Heterogeneity in risky choice behavior in a broad population
von Gaudecker, H.M.; van Soest, Arthur; Wengstrom, E.

Published in:
American Economic Review

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2011

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 10. Oct. 2018
Heterogeneity in Risky Choice Behavior in a Broad Population

By Hans-Martin von Gaudecker, Arthur van Soest, and Erik Wengström*

We analyze risk preferences using an experiment with real incentives in a representative sample of 1,422 Dutch respondents. Our econometric model incorporates four structural parameters that vary with observed and unobserved characteristics: utility curvature, loss aversion, preferences toward the timing of uncertainty resolution, and the propensity to choose randomly rather than on the basis of preferences. We find that all four parameters contribute to explaining choice behavior. The structural parameters are significantly associated with socioeconomic variables, but it is essential to incorporate unobserved heterogeneity in each of them to match the rich variety of choice patterns in the data. (JEL D12, D81)

We describe and analyze an experiment to elicit risk preferences in a representative sample from the Dutch population with more than 1,400 individuals. The main innovation of our work is that we estimate a structural model of the distribution of preferences in the population, distinguishing parameters for risk aversion, loss aversion, and a preference for the timing of uncertainty resolution. The rich nature of the data allows us to account for heterogeneity in all these parameters using a random coefficients model. In addition, we model heterogeneity in the tendency to make random optimization errors that explain why reported choices often are in conflict with fully rational decision making.

Analyzing decisions under risk is an important theme in economic research. Economic theory has generated numerous models for it, starting with the early work of John von Neumann and Oskar Morgenstern (1947). The class of models that combine linearity in probabilities with a nonlinear utility function remains the workhorse in much of modern economics. The seminal contributions of Kenneth J. Arrow (1965)
and John W. Pratt (1964) provided the foundations for measuring individual risk attitudes using the curvature of the utility function. Following this approach, a large literature on the estimation of individual risk aversion preferences using experimental methods has evolved over the last decades (see, for example, Hans P. Binswanger 1980; Charles A. Holt and Susan K. Laury 2002; or Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström 2007). Viewed in the light of expected utility, these studies show that people display considerable risk aversion even over modest stakes.

Starting with Maurice Allais (1953), however, many violations of the basic expected utility models have been documented. Chris Starmer (2000) reviews this literature. One particularly persistent finding in experiments is a greater sensitivity to losses than to gains of similar size. This was widely popularized through prospect theory (Daniel Kahneman and Amos Tversky 1979) and defined formally as a kink of the utility function at the reference point by Veronika Köbberling and Peter P. Wakker (2005). Heterogeneity in utility curvature has been documented in many studies. Important examples include John D. Hey and Chris Orme (1994), Thomas Dohmen et al. (forthcoming), and Steffan Andersen et al. (2008). Nearly all studies that incorporate loss aversion concentrate on estimating mean or median parameters. Recent exceptions include Ernst Fehr and Lorenz Götte (2007) and Simon Gächter, Eric J. Johnson, and Andreas Herrmann (2007), who assume linear utility on each side of the reference point. A novel aspect of our analysis is the joint estimation of individual-specific parameters measuring the utility curvature and the strength of the kink in a structural model.

It has long been recognized that this class of preferences cannot capture what is arguably the most common situation of choice under uncertainty: a decision-maker needs to make a consumption choice before all uncertainty he faces is resolved (Michael Spence and Richard J. Zeckhauser 1972). The class of preferences developed by David M. Kreps and Evan L. Porteus (1978) in order to deal with this deficiency has spurred an enormous amount of literature in the economics of the life cycle (Larry G. Epstein and Stanley E. Zin 1989). More recently, economists have explored a second route to why the timing of uncertainty resolution matters, namely through anticipatory emotions such as hope or anxiety (George Wu 1999; Andrew Caplin and John Leahy 2001). This channel has received considerable attention by experimental economists in recent years (Soo Hong Chew and Joanna L. Ho 1994; Martin Ahlbrecht and Martin Weber 1996; Dan Lovallo and Kahneman 2000; Charles Noussair and Ping Wu 2006; Kfir Eliaz and Andrew Schotter 2007; Frans van Winden, Michal Krawczyk, and Astrid Hopfensitz 2008). This literature has concentrated on estimating average parameters and has produced mixed results.

