility discount rates appear to exhibit no quantitatively significant magnitude effect. From Figure 5E (Chapman 1996, p. 784), the rates are between 20% and 5% for the 1-year horizon and the various principals, and between 15% and 5% for the 3-year horizon. Although these magnitude effects are statistically significant in terms of an ANOVA test (Chapman 1996, p. 785, Table 8), Chapman (1996, p. 783) notes that:

'Because of the very curved utility function for money, utility discount rates for money were so low that differences were hard to detect.' No data on the estimated exponents of these utility functions are presented, although the graph of the average value of the estimates (Chapman 1996, p. 782, Figure 3) does suggest significant concavity.

Kirby (1997) is another remarkable study. It used real incentives, used payments by subjects out of their own cash, used an incentive-compatible second-price sealed-offer auction to elicit present values, considered the effect of varying the deferred amount (\$10 or \$20), and considered all odd-numbered horizons between 1 and 29 days. Each subject entered 30 bids, and was told that one of these bids would be selected at random for payment if the bid was the winning bid. Each auction apparently consisted of the entire sample in an experiment, which does not affect the incentive compatibility of the procedure. Subjects in Experiment 1 were 'pseudo-volunteers' receiving extra credit in a psychology class for attending, but apart from the show-up rewards, all payments were salient. Subjects in Experiments 2 and 3 were 'people from the Williams College community, including summer students, college staff, and persons unaffiliated with the college', and recruitment was by sign-up fliers and newspaper advertisements. The daily rates averaged 5.6% for the \$10 amount and only 3.6% for the \$20 amount in one experiment with 24 students completing a class credit; 2.4% and 1.4%, respectively, for 28 non-students recruited from the field; and 3.5% and 2.4%, respectively, for 20 non-students. These were statistically significantly different from each other (Kirby 1997, pp. 60, 62, 63), and the implied annual rates would be extremely large.

Kirby *et al.* (1999) implement the procedures of Kirby and Maraković (1996) with 56 heroin addicts and 60 control subjects. The control subjects were recruited by newspaper advertisement, and then matched approximately to the demographics of the sample of heroin addicts. Each subject was given 27 binary choices, and told that there was a one-in-six chance that one of the 27 choices would be paid out. Thus these are incentivized tasks.[6] The later amounts of money were grouped into three reward sizes—small (\$25 to \$35), medium (\$50 to \$60) and large (\$75 to \$85)—and horizons varied between 7 and 186 days. The results are reported (p. 81, Table 3) in terms of the parameter $K$ in the hyperbolic specification $D(t) = 1/(1 + Kt)$ of the discount factor for horizon $t$, and the number of subjects that were indifferent at that value (intervals were averaged). For this specification, the implied discount rate is $d(t) = (1 + Kt)^{1/t} - 1$. Thus for each discount rate, one can infer the fraction of subjects who have a discount rate less than that discount rate, and draw inferences about the discount rates of the sample. In effect, this is an 'interval regression' with a constant, and with the sample weighted by the sample size. The raw data from Table 3 of Kirby *et al.* (1999) are collated in our Appendix A (available on request). A grouped logistic analysis of those data shows a statistically significant magnitude effect.

To test the hypothesis of a fixed premium for delayed outcomes, Benhabib *et al.* (2010) present subjects with two types of fill-in-the-blank tasks. In one type, the subject was asked 30 questions of the form: 'What amount of money \$$x$, if paid to you today, would make you indifferent to \$$y$ paid to you in $t$ days?' In this case the amount \$$y$ and the horizon $t$ would be filled in: $y \in \{10, 20, 30, 50, 100\}$ and $t \in$ {3 days, 1 week, 2 weeks, 1 month, 3 months, 6 months}. The response \$$x$ was incentivized with a Becker *et al.* (1964) (BDM) auction for one of the 30 choices selected at random. A price would be drawn from a uniform distribution between \$0 and \$$y$, and if the random price was greater than the stated amount \$$x$, then the subject would receive that random price immediately; otherwise, the subject would receive \$$y$ in $t$ days. So the upper bound of the BDM auction was the larger amount to be provided in the future. The other type of fill-in-the-blank question involved 30 questions of the form: 'What amount of money \$$y$ would make you indifferent between \$$x$ today and \$$y$ in $t$ days from now? [Upper bound = \$$z$]', where the text in brackets was given to subjects as notation instead of these words. In this case the values of $t$ were the same as in the first fill-in-the-blank task, and the values of $x \in \{1, 2, 3, 5, 6, 7\}$ for $z = 10$, $x \in \{4, 7, 8, 10, 12, 14\}$ for $z = 20$, $x \in \{8, 14, 17, 19, 22, 24\}$ for $z = 30$, $x \in \{15, 20, 28, 32, 36, 39\}$ for $z = 50$, and $x \in \{40, 60, 65, 70, 75, 80\}$ for $z = 100$. The same subjects were given both sets of questions on different days. The data were evaluated using a flexible specification, and the model estimated for each individual using non-linear least squares. The individual estimates are very erratic, with a wide range of behaviours being inferred. The general theme is of extremely high discount rates, reported to average 472%, and considerable noise. Although Benhabib *et al.* (2010, p. 209) report evidence of significant magnitude effects, they show data from only three subjects; presumably, those data are representative.

One major issue in this design is the subject comprehension of the BDM procedure, which is often asserted by experimenters to be understood by subjects but often is not.[7] In this design the upper bound for the BDM procedure was varied with magnitude in a complicated manner, as one can discern from inspection of the parameters listed above. This would not be an issue apart from the fact that it appears to have been a behavioural focal point for elicited valuations. The striking result is that the modal valuation was to bid the upper bound, and the majority were close to that upper bound. This is a serious concern about the subject comprehension of the elicitation procedure.

Finally, Halevy (2012) presents subjects with four sets of binary choice tasks using a 1-week time horizon between the sooner and later amounts. Treatments included prizes of \$10 and \$100, and the sooner amount to be paid out immediately or after 4 weeks. Subjects were asked to fill out four multiple price lists where the later amount varied between \$9.90 and \$11 for the low stakes treatment and between \$99 and \$110 for the high stakes treatment. One row in each payoff table was selected at random for payment, and the chosen amount was paid out at the preferred date by the subject. The average switching points are \$10.59 and \$103.7, which points to a magnitude effect, but the standard errors of the estimated coefficients are not presented, and a formal statistical test is not reported.

Magnitude effects are present in all of these studies, and are in most cases many times larger than what we find here. Halevy (2012) is the study that most closely matches our design since it uses an incentive-compatible elicitation mechanism, and real stakes with principals spanning \$10 and \$100. He finds that discount rates are three times higher for the smaller compared to the larger principals. While our design uses Danish kroner—converted to US dollars at the time of the experiment our principals are \$300 and \$600—none of the studies using real rewards employ principals that are close to ours.

## II. Experiments

To set the stage minimally for the discussion about experimental design, we define the discount factor for a given horizon $\tau$ to be the scalar $D(\tau)$ that equates the utility of the income received at time $t$ with the income received at time $t + \tau$:

$$U(y_t) = D(\tau)\, U(y_{t+\tau})$$

for some utility function $U(.)$. This general definition permits the special case, much studied in the experimental literature, in which $U(.)$ is linear.