In our experiment, all payoffs take place three months after the experiment. The timing of uncertainty resolution varies within subjects between the beginning and the end of this period. This allows us to identify, at the individual level, preferences for the timing of uncertainty resolution. We estimate a specification of the Kreps and Porteus (1978) model that explicitly distinguishes parameters for utility curvature, loss aversion, and preferences for the timing of uncertainty resolution.

Most theories of utility are deterministic in nature. Provided knowledge of the relevant parameters, they imply a unique choice from any set of options except for very special cases of indifference. Wherever repeated or sufficiently similar choices allow a violation of revealed preference conditions, a significant fraction of subjects
will violate them. Prominent studies demonstrating this fact include Hey and Orme (1994) and Syngjoo Choi et al. (2007). Explanations for this include changing tastes, lack of motivation, and difficulty understanding the choice problem. Empirical models have therefore added noise to the process that generates observed choices. There has been considerable progress on modeling this noise. Key contributions include David W. Harless and Colin F. Camerer (1994), Hey and Orme (1994), Graham Loomes and Robert Sugden (1995), T. Parker Ballinger and Nathaniel T. Wilcox (1997), Loomes, Peter G. Moffatt, and Sugden (2002), and the papers in a special issue of Experimental Economics (Starmer and Nicholas Bardsley 2005).

We build on this literature and estimate individual-specific error propensities along with the preference parameters. This allows us to assess how much confidence we can have in the estimates of individuals’ preference parameters. Put differently, our estimates yield signal-to-noise ratios for any observed set of choices.

Recently, there has been an increased interest in using economic experiments to draw inference on the distribution of economically important preference parameters in a broad heterogeneous population (among many others, Han Bleichrodt, José Luis Pinto, and Wakker 2001; Harrison, Lau, and Melonie B. Williams 2002; Andersen et al. 2008). There has been a growing concern that the standard recruitment procedure—an experimenter inviting college students via e-mail or posters—restricts sociodemographic variation too severely to allow for meaningful inference on the broad population of interest.

We conduct our experiment on a heterogeneous subject pool that is representative of the Dutch population. Moreover, we model heterogeneity in all preference and error parameters, as a function of observed as well as unobserved characteristics. Our experiment is based on the CentERpanel, an Internet-based household survey in the Netherlands. It has been used before for related questions: Bas Donkers, Bertrand Melenberg, and Arthur van Soest (2001) and Adam S. Booij and Gijs van de Kuilen (2009) use hypothetical questions to estimate risk preference functionals and analyze their associations with observed covariates. Steffen Huck and Wieland Müller (2007) provide evidence that violations of Expected Utility Theory in the Allais paradox occur more frequently in some population subgroups than in others.

Our main results can be summarized as follows. First, utility curvature and loss aversion turn out to be the key determinants of individuals’ choices under risk. Second, the influence of preferences toward the resolution of uncertainty is less important and the median subject is almost neutral between early and late resolution. However, this is not true for everyone and the large preference heterogeneity that we document is likely to stand behind the mixed evidence accumulated so far. Third, while many people exhibit consistent choice patterns, some have very high error propensities. Fourth, preference parameters vary strongly with sociodemographics, but this variation is small compared to the variance ascribed to unobserved heterogeneity. This implies that controls for individual preferences contain useful information in addition to sociodemographics and that it would be desirable to make them more widely available for empirical work based upon micro-data from socioeconomic surveys.

The remainder of this paper is organized as follows. In the next section, we describe the experimental design and motivate the subsequent analysis by means of
aggregate features of the data and by describing the choice behavior of five “benchmark” subjects, highlighting some common and some extreme examples of choice behavior. In Section II we describe the theoretical and empirical models. Section III contains our results, describing average preference parameters and their sociodemographic correlates, the nature of unobserved heterogeneity, and the implications of the choices of the five selected subjects for our estimates of their preference parameters. Section IV concludes.

I. Data and Experimental Setup

We implemented the experiment in the CentERpanel, a Dutch household survey administered via the Internet. In order to avoid selection problems due to lack of Internet access, respondents without a computer are equipped with a box for their television set (and with a television if they do not have one). The panel consists of roughly 2,000 households representative of the Dutch population in terms of observable characteristics. It has rich background information on important demographic and socioeconomic variables. Respondents are reimbursed regularly for their costs of using the Internet and we used the existing reimbursement system to make payments to the participants in the experiment.

We use data on 1,422 individuals in the CentERpanel and provide a description of the covariates and the construction of the sample used for estimation in the Web Appendix. Parallel experiments were conducted with economics students at Tilburg University’s economics laboratory. Von Gaudecker, van Soest, and Wengström (2008) perform an extensive analysis comparing the data from the laboratory and the Internet experiments and find virtually no differences between the behavior of students in the lab and young highly educated subjects in the Internet experiments.

A. Experimental Design

Our experiment uses an adapted version of the well-established multiple price list format, applied earlier by, for example, Binswanger (1980), Tversky and Kahneman (1992), Holt and Laury (2002), and Harrison, Lau, and Rutström (2007). Andersen et al. (2006) provide a detailed description, and we limit ourselves to a brief introduction. Each subject is shown four pairs of lotteries such as the ones presented in Figure 1. We call the components of these pairs option A and option B, where option B always involves at least as much risk as option A. Subjects may opt for either option in each of the four choice tasks. The payoffs of both options do not change, but the probabilities of the high payoff in each option vary from 25 percent to 100 percent as one moves down the screen. The table is designed such that the expected value of option A starts out higher but moves up more slowly than the expected value of option B.

Utility-maximizing participants switch at some point from option A to option B or choose option B throughout the screen. If such consistent behavior is observed, the subject is routed to a screen containing lotteries with the same payoffs, but a finer probability grid. Andersen et al. (2006) recommend using a closely related method and call it “iterated multiple price list,” others have used the term “chained method” for similar strategies (Wakker and Daniel Deneffe 1996). The grid now consists of steps of 10 percentage points located roughly between the subject’s highest choice
of A and his lowest choice of B on the first screen. For a set of lotteries that differ only in the probabilities, we use the term “payoff configuration.” All subjects were given the seven payoff configurations listed in Table 1. For each configuration, subjects made either eight or four decisions, depending on whether their answers on the first screen were consistent or not. Our data therefore constitute an unbalanced panel of 28 to 56 binary choices for each respondent.

A modification compared to previous studies is the inclusion of pie charts as a graphical tool to help describe the probabilities of the outcomes. Pilot experiments showed that this supplement to the verbal descriptions of the decision tasks was appreciated by subjects who were not familiar with probability judgements. Moreover, we restricted the number of decision tasks per screen from the usual 10 to 4 to avoid the need for scrolling. Subjects could not go back to revise decisions on previous screens. We tested the instructions thoroughly. Unlike in typical laboratory settings, there was no experimenter assisting the respondents and answering questions they might have. In order to compensate for this, subjects had access to the instructions and specially designed help screens throughout the experiment.
Participants were allocated to one of three different incentive treatments. One of these was entirely hypothetical, with lotteries based upon the payoff configurations in Table 1. The two treatments with real incentives involved a participation fee paid to everyone who completed the experiment. Additionally, for one in every ten participants in these two treatments, one lottery was randomly selected and played out, and the payoff of that lottery was paid out. The procedure of paying only one out of ten participants is in line with earlier large-scale experimental work such as Harrison, Lau, and Rutström (2007) and Dohmen et al. (forthcoming). Yet, it is worth noting that the probability of a given decision to be selected for payments is lower than in laboratory studies using the random lottery incentive scheme. For choice tasks similar to ours, Andersen et al. (2010) find no significant differences in choices between the two payment schemes, although this may be different for other experimental designs (Guido Baltussen et al. 2008). The iteration introduces a distortion of incentive compatibility for some subjects. Sufficiently risk-tolerant participants (i.e., those whose preferences imply a switch point at or below 75 percent) had an incentive to switch at a higher probability for the high payoff because this would raise expected payoffs in the second step. This does not appear to be an issue, as the iteration procedure was not announced and we do not observe learning in the sense of increasing switch points over the course of the experiment (see Figure A1 in the Web Appendix). In the high-incentive treatment, the completion fee was €15, with lottery payoffs listed in Table 1. In the low-incentive treatment, payoffs and participation fee were one-third of those of the high-incentive treatment. We allocated subjects to one of two randomly determined orderings of the seven payoff configurations.