The discount factor for the exponential specification is defined as

$$D(\tau) = 1/(1+\delta)^{\tau} \tag{1}$$

for $t \geq 0$, where the discount rate $d(\tau)$ is simply

$$d(\tau) = \delta. \tag{2}$$

Although these characterizations are abstract, we view the discount rate on an annualized basis throughout. The key feature of this model, of course, is that the discount rate is a constant over time. The percentage rate at which utility today and utility tomorrow are discounted is exactly the same as the rate at which utility in 7 days and utility in 8 days are discounted. For our immediate purposes, the exact form of the discounting function is of no consequence: one can view the exponential specification descriptively as simply a convenient summary statistic for the effect of magnitude on elicited discount rates.[8]

### *Design*

An important aspect of our methodological contribution to the literature on magnitude effects is the provision of sizeable monetary incentives that are paid out, rather than stated as hypothetically paid. This sharpens the incentives for respondents to truthfully report their preferences. Subjects are presented with two tasks.[9] The first task identifies individual discount rates, and the second task identifies atemporal risk attitudes and hence the concavity of the utility function. Observed choices from both tasks are then used to jointly estimate structural models of the discounting function defined over utility of income.

Using real monetary incentives comes with budgetary consequences for the researcher. Two different procedures could be used to keep such budgets reasonable: one could either use relatively small money amounts for both the smaller and the larger principal and pay every subject, or one could use larger amounts but pay for the task only with some probability. There are drawbacks with both of these approaches. Using small amounts can lead to confounding behavioural effects if subjects are rounding the money amounts. Using stochastic payments requires sophisticated portfolio analysis of the data or a maintained hypothesis that the independence axiom of EUT holds. We have opted for avoiding the former approach due to the apparent strong presence of rounding that is prevalent in the Benhabib *et al.* (2010) data. We use a 10% chance for a subject to be paid, allowing us to use principals with relatively high numerical values.[10]

We illustrate the potential seriousness of rounding behaviour in Table 2. The first four columns show the behaviour that would be expected of a subject in a fill-in-the-blank elicitation task if that subject had a 10% annualized discount rate. In this task the amount of money today in column (1) is given to the subject, and the future amount is the response that would be expected of someone who had a 10% discount rate, assuming that no rounding occurs. The horizon is shown in days, then in column (4) we see the implied annualized effective rate (AER) of discount, which by construction here had better be 10%. Each block of rows in Table 2 shows a series of binary choices with the same horizons, with the principal roughly doubling from block to block. Clearly the monetary premia in column (5) are all *very* small.

TABLE 2
SMALL STAKES AND THE EFFECTS OF ROUNDING

| Amount today (1) | Future amount (2) | Horizon in days (3) | AER (4) | Monetary premium (5) | Amount today (6) | Future amount (7) | Horizon in days (8) | AER (9) |
|---|---|---|---|---|---|---|---|---|
| $7 | $7.01 | 3 | 10% | $0.01 | $7 | $8 | 3 | 1661% |
| $6 | $6.01 | 7 | 10% | $0.01 | $6 | $7 | 7 | 813% |
| $5 | $5.02 | 14 | 10% | $0.02 | $5 | $6 | 14 | 478% |
| $3 | $3.02 | 30 | 10% | $0.02 | $3 | $4 | 30 | 352% |
| $2 | $2.05 | 91 | 10% | $0.05 | $2 | $3 | 91 | 163% |
| $1 | $1.05 | 181 | 10% | $0.05 | $1 | $2 | 181 | 140% |
| $14 | $14.01 | 3 | 10% | $0.01 | $14 | $15 | 3 | 849% |
| $12 | $12.02 | 7 | 10% | $0.02 | $12 | $13 | 7 | 420% |
| $10 | $10.04 | 14 | 10% | $0.04 | $10 | $11 | 14 | 249% |
| $8 | $8.07 | 30 | 10% | $0.07 | $8 | $9 | 30 | 144% |
| $7 | $7.18 | 91 | 10% | $0.18 | $7 | $8 | 91 | 54% |
| $4 | $4.20 | 181 | 10% | $0.20 | $4 | $5 | 181 | 45% |
| $24 | $24.02 | 3 | 10% | $0.02 | $24 | $25 | 3 | 500% |
| $22 | $22.04 | 7 | 10% | $0.04 | $22 | $23 | 7 | 233% |
| $19 | $19.07 | 14 | 10% | $0.07 | $19 | $20 | 14 | 134% |
| $17 | $17.14 | 30 | 10% | $0.14 | $17 | $18 | 30 | 70% |
| $14 | $14.36 | 91 | 10% | $0.36 | $14 | $15 | 91 | 28% |
| $8 | $8.41 | 181 | 10% | $0.41 | $8 | $9 | 181 | 24% |
| $39 | $39.03 | 3 | 10% | $0.03 | $39 | $40 | 3 | 309% |
| $36 | $36.07 | 7 | 10% | $0.07 | $36 | $37 | 7 | 143% |
| $32 | $32.12 | 14 | 10% | $0.12 | $32 | $33 | 14 | 80% |
| $28 | $28.23 | 30 | 10% | $0.23 | $28 | $29 | 30 | 43% |
| $20 | $20.50 | 91 | 10% | $0.50 | $20 | $21 | 91 | 20% |
| $15 | $15.76 | 181 | 10% | $0.76 | $15 | $16 | 181 | 13% |
| $80 | $80.07 | 3 | 10% | $0.07 | $80 | $81 | 3 | 151% |
| $75 | $75.14 | 7 | 10% | $0.14 | $75 | $76 | 7 | 69% |
| $70 | $70.27 | 14 | 10% | $0.27 | $70 | $71 | 14 | 37% |
| $65 | $65.54 | 30 | 10% | $0.54 | $65 | $66 | 30 | 19% |
| $60 | $61.51 | 91 | 10% | $1.51 | $60 | $62 | 91 | 13% |
| $40 | $42.03 | 181 | 10% | $2.03 | $40 | $43 | 181 | 15% |

The last four columns of Table 2 show what happens if the subject rounds up the filled-in response to the next whole dollar. Thus column (6) is the same as column (1), and column (7) is column (2) rounded up to the nearest dollar. The reason for rounding *up*, if one is going to round at all, is apparent from column (5): rounding to the *nearest* dollar would mean in most cases that the future amount was literally the same as the principal, and that is intuitively false to even the most cognitively challenged subject. But the impacts of this modest amount of rounding on the implied discount rates in column (9) are staggering. Two frequently reported anomalies immediately emerge: the appearance of declining discount rates with horizon, and declining discount rates with the magnitude of the principal.

Some evidence of rounding behaviour is found in Benhabib *et al.* (2010). In their experiment, 27 subjects were given the task to fill in the blank for future payments, based on exactly the horizons and principals shown in Table 2. They also did a comparable fill-in-the-blank task with the same 27 subjects in which a future amount was given and the subject had to choose the current amount to match. Out of 1620 observed responses, a staggering 95.6% reported amounts only in whole dollars.[11]

This chilling arithmetic follows from the very argument that motivates the theory exposition in Benhabib *et al.* (2010). What if subjects discount the future with some fixed monetary premium, irrespective of horizon, rather than some premium that is a proportion of the principal, as assumed in quasi-hyperbolic specifications? The upshot is implied by Table 2, but can be seen immediately by assuming a principal of $10 and the horizons of 3, 7, 14, 30, 91 and 181 days, and a fixed monetary premium of $1. The implied discount rates are 1178%, 500%, 249%, 116%, 38% and 19%, respectively, showing striking evidence of 'hyperbolicky' discounting. Then change the principal to $100, and the same monetary premium implies discount rates of only 121%, 52%, 26%, 12%, 4% and 2%, respectively, and a 'magnitude effect'.

### *Individual discount rates*

Individual discount rates are examined by asking subjects to make a series of choices over two certain outcomes that differ in terms of when they will be received. For example, one option can be 3000 kroner today, and another option can be 3300 kroner in 60 days.[12] If the subject picks the earlier option, then we can infer that their monetary discount rate is below 10% for 60 days, and if the subject picks the later option, then we can infer that their monetary discount rate is above 10% for that horizon. By varying the amount of the later option we can identify the utility discount rate of the individual, conditional on knowing the utility of those amounts to this individual. One can also vary the time horizon to identify the discount rate function. This method has been widely employed in the USA (e.g. Coller and Williams 1999), Denmark (e.g. Harrison *et al.* 2002) and Canada (e.g. Eckel *et al.* 2005).

We ask subjects to evaluate choices over several time horizons. We consider time horizons between 2 weeks and 1 year. Each subject is presented with choices over four time horizons, and those horizons are drawn at random, without replacement, from a set of thirteen possible horizons (2 weeks, and 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 and 12 months). This design will allow us to obtain a smooth characterization of the discount rate function across the sample for horizons up to one year.

Of course, the key treatment here is to vary the principal. We employ two levels of the principal—1500 kroner and 3000 kroner—on a between-subjects basis.[13] We also vary the order in which the time horizon is presented to the subject: in either ascending order or descending order. Similarly, we vary the provision of implied interest rates for each choice on a between-subjects basis, and independently of the two other treatments.

These three treatments—the level of the principal, the order of presentation of the horizon, and information on implied interest rates—result in a $2 \times 2 \times 2$ design. Roughly one-eighth of the sample was assigned at random to any one particular combination.

### *Risk attitudes*

Risk attitudes are evaluated by asking subjects to make a series of choices over outcomes that involve some uncertainty. To be clear, risk attitudes are elicited here simply as a convenient vehicle to estimate the non-linear utility function of the individual. The theoretical requirement, from the definition of a discount factor in (0), is for us to know the utility function over income if we are to correctly infer the discount rate used by the individual. The discount rate choices described above are not defined over lotteries.

Our design poses a series of binary lottery choices. For example, lottery A might give the individual a fifty-fifty chance of receiving 1600 kroner or 2000 kroner to be paid today, and lottery B might have a fifty-fifty chance of receiving 3850 kroner or 100 kroner today. The subject picks A or B. One series of 10 choices would offer these prize sets with probabilities on the high prize in each lottery starting at 0.1, then increasing by 0.1 until the last choice is between two certain amounts of money. We present these pairwise choices, one pair at a time, to the subject as a 'pie chart' showing prizes and probabilities. We give subjects 40 choices, in four sets of 10, where each set has the same prizes. The prize sets employed are as follows: [A1: 2000 and 1600; B1: 3850 and 100], [A2: 1125 and 750; B2: 2000 and 250], [A3: 1000 and 875; B3: 2000 and 75] and [A4: 2500 and 1000; B4: 4500 and 50]. The order of these four sets is random for each subject, but within each set the choices are presented in an ordered manner, with increments of the high prize probability of 0.1.

The typical findings from lottery choice experiments of this kind are that individuals are generally averse to risk, and that there is considerable heterogeneity in risk attitudes across subjects: see Harrison and Rutström (2008) for an extensive review. Much of that heterogeneity is correlated with observable characteristics, such as age and education level (Harrison *et al.* 2007), although here we represent it with a random distribution that describes the full extent of preference heterogeneity.

This design does not assume that behaviour is better characterized by expected utility theory (EUT) or some other model.

### *Sample and procedures*

Between 28 September and 22 October 2009, we conducted experiments with 198 Danes, of which 89 received the lower principal of 1500 kroner and 109 received the higher principal of 3000 kroner. The sample was drawn to be representative of the adult population as of 1 January 2009, using sampling procedures that are virtually identical to those documented at length in Harrison *et al.* (2005). We received a random sample of the population aged between 18 and 75, inclusive, from the Danish Civil Registration Office, stratified that by geographic area, and sent out 1969 invitations.[14] The experiments reported here are a subsample of the complete sample (which was 413 subjects), including only those who did not have a front-end delay to the sooner option.

Our experiments were all conducted in hotel meeting rooms around Denmark, so that travel logistics for the sample would be minimized. Various times of day were also offered to subjects, to facilitate a broad mix of attendance. The largest session had 15 subjects, but most had fewer. The procedures were standard: Appendix B (available on request) documents an English translation of the instructions, and shows typical screen displays. Subjects were given written instructions, which were also read out, and then made choices in a trainer task, which was 'played out' so that the full set of consequences of each choice was clear. In fact, subjects were paid *Big Ben* caramels instead of money for all training tasks, and the payments were happily consumed when delivered. All interactions were by computer. The order of the block of discount rate tasks and the block of risk attitude tasks was randomized for each session. After all choices had been made, the subject was asked a series of standard socio-demographic questions.

There were 40 discounting choices and 40 risk attitude choices, and each subject had a 10% chance of being paid for one of each set. Average payments on the first block were 228 kroner (although some were for deferred receipt), and on the second block the average was 218 kroner, for a combined average of 446 kroner. The exchange rate at the time was close to 5 kroner per US dollar, so earnings averaged $89 per two-hour session for these tasks. Subjects were also paid a show-up fee of 300 kroner or 500 kroner, and earnings from additional tasks completed after the tasks of interest here were completed.[15]

For payments to be made in the future, the following language explained the procedures:

> You will receive the money on the date stated in your preferred option. If you receive some money today, then it is paid out at the end of the experiment. If you receive some money to be paid in the future, then it is transferred to your personal bank account on the specified date. In that case you will receive a written confirmation from Copenhagen Business School which guarantees that the money is reserved on an account at Danske Bank. You can send this document to Danske Bank in a prepaid envelope, and the bank will transfer the money to your account on the specified date.

Payments by way of bank transfer are common in Denmark, Copenhagen Business School is well known in Denmark, and Danske Bank is the largest financial enterprise in Denmark as measured by total assets.[16]

## III. ECONOMETRICS

Although the core treatment is a simple one, i.e. varying the principal, inferences about implied discount rates require attention to the definition of a discount rate in terms of utility streams, as shown earlier in (0). The approach that we adopt is direct estimation by maximum simulated likelihood of some structural model of a latent choice process in which the core parameters defining risk attitudes and discounting behaviour can be estimated. We review the basic inferential logic for estimating risk attitudes, and discuss the extension to discounting behaviour. We employ maximum simulated likelihood because we estimate the main structural parameters as random coefficients, to reflect unobserved individual heterogeneity.

### *Estimating the utility function*

Assume that utility of income is defined by

$$U(y) = M^{1-r}/(1-r), \tag{3}$$

where $M$ is the lottery prize and $r \neq 1$ is a parameter to be estimated. For $r = 1$, assume $U(M) = \ln(M)$ if needed. Thus $r$ is the coefficient of constant relative risk aversion (CRRA): $r = 0$ corresponds to risk neutrality, $r < 0$ to risk loving, and $r > 0$ to risk aversion. Let there be two possible outcomes in a lottery. Under EUT, the probabilities $p(M_j)$ for each outcome $M_j$ are those that are induced by the experimenter, so expected utility is simply the probability-weighted utility of each outcome in each lottery $i$ plus some level of background consumption $\omega$:

$$EU_i = [p(M_1) \times U(\omega + M_1)] + [p(M_2) \times U(\omega + M_2)]. \tag{4}$$

The EU for each lottery pair is calculated for a candidate estimate of $r$, and the index

$$\nabla EU = EU_A - EU_B \tag{5}$$

is calculated, where $EU_A$ is the lottery in Option $A$ and $EU_B$ is the lottery in Option $B$ as presented to subjects. This latent index, based on latent preferences, is then linked to observed choices using the cumulative logistic distribution function $\Lambda(\nabla EU)$. This 'logit' function takes any argument between $\pm\infty$ and transforms it into a number between 0 and 1. Thus we have the logit link function

$$\text{prob}(\text{choose Option} A) = \Lambda(\nabla EU). \tag{6}$$

The index defined by (5) is linked to the observed choices by specifying that the lottery in Option $A$ is chosen when $\Lambda(\nabla EU) > 1/2$, which is implied by (6).