Some of the lotteries included negative outcomes and we use a zero lottery payoff as the natural reference point. We avoided negative overall payoffs by setting the participation fees equal to the maximum losses that could be incurred. All payoffs

\(^2\) Although this procedure is used by all experimental studies we are aware of, it should be noted that this has the potential drawback that loss aversion might be underestimated, since some respondents may combine payoff and participation fee and use a different reference point so that they never can experience a loss. Two facts suggest that this not so important. First, the high degree of narrow framing documented by Matthew Rabin and Georg Weizsäcker (2009) in a related experiment implies that most subjects will evaluate each screen separately. Second, we still find strong evidence of loss aversion, and the less risky option A always has weakly positive outcomes, so that subjects were able to avoid losses altogether.

### Table 1—Characteristics of the Seven Payoff Configurations

<table>
<thead>
<tr>
<th>Payoff configuration</th>
<th>Option A</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uncertainty</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td>resolution</td>
<td>payoff</td>
</tr>
<tr>
<td>1</td>
<td>early</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>early</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>early</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>early</td>
<td>33</td>
</tr>
<tr>
<td>5</td>
<td>early</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>early</td>
<td>24</td>
</tr>
<tr>
<td>7</td>
<td>late</td>
<td>15</td>
</tr>
</tbody>
</table>

Notes: Subjects were asked to choose between the two options on each row four or eight times, with different probabilities of obtaining the high outcome. These values were shown in the high incentive and hypothetical treatments. For the low-incentive treatment they were divided by three. The order was randomized.
were made three months after the experiment. The timing of uncertainty resolution was set to either directly after the experiment or just before the payment was made. We emphasized to subjects that this concerned only the timing of uncertainty resolution and not the timing of payment: an entire screen of the introduction was devoted to this aspect and the timing of uncertainty resolution featured prominently at the top of each payoff configuration (see Figure 1).

B. Heterogeneity in the Choice Data

In this section we highlight the features of our experimental data that motivate the specification of our empirical model. We first present some descriptive statistics of the complete dataset. Then we describe the choices of five exemplary individuals to show the need to incorporate heterogeneity and the possibility of making inconsistent decisions in the econometric model.

A convenient way to summarize the experimental choices are “switch points,” i.e., the probability at which a subject switches from choosing the safer option A to the more risky option B (see, for example, Holt and Laury 2002). To be more precise, we define a switch point as the highest probability corresponding to an A choice that is lower than the minimum probability at which B is chosen. Figure 2 contains the average of these switch points for each of the seven payoff configurations, using the aggregate data from all three treatments. Looking at each treatment separately does not yield any additional insights, as the patterns of mean switch points across payoff configurations is similar in the three treatments. Following Holt and Laury (2002), all payoff configurations are designed in such a way that switch points would all be 40 percent for individuals who are risk neutral, have no loss aversion, are indifferent to the timing of uncertainty resolution, and make no errors. The much higher values are a clear indication that the average participant is quite different from this risk-neutral benchmark respondent. Average switch points are typically higher if option B involves the possibility of a loss (payoff configurations 3 and 5–7). The timing of uncertainty resolution for option B (after three months in payoff configurations 4, 5, and 7, and immediately otherwise) does not influence average choices in a clear way. In particular, configurations 4 and 5, where the uncertainty is revealed early in option A and late in option B, have the lowest and highest average switch points.

Since we let subjects make independent choices on each row, instead of forcing subjects to have a unique switch point, we allowed them to behave in an inconsistent manner. More specifically, three types of choice patterns are inconsistent with virtually all deterministic utility specifications. First, a dominance violation occurs if somebody chooses option B when the probability for the high outcome is zero, or option A when this probability is one. The second type of inconsistency emerges when subjects switch back to option A on the same screen after having chosen B for a lower \( p^\text{high} \). The third consists of such back-switching behavior or different choices for identical lotteries on the initial screen and the follow-up screen.

This is one of several ways to handle the optimization errors described below. Some alternatives are explored in von Gaudecker, van Soest, and Wengström (2008). They lead to the same ranking of payoff configurations.