Thus the likelihood of the observed responses, conditional on the EUT and CRRA specifications being true, depends on the estimates of $r$ given the above statistical specification and the observed choices. Assuming non-random coefficients for the moment, the conditional log-likelihood is then

$$\begin{aligned} \ln L^{RA} &= \ln L(r; y, \omega) \\ &= \sum_i [(\ln \Lambda(\nabla EU) \times I(y_i = 1)) + (\ln(1 - \Lambda(\nabla EU)) \times I(y_i = -1))], \end{aligned} \tag{7}$$

where $I(.)$ is the indicator function, and $y_i = 1$ or $-1$ denotes the choice of the Option $A$ or $B$ lottery in risk aversion task $i$. Harrison and Rutström (2008, Appendix F) review procedures and syntax from the popular statistical package *Stata* that can be used to estimate structural models of this kind, as well as more complex non-EUT models.

We employ a *behavioural* error specification originally due to Fechner and popularized by Hey and Orme (1994). This error specification posits the latent index

$$\nabla EU = (EU_A - EU_B)/\mu \tag{5$'$}$$

instead of (5), where $\mu$ is a structural 'noise parameter' used to allow some errors from the perspective of the deterministic EUT model. This is just one of several different types of error story that could be used, and Wilcox (2008) provides a masterful review of the implications of the alternatives.[17] As $\mu \to 0$, this specification collapses to the deterministic choice EUT model, where the choice is strictly determined by the EU of the two lotteries; but as $\mu$ gets larger and larger, the choice essentially becomes random. When $\mu = 1$, this specification collapses to (5). Thus $\mu$ can be viewed as a parameter that flattens out the link functions as it gets larger.

An important contribution to the characterization of behavioural errors is the 'contextual error' specification proposed by Wilcox (2011). It is designed to allow robust inferences about the primitive 'more stochastically risk averse than'. It posits the latent index

$$\nabla EU = [(EU_A - EU_B)/v]/\mu \tag{5$''$}$$

instead of (5 ), where $v$ is a new, normalizing term for each lottery pair. The normalizing term $v$ is defined as the maximum utility over all prizes in this lottery pair minus the minimum utility over all prizes in this lottery pair. The value of $v$ varies, in principle, from lottery choice to lottery choice: hence it is said to be 'contextual'. For the Fechner specification, dividing by $v$ ensures that the *normalized* EU difference $(EU_A - EU_B)/v$ remains in the unit interval.

*Estimating the discounting function*

Assume that EUT holds for choices over risky alternatives and that discounting is exponential. A subject is indifferent between two income options $M_t$ and $M_{t+\tau}$ if and only if

$$U(\omega + M_t) + (1/(1+\delta)^{\tau})\, U(\omega) = U(\omega) + (1/(1+\delta)^{\tau})\, U(\omega + M_{t+\tau}), \tag{8}$$

where $U(\omega + M_t)$ is the utility of monetary outcome $M_t$ for delivery at time $t$ plus some measure of background consumption $\omega$, $\delta$ is the discount rate, $\tau$ is the horizon for delivery of the later monetary outcome at time $t + \tau$, and the utility function $U$ is separable and stationary over time. The left-hand side of equation (8) is the sum of the discounted utilities of receiving the monetary outcome $M_t$ at time $t$ (in addition to background consumption) and receiving nothing extra at time $t + \tau$; the right-hand side is the sum of the discounted utilities of receiving nothing over background consumption at time $t$ and the outcome $M_{t+\tau}$ (plus background consumption) at time $t + \tau$. Thus (8) is an indifference condition and $\delta$ is the discount rate that equalizes the present value of the *utility* of the two monetary outcomes $M_t$ and $M_{t+\tau}$, after integration with an appropriate level of background consumption $\omega$.[18]

We can write out the likelihood function for the choices that our subjects made, and jointly estimate the risk parameter $r$ in (3) and the discount rate parameter $\delta$ in (8). We use the same stochastic error specification as in (5 ).[19] Instead of (5 ) we have

$$\nabla PV = (PV_A - PV_B)/\mu, \tag{9}$$

where the discounted utility of Option $A$ is given by

$$PV_A = (\omega + M_A)^{1-r}/(1-r) + (1/(1+\delta)^{\tau})\,\omega^{1-r}/(1-r), \tag{10}$$

and the discounted utility of Option $B$ is

$$PV_B = \omega^{1-r}/(1-r) + (1/(1+\delta)^{\tau})(\omega + M_B)^{1-r}/(1-r), \tag{11}$$

and $M_A$ and $M_B$ are the monetary amounts in the choice tasks presented to subjects. We assume here that the utility function is stable over time and is perceived *ex ante* to be stable over time.[20]

Thus the likelihood of the discount rate responses, conditional on the EUT, CRRA and exponential discounting specifications being true, depends on the estimates of $r$, $\delta$ and $\mu$, given the assumed value of $\omega$ and the observed choices. The conditional log-likelihood is

$$\begin{aligned} \ln L^{DR} &= \ln L(r, \delta, \mu; y, \omega) \\ &= \sum_i [(\ln \Lambda(\nabla PV) \times I(y_i = 1)) + (\ln(1 - \Lambda(\nabla PV)) \times I(y_i = -1))], \end{aligned} \tag{12}$$

where $y_i = 1$ or $-1$ again denotes the choice of Option $A$ or $B$ in discount rate task $i$.

The joint likelihood of the risk aversion and discount rate responses can then be written as

$$\ln L(r, \delta, \mu; y, \omega) = \ln L^{RA} + \ln L^{DR}, \tag{13}$$

where $L^{RA}$ is defined by (7) and $L^{DR}$ is defined by (12). This expression can then be maximized using standard numerical methods. The parameter $\omega$ is set exogenously: using data from the household expenditure survey at Statistics Denmark, Andersen *et al.* (2008a, p. 600, Appendix D) calculate per capita consumption of private non-durable goods on an average daily basis as being equal to 118 kroner in 2003.[21] We adjust that amount for inflation to the time of our experiments, and assume that $\omega = 130$ kroner.

Nothing in this inferential procedure relies on the use of EUT, or the CRRA functional form. Nor does anything rely on the use of the exponential discounting function. In fact, following the literature on the magnitude effect, we can employ the exponential discounting function *descriptively*, since all we are interested in here is whether the discount rate differs as we vary the principal. Our methods generalize immediately to alternative models of decision-making under risk, and especially to alternative discounting functions, as demonstrated by Andersen *et al.* (2011).

### *Random coefficients*

We account for unobserved individual heterogeneity through the possibility that the coefficients $r$ and $\delta$ are *random coefficients* following some parametric distribution. In other words, one can allow the coefficients $r$ and $\delta$ to be distributed in a random manner across the population: each subject behaves as if they have a specific $r$ and $\delta$, but there is variation across subjects, and that variation is assumed to be characterized by some parametric distribution.

For example, if $\delta$ is assumed to vary according to a normal distribution, then one would estimate two 'hyper-parameters' to characterize that distribution: a population mean of $\delta$ and a population standard *deviation* of $\delta$.[22] Each of these hyper-parameters would have a point estimate and a standard *error*, where the latter derives from familiar sampling variability. As the sample size increases, and assuming consistent estimators, the *sampling error* on the population mean and the population standard deviation would converge to 0, but there is no presumption that the point estimate for the *population standard deviation* converges to 0, since it is a characteristic of the population and not sample variability.

We use non-linear 'mixed logit' methods developed by Andersen *et al.* (2012) to estimate such specifications. In fact, we also allow the distribution for $r$ and $\delta$ to be a logit-normal distribution, which is a logistic transform of a normally distributed variable. Due originally to Johnson (1949), and familiar in biostatistics, this transformation allows the resulting distribution to closely approximate a flexible Beta distribution: it allows skewness and bimodality. The domain is restricted to the unit interval, but it is a simple matter to expand that to any finite interval.

## IV. Results

A direct test of the magnitude effect can be made by varying the principal offered to respondents. The analysis of the data is therefore a simple matter: did the higher principal generate discount rates that were lower than the discount rates generated by the lower principal? We elicited discount rates over a horizon between 2 weeks and 1 year. Although this analysis could be performed on quasi-hyperbolic or hyperbolic discounting functions, the exponential discounting model offers a parsimonious one-parameter alternative, in line with our previous findings in Andersen *et al.* (2011). Hence we consider a single discount rate across time horizons when comparing the effect of the principal.

Figure 1 presents the raw data from the discounting tasks and shows the fraction of sooner choices across different levels of annual interest rates for the low and high magnitude treatments. The fraction of sooner choices generally falls when the interest rate increases, and we observe some differences in responses between the two treatments. The raw data point to a larger fraction of sooner choices for the high magnitude treatment when the interest rate is between 10% and 45%, and a larger fraction of sooner choices for the low magnitude treatment at either end of the interest rate interval.

Figure 2 shows the population distributions of estimated discount rates obtained if we (incorrectly) assume linear utility functions. With the low principal, the median discount rate is 10.5%, the mean discount rate is 24.7%, and the standard deviation is 29.3%; with the high principal, the median discount rate is 12.1%, the mean discount rate is 24.3%, and the standard deviation is 27.3%. The mean discount rate for the low principal is indeed higher than the discount rate for the high principal under the assumption of linear utility, but the difference is only 0.4 percentage points, which is very small in comparison to previous studies. We use a two-group mean-comparison test and find that the difference of 0.3 percentage points is significant with a *p*-value of 0.004.

Table 3 shows the random coefficient estimates for the pooled data spanning our two treatments when we allow for non-linear utility functions, which is our preferred model. The table shows parameter values for the underlying normal distribution in panel A and the logistic transformation in panel B. The results in panel B show that Danes are generally risk averse, with a mean risk aversion of 0.49 and a standard deviation of 0.32. The population distribution for the discount rate is sharply, positively skewed, as can be seen in Figure 3.[23] The mean discount rate is 0.059, and the standard deviation is 0.174. Of

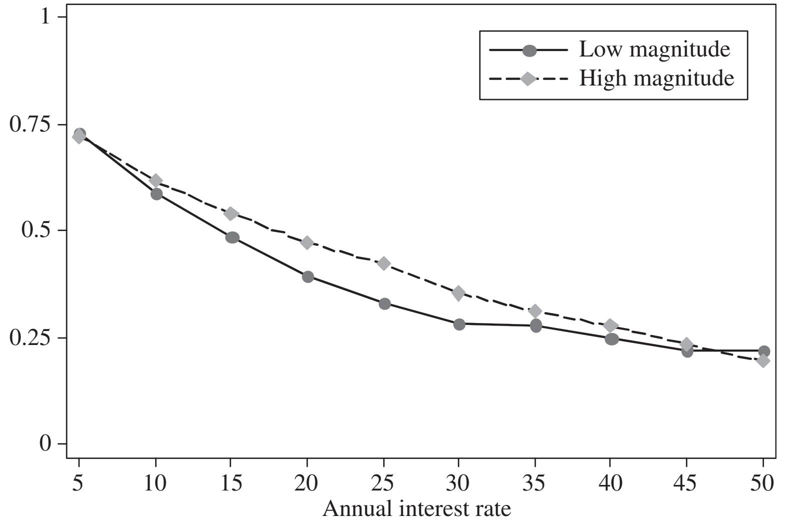


FIGURE 1. Fraction of sample choosing the sooner option.

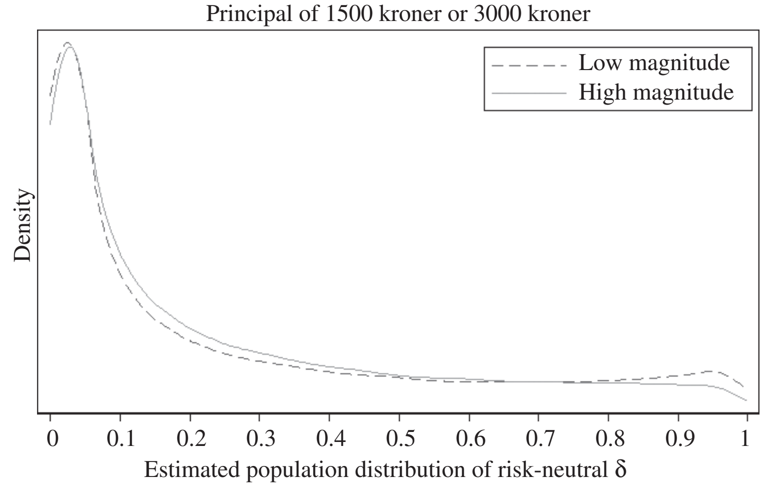


FIGURE 2. Estimated population distribution of risk-neutral discount rates by magnitude of principal (principal of 1500 or 3000 kroner).

TABLE 3
RANDOM COEFFICIENT ESTIMATES BASED ON POOLED DATA ($LL = -6716.3$, $N = 198$)

| | A. Underlying normal distribution | | | B. Logistic transformation | | |
|---|---|---|---|---|---|---|
| Parameter | Point estimate | Standard error | $p$-value | Point estimate | Standard error | $p$-value |
| $r_{mean}$ | −0.121 | 0.072 | 0.093 | 0.530 | 0.018 | < 0.001 |
| $r_{sd}$ | 0.709 | 0.022 | < 0.001 | 0.330 | 0.005 | < 0.001 |
| $\delta_{mean}$ | 1.355 | 0.177 | < 0.001 | 0.082 | 0.011 | < 0.001 |
| $\delta_{sd}$ | 2.365 | 0.170 | < 0.001 | 0.038 | 0.007 | < 0.001 |
| $\rho$ | | | | 0.027 | 0.023 | 0.049 |
| $\mu$ | | | | 0.025 | 0.021 | 0.001 |

course, since this is a skewed distribution, one should not infer statistical insignificance from the standard deviation equalling or exceeding the mean. For the same reason, the appropriate measure of central tendency of the population distribution is the median rather than the mean. Figure 3 displays the estimated population distribution for the logistic transformation $\Lambda(N(-0.12, 0.71))$ for risk attitudes in the left panel, and the transformation $0.6 \times \Lambda(N(1.36, 2.36))$ for discount rates in the right panel. The covariance between the two random coefficients is 0.059, with a 95% confidence interval between −0.04 and 0.16; so we reject the hypothesis that the two coefficients are independent.