For example, if on the first screen a subject switches from A to B in the second row at \( p^\text{high} = 0.5 \), the rows on the second screen have \( p^\text{high} = \{0.2, 0.3, 0.4, 0.5\} \). An inconsistency then arises if the subject chooses B at probability 0.2 or if he chooses A at probability 0.5.
In Table 2, the frequencies of these three types of violations are broken down by payoff configuration. Overall, optimization errors are found in 34.6 percent of the payoff configurations (where a maximum of one violation is counted per payoff configuration). Inconsistencies between the first and the second screen are the most common type of inconsistency; 24.7 percent of the times subjects were routed to the second screen, they made a choice on the second screen that was not consistent with their choices of the first screen. Notably, dominance violations are also common and found on 12.3 percent of the screens in which such a violation was possible. Violations due to switching back and forth within the same screen were less frequent and occurred on 4.9 percent of the screens. As noted by a referee, the pattern shows that our design probably reduced the number of inconsistencies compared to a setting with only one choice per screen. There is some variation in error rates between payoff configurations, and we can only speculate on the sources of differences among the screens. One factor that seems to influence error rates is the possibility to make losses or receive a zero payoff. The payoff configurations with only strictly positive outcomes, 1 and 2, are associated with the highest error rates. One very tentative explanation may hence be that fear of losses increases subjects’ cognitive effort, resulting in more consistent decision making.

5 The three different types of violations are described in detail by von Gaudecker, van Soest, and Wengström (2008), who also show that the high number of dominance violations when compared to lab experiments is due to the composition of the sample (with more low-educated and older people than the usual lab experiment) rather than the Internet environment.
Another issue related to the pattern of inconsistencies is how the error frequency develops as subjects gain experience in the course of the experiment. Although there is no feedback, one might expect some learning. Table A2 in the Web Appendix reveals that there are no clear effects of experience on error rates. The empirical model will take these findings into account, but, in addition, the averages do not show the enormous amount of heterogeneity in individual choices. This is why we also describe the choices of five selected individuals, labelled respondents R1–R5. We will also come back to these five examples later, to analyze what our model estimates and their choices imply concerning their preference parameters and their tendencies to make mistakes.

The pattern of R1, described in panel A of Figure 3, bears some similarity to the average figures. We see risk aversion for all configurations. The most risky choices are made in payoff configurations 1 and 2, characterized by early resolution and nonnegative payoffs. More safe choices were made in configuration 3 (with a probability of a negative payoff in option B and early resolution of uncertainty in both options) and configuration 4 (late resolution of option B, positive outcomes only). R1 makes inconsistent decisions in configurations 5 and 7 (nonmonotonicities) and 6 (different choices when faced with the same decision twice). The last could, in principle also reflect indifference between the two options and is therefore not necessarily an inconsistency. However, preference parameters will have continuous distributions in our econometric model, and the probability of drawing a respondent who is indifferent between two options is always equal to zero.

The other four benchmark respondents have choice patterns that are less common but not exceptional. The choices of R2 (panel B of Figure 3) are very consistent and mostly safe. In all cases where the outcome set included a nonzero probability of a negative or zero payoff, R2 avoided this possibility by choosing the safer lottery. R3 (panel C) makes the same choices as R2 for payoff configurations 5 and 7, but the choices on the other screens suggest that these are generated from different underly-

<table>
<thead>
<tr>
<th>Payoff configuration</th>
<th>Dominance</th>
<th>Within</th>
<th>Between</th>
<th>Any type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14.3 percent</td>
<td>8.1 percent</td>
<td>19.3 percent</td>
<td>34.1 percent</td>
</tr>
<tr>
<td>2</td>
<td>11.2 percent</td>
<td>7.5 percent</td>
<td>26.4 percent</td>
<td>37.1 percent</td>
</tr>
<tr>
<td>3</td>
<td>16.4 percent</td>
<td>4.5 percent</td>
<td>16.3 percent</td>
<td>30.9 percent</td>
</tr>
<tr>
<td>4</td>
<td>10.4 percent</td>
<td>6.4 percent</td>
<td>29.6 percent</td>
<td>39.0 percent</td>
</tr>
<tr>
<td>5</td>
<td>16.0 percent</td>
<td>5.2 percent</td>
<td>17.7 percent</td>
<td>34.2 percent</td>
</tr>
<tr>
<td>6</td>
<td>16.0 percent</td>
<td>4.5 percent</td>
<td>19.2 percent</td>
<td>32.9 percent</td>
</tr>
<tr>
<td>7</td>
<td>12.6 percent</td>
<td>8.2 percent</td>
<td>20.2 percent</td>
<td>34.1 percent</td>
</tr>
<tr>
<td>Total</td>
<td>12.3 percent</td>
<td>4.9 percent</td>
<td>24.7 percent</td>
<td>34.6 percent</td>
</tr>
</tbody>
</table>