Figure 4 then displays the estimated population distribution for the discount rate according to the level of the principal. We observe a statistically significant difference, in the direction of the magnitude effect, although the quantitative magnitude is much smaller than conventionally reported. The median discount rate for the low principal is 7.7%, and the average is 13.6%; the median discount rate for the high principal is 8.0%, and the average is 11.9%. The means of these distributions are significantly different from

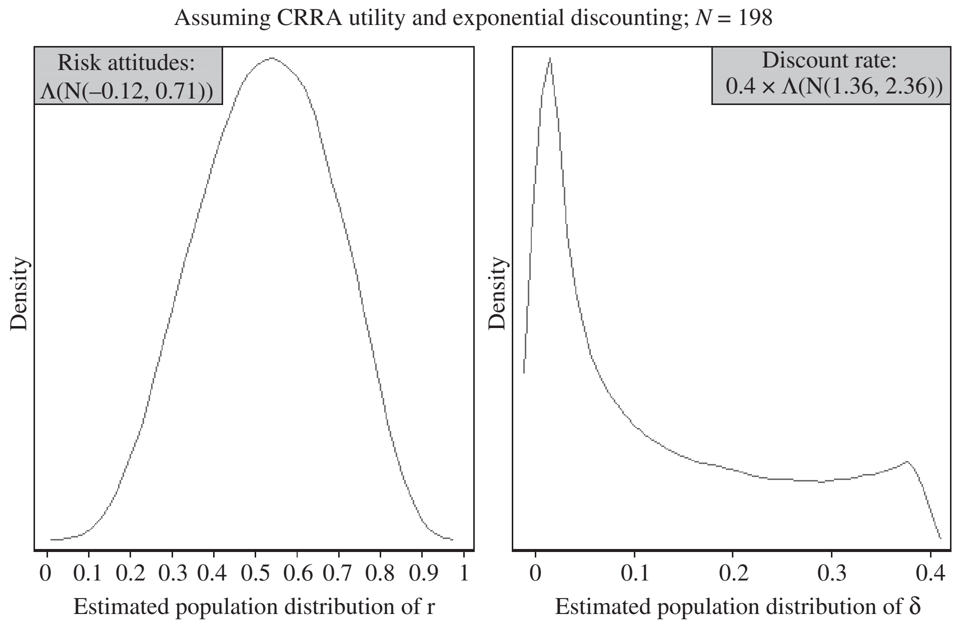


FIGURE 3. Random coefficient estimates of risk attitudes and discount rates (assuming CRRA utility and exponential discounting, $N = 198$).

each other by 1.7 percentage points ($p < 0.001$), consistent with the visual impression from Figure 4.[24] However, the median discount rates are almost identical, with a difference of only 0.3 percentage points, and this difference is significant with $p = 0.016$ using a non-parametric equality-of-medians test. Hence we see evidence of a modest magnitude effect based on the mean estimates, and a small effect based on the median estimates.

Finally, we have actually thrown away roughly one-half of the data that we collected by focusing solely on tasks in which the sooner option was to be delivered now. The missing half involved choices in which the sooner option was to be delivered 30 days from the date of choice, so that both options had some delay. This 'front-end delay' is intended to reduce the differences in subjective transactions costs of realizing either option. Collecting money at the end of the session is simply more credible than relying on someone to send a bank deposit in one month. The vast bulk of the literature on the magnitude effect assumes no front-end delay, hence our main analyses focused on that case. Figure 5 shows the effect of adding in the choices with a front-end delay. The magnitude effect does not disappear. The median discount rates are 10.2% and 7.3%, and the mean discount rates are 5.1% and 11.6%, for the low and high principals, respectively. So there is a small, but significant, difference in the estimated means by treatment ($p < 0.001$).[25] Choices with a front-end delay are appropriate characterizations of many financial decisions, but not all economic decisions. We view each as appropriate in different settings.

How do we reconcile our results with the received wisdom? The first response is to question if the evidence for the magnitude effect is all that impressive, given the contexts used in the previous literature. We see nothing in our experimental procedures that might bias behaviour in undesirable ways, or that deviates in any novel manner from the types of procedures used in the past. We avoid eliciting present values in an open-ended manner, because we are suspicious of the behavioural accuracy of those responses.[26] We use

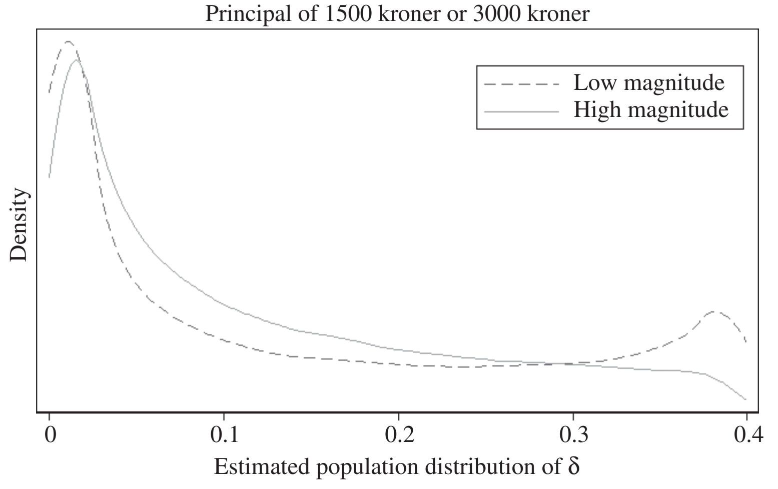


FIGURE 4. Estimated population distribution of discount rates by magnitude of principal (principal of 1500 or 3000 kroner).

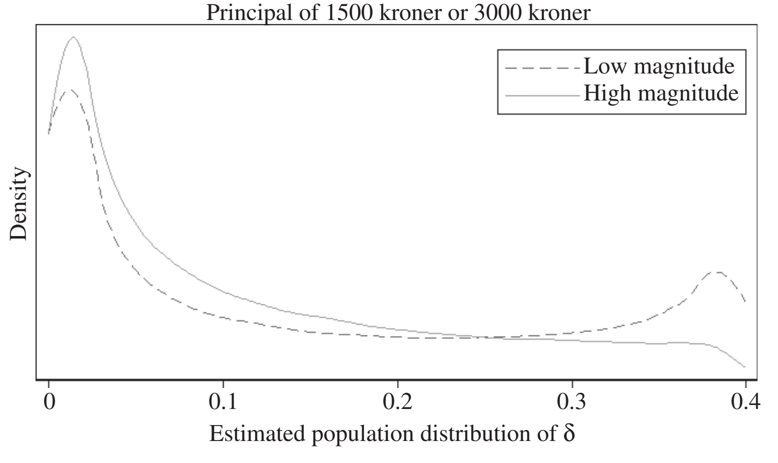


FIGURE 5. Estimated population distribution of discount rates by magnitude of principal, adding responses with front-end delay (principal of 1500 or 3000 kroner).

displays of the tasks that make them relatively transparent in terms of the choice alternatives, rather than relying entirely on the ability of subjects to read numbers and words. And, obviously, we pay the subjects in a salient manner.

Our basic econometric procedures are familiar from the binary choice literature, and have a long tradition in experimental economics (e.g. Camerer and Ho 1994; Hey and Orme 1994).

We use relatively flexible parametric methods, but if the magnitude effect is as robust as claimed, then the use of parametric functional forms should not be a factor causing it to disappear. And the notion of joint estimation of utility functions and discounting functions is driven by theory, and implies nothing fundamental from an econometric perspective.

The only factor that is left, in comparison with those (few) studies that use salient rewards and incentive-compatible elicitation methods, is the stake size. One reason that we pay the subject with only a 10% chance is to allow them to face a principal with relatively high numerical values. The rationale is that this avoids rounding effects that can be severely confounding for inferences about discounting behaviour, particularly when comparing behaviour across stakes as we do here.

## V. Conclusions

We use real incentives with large monetary rewards to avoid possible rounding effects. With some exceptions, noted in our literature review, all evidence of the magnitude effect that meets certain minimal standards of salience and design occurs in samples of college-age students. We do observe a statistically significant magnitude effect in the discounting behaviour of adult Danes making choices of deferred monetary payments, but it is not very large: the size of the effect depends on whether one looks at median estimates or mean estimates. The preferred median estimates, correcting for unobserved individual heterogeneity and non-linear utility functions, differ by only 0.7 percentage points, from 4.2% to 3.5%, and the mean estimates differ by 3.4 percentage points, from 11.3% to 7.9%. Given the sharp skewness of the population distribution, there is an argument for focusing on the median, but we report each measure of central tendency. By using relatively large principals of 1500 kroner and 3000 kroner, we avoid confounding effects from rounding behaviour, and can more clearly test the presence of magnitude effects.

We conclude that the alleged magnitude effect in discounting is not economically significant for the design of policy for Danes. Within the range of stakes used here, there is no apparent reason to apply different discount rates. Of course, it is an open question whether this conclusion holds when the stakes get even larger, into the tens or hundreds of thousands of kroner. If it does, then it would be appropriate to use the same discount rates for policies concerning small principals, such as small tax refunds or small consumer loans, as for those concerning large principals, such as house mortgages. Our findings demonstrate to practitioners that the magnitude effects found elsewhere in the literature need to be tested for robustness with respect to the level of real rewards before they can inform policy.

## ACKNOWLEDGMENTS

We thank the US National Science Foundation for research support under grants NSF/HSD 0527675 and NSF/SES 0616746, the Danish Social Science Research Council for research support under project 275-08-0289, and the Carlsberg Foundation under grant 2008_01_0410.

## NOTES

1. Baucells and Heukamp (2012) consider probability discount rates as opposed to monetary discount rates. They attribute the magnitude effect to sub-endurance, meaning that subjects are willing to wait longer for amounts that are paid out with higher probability.
2. Kirby (1997) also considers two other samples with 'people from the Williams College community, including summer students, college staff, and persons unaffiliated with the college', and he finds similar magnitude effects.
3. Another group was presented with comparable tasks framed as losses. We focus here solely on choices framed as gains.
4. No statistical test is typically needed when the median or means are so different, although one would like to know if the distributions had large variances or skewness.

5. Appendix A (available on request) collates the data reported in the tables of Holcomb and Nelson (1992) and demonstrates that there is a statistically significant magnitude effect.
6. One immediate concern with Kirby and Maraković (1996) is that their study is 'almost hypothetical'. Questionnaires were sent out to every undergraduate student at Williams College, which had a student population of roughly 2000 at the time. Subjects were told that the questionnaires could be returned before one of two days, and that on each day one would be drawn at random and one of the 21 choices played out for real.
7. Benhabib *et al.* (2010, p. 208) assert that: 'We had no doubt that the subjects understood the incentive properties of the mechanism.' Concerns about the behavioural reliability of the BDM procedure are voiced by Harrison (1992), Plott and Zeiler (2005), and Harrison and Rutström (2008, Section 1.4, Appendix D).
8. Andersen *et al.* (2011) look at the menagerie of quasi-hyperbolic and hyperbolic specifications and find no statistical evidence in favour of non-exponential discounting using the same data as here.
9. A complete list of parameter values for all choices is presented in Appendix B (available on request).
10. Andersen *et al.* (2011) vary the probability of payment for the discounting task from 10% to 100% and find no significant difference in the responses to these treatments.
11. Moreover, 3.2% were rounded to the nearest \$0.50, and 0.6% to the nearest \$0.25, leaving only 11 responses, or 0.7%, to be reported with more precision.
12. Much of the literature in economics employs a front-end delay such that the earlier option is delayed by some time period, such as one week or one month (e.g. Coller and Williams 1999; Harrison *et al.* 2002). Virtually none of the studies documenting a magnitude effect employs a front-end delay, so we also avoid it here in our baseline design.
13. Of course, one can look at greater differences in principals compared to the 2:1 ratio that we consider in our design. We opted for 1500 kroner as the lower principal to avoid rounding effects as explained above, and budgetary constraints prevented us from using principals larger than 3000 kroner. These stakes are large compared to those used in the literature, and we do not want to consider hypothetical questions as an alternative method to consider 'larger' stakes.
14. The recruiting sample was drawn by us from a random sample that includes information on gender, age, residential location, marital status, and whether the individual is an immigrant. At a very broad level, our sample was representative on average: the sample of 50,000 had an average age of 49.8 years, 50.1% of them were married, and 50.7% were female; our final sample of 413 subjects had an average age of 48.7 years, 56.5% of them were married, and 48.2% were female.
15. An extra show-up fee of 200 kroner was paid to 35 subjects who had received invitations stating 300 kroner but then received a final reminder that accidentally stated 500 kroner. In general, the additional tasks earned subjects an average of at least 370 kroner (the exact amount depended on later decisions by other subjects), so total earnings from choices made in the session averaged 722.9 kroner, or roughly \$145.
16. The payment methods do lead to small differences in transaction costs between payments made at the experiment and payments made by bank transfer. However, if all payments are made by bank transfer, then one effectively introduces a front-end delay on the sooner income option since transfers are done overnight in the Danish banking system.
17. Some specifications place the error at the final choice between one lottery or after the subject has decided which one has the higher expected utility; some place the error earlier, on the comparison of preferences leading to the choice; and some place the error even earlier, on the determination of the expected utility of each lottery.
18. Andersen *et al.* (2008a) show that the addition of background consumption is a sufficient condition to avoid negative discount rates when the intertemporal utility function is additively separable.
19. We do not need to apply the contextual utility correction $v$ for these choices since they are over deterministic monetary amounts.
20. Direct evidence for the former proposition is provided by Andersen *et al.* (2008b), who examine the temporal stability of risk attitudes in the Danish population. The second proposition is a more delicate matter: even if utility functions are stable over time, they may not be subjectively perceived to be, and that is what matters for us to assume that the same $r$ that appears in (3) appears in (10) and (11). When there is no front-end delay, as here, this assumption is immediate for (10), but not otherwise.
21. Andersen *et al.* (2008a, p. 602) show that estimates are robust to variations in $\omega$ between 50 and 200 kroner.
22. In fact, we allow for a non-zero correlation between these two random coefficients, so their covariance is a third hyper-parameter to be estimated.
23. There is one technical issue of importance here, however. As flexible as the logit-normal is, it allows bimodality only at the *endpoints* of the finite interval allowed. In this case we constrained the domain to be between 0 and 0.6, and hence the mode close to 0 *might* be an artefact of that assumption. Although we know *a priori* that $\delta \geq 0$, we do not know the upper bound. One can loop through alternative parametric assumptions of the upper bound and evaluate the maximum likelihood at each: these are known as profile likelihoods. In our case the qualitative results are invariant to assuming upper bounds lower than 0.6. A better solution, which is common in the statistical literature to allow 'internal modes', is to allow mixtures.
24. We find similar qualitative results for the hyperbolic discounting model $D(t) = 1/(1 + Kt)$. The mean estimate of the parameter $K$ for the low principal is 13.1 with a median of 5.6, and the mean estimate of $K$

for the high principal is 9.5 with a median of 4.7. The means of these distributions are significantly different, with $p < 0.001$.

25. The results are similar when we consider only choices with a front-end delay. The estimated mean discount rates are 12.6% and 12.9% for the low and high principals, respectively, and the median discount rates are 6.1% and 6.5%. The difference in the estimated means by treatment is significant, with $p < 0.001$. The fixed-cost model proposed by Benhabib *et al.* (2010) cannot explain this pattern of behaviour, since the model posits a risk premium to any delayed outcome and constant discounting between any two delayed outcomes with the same time horizon.
26. We strongly encourage systematic studies of the effects of using choice and open-ended fill-in-the-blank procedures, along the lines of Ahlbrecht and Weber (1997) and Read and Roelofsma (2003), but for discounting tasks in which subjects are making salient, non-hypothetical choices.

## REFERENCES

AHLBRECHT, M. and WEBER, M. (1997). An empirical study on intertemporal decision making under risk. *Management Science*, **43**(6), 813–26.

AINSLIE, G. and HAENDEL, V. (1983). The motives of the will. In E. Gottheil, K. Druley, T. Skodola and H. Waxman (eds), *Etiology Aspects of Alcohol and Drug Abuse*. Springfield, IL: Charles C. Thomas.

ANDERSEN, S., HARRISON, G. W., HOLE, A. R., LAU, M. I. and RUTSTRÖM, E. E. (2012). Non-linear mixed logit. *Theory and Decision*, **73**(1), 77–96.

———, ———, LAU, M. I. and RUTSTRÖM, E. E. (2008a). Eliciting risk and time preferences. *Econometrica*, **76** (3), 583–619.

———, ———, ——— and ——— (2008b). Lost in state space: are preferences stable? *International Economic Review*, **49**(3), 1091–112.

———, ———, ——— and ——— (2011). Discounting behavior: a reconsideration. Working Paper 2011-03, Center for the Economic Analysis of Risk, Robinson College of Business, Georgia State University.

BAUCELLS, M. and HEUKAMP, F. H. (2012). Probability and time trade-off. *Management Science*, **58**(4), 831–42.

BECKER, G. M., DEGROOT, M. H. and MARSCHAK, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, **9**, 226–32.

BENHABIB, J., BISIN, A. and SCHOTTER, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games & Economic Behavior*, **69**(2), 205–23.

BENZION, U., RAPOPORT, A. and YAGIL, J. (1989). Discount rates inferred from decisions: an experimental study. *Management Science*, **35**, 270–84.

CAMERER, C. and HO, T. (1994). Violations of the betweenness axiom and nonlinearity in probability. *Journal of Risk & Uncertainty*, **8**, 167–96.

CHAPMAN, G. B. (1996). Temporal discounting and utility for health and money. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **22**, 771–91.

——— and ELSTEIN, A. S. (1995). Valuing the future: discounting health and money. *Medical Decision Making*, **15**, 373–86.

——— and WINQUIST, J. R. (1998). The magnitude effect: temporal discount rates and restaurant tips. *Psychonomic Bulletin & Review*, **5**, 119–23.

COLLER, M. and WILLIAMS, M. B. (1999). Eliciting individual discount rates. *Experimental Economics*, **2**, 107–27.

DU, W., GREEN, L. and MYERSON, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *Psychological Record*, **52**, 479–92.

ECKEL, C., JOHNSON, C. and MONTMARQUETTE, C. (2005). Savings decisions of the working poor: short- and long-term horizons. In J. Carpenter, G. W. Harrison and J. A. List (eds), *Field Experiments in Economics, Vol. 10, Research in Experimental Economics*. Greenwich, CT: JAI Press.

ESTLE, S. J., GREEN, L., MYERSON, J. and HOLT, D. D. (2007). Discounting of monetary and directly consumable rewards. *Psychological Science*, **18**(1), 58–63.

GREEN, L., FRISTOE, N. and MYERSON, J. (1994). Temporal discounting and preference reversals in choice between delayed outcomes. *Psychonomic Bulletin & Review*, **1**, 383–9.

———, FRY, A. F. and MYERSON, J. (1994). Discounting of delayed rewards: a life-span comparison. *Psychological Science*, **5**, 33–6.

———, MYERSON, J. and MCFADDEN, E. (1997). Rate of temporal discounting decreases with amount of reward. *Memory & Cognition*, **25**(5), 715–23.

HALEVY, Y. (2012). Time consistency: stationarity and time invariance. Mimeo, University of British Columbia.

HARRISON, G. W. (1992). Theory and misbehavior of first-price auctions: reply. *American Economic Review*, **82**, 1426–43.

——— and LIST, J. A. (2004). Field experiments. *Journal of Economic Literature*, **42**(4), 1013–59.
——— and RUTSTRÖM, E. E. (2008). Risk aversion in the laboratory. In J. C. Cox and G. W. Harrison (eds), *Risk Aversion in Experiments, Vol. 12, Research in Experimental Economics*. Bingley: Emerald.
———, LAU, M. I. and RUTSTRÖM, E. E. (2007). Estimating risk attitudes in Denmark: a field experiment. *Scandinavian Journal of Economics*, **109**(2), 341–68.
———, ———, ——— and SULLIVAN, M. B. (2005). Eliciting risk and time preferences using field experiments: some methodological issues. In J. Carpenter, G. W. Harrison and J. A. List (eds), *Field Experiments in Economics, Vol. 10, Research in Experimental Economics*. Greenwich, CT: JAI Press.
———, ——— and WILLIAMS, M. B. (2002). Estimating individual discount rates for Denmark: a field experiment. *American Economic Review*, **92**(5), 1606–17.
HEY, J. D. and ORME, C. (1994). Investigating generalizations of expected utility theory using experimental data. *Econometrica*, **62**(6), 1291–326.
HOLCOMB, J. H. and NELSON, P. S. (1992). Another experimental look at individual time preference. *Rationality and Society*, **4**, 199–220.
JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149–76.
KIRBY, K. N. (1997). Bidding on the future: evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General*, **126**, 54–70.
——— and MARAKOVIĆ, N. N. (1995). Modeling myopic decisions: evidence of hyperbolic delay-discounting within subjects and amounts. *Organizational Behavior & Human Decision Processes*, **64**, 22–30.
——— and ——— (1996). Delay-discounting probabilistic rewards: rates decrease as amounts increase. *Psychonomic Bulletin & Review*, **3**(1), 100–4.
———, PETRY, N. M. and 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.
LOEWENSTEIN, G. (1987). Anticipation and the valuation of delayed outcomes. *Economic Journal*, **97**, 666–84.
PLOTT, C. R. and ZEILER, K. (2005). The willingness to pay–willingness to accept gap, the 'endowment effect,' subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review*, **95** (3), 530–45.
RAINERI, A. and RACHLIN, H. (1993). The effect of temporal constraints on the value of money and other commodities. *Journal of Behavioral Decision Making*, **6**, 77–94.
READ, D. and ROELOFSMA, P. H. M. P. (2003). Subadditive versus hyperbolic discounting: a comparison of choice and matching. *Organizational Behavior and Human Decision Processes*, **91**, 140–53.
SCHOLTEN, M. and READ, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review*, **117**(3), 925–44.
SHELLEY, M. K. (1993). Outcome signs, question frames, and discount rates. *Management Science*, **39**, 806–15.
THALER, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics Letters*, **8**, 201–7.
VON WINTERFELDT, D. and EDWARDS, W. (1986). *Decision Analysis and Behavioral Research*. New York: Cambridge University Press.
WILCOX, N. T. (2008) Stochastic models for binary discrete choice under risk: a critical primer and econometric comparison. In J. COX and G. W. HARRISON (eds), *Risk Aversion in Experiments, Vol. 12, Research in Experimental Economics*. Bingley: Emerald.
——— (2011). 'Stochastically more risk averse': a contextual theory of stochastic discrete choice under risk. *Journal of Econometrics*, **162**(1), 89–104.