**Notes:** The figures represent the average percentage of errors given the number of possible violations, by payoff configuration and type of error. The numbers for the dominance violations are obtained by first dividing the number of dominance violations for each subject by the number of screens shown to the subject on which dominance violations could be made, and then taking the average of this fraction across all subjects. The figures for the within category are calculated in a similar way by taking the average (across subjects) of the number of within violations divided by the total number of screens shown to the subject. The numbers for the between column are obtained by dividing the number of between errors for each subject by the number of times the second screen was displayed to the subject, and then taking the average across all subjects. Only one inconsistency was counted per payoff configuration.
ing preferences: in the other configurations the choices of R3 are much riskier than those of R2. This suggests that R3’s safe choices in these configurations are driven by a combination of moderate risk aversion, loss aversion, and a preference for early resolution of uncertainty.

R4 in panel D shows the opposite of R2: this subject exhibits risk-loving behavior on all screens. The substantial number of such patterns in the data justifies a specification of heterogeneous preferences that can accommodate this kind of
behavior. Finally, R5 in panel E makes seemingly idiosyncratic choices without any clear pattern. Although few individuals behave like this, there are many intermediate cases with a pattern suggesting an error tendency between that of subjects R2 and R5. The wide range of choice patterns with inconsistencies makes careful modelling of heterogeneity as well as errors important (e.g., Moffatt 2005 for a similar motivation). At the end of the day, the estimates will tell us about the informational content of observed choices for individual preferences for every respondent in our sample. In contrast to other studies, our data suggest no clear cut-off point to exclude some individuals a priori from the sample (see, e.g., Mohammed Abdellaoui, Carolina Barrios, and Wakker 2007; or Choi et al. 2007).

II. Theoretical Framework and Empirical Model

In this section, we first lay out the utility specifications that form the basis of our econometric analysis. We start with an expected utility of income specification, which incorporates loss aversion, and add a temporal component in a two-period model. While all payments are made in the second period, uncertainty may be resolved either in the first or the second period. We use a parsimonious version of the Kreps and Porteus (1978) model to allow for preferences toward the timing of uncertainty resolution. The last step consists in developing an empirical model that allows for sufficient heterogeneity.

A. A Simple Model of choice under Risk

We start from a standard expected utility formulation with an exponential utility function:

\[ u(z, \gamma) = -\frac{1}{\gamma} e^{-\gamma z}, \]

where \( z \in \mathbb{R} \) denotes a lottery outcome and \( \gamma \in \mathbb{R} \) is the Arrow-Pratt coefficient of absolute risk aversion. We prefer exponential utility over power utility because it lends itself better to incorporating the timing of uncertainty resolution, as discussed below. In the Web Appendix, we compare the results to alternative specifications using power utility, and find a better fit for exponential utility.

The first extension of (1) is to incorporate loss aversion, following prospect theory (Kahneman and Tversky 1979) and in line with the widely recognized stylized fact that “losses loom larger than gains” (see, e.g., Starmer 2000 for a review). In line with the literature, we augment (1) with a loss aversion parameter \( \lambda \in \mathbb{R}_+ \):

\[ u(z, \gamma, \lambda) = \begin{cases} -\frac{1}{\gamma} e^{-\gamma z} & \text{for } z \geq 0 \\ \frac{\lambda - 1}{\gamma} - \frac{\lambda}{\gamma} e^{-\gamma z} & \text{for } z < 0. \end{cases} \]

The degree of loss aversion is measured by the ratio of the left and the right derivatives of the utility function at zero, as suggested by Köbberling and Wakker (2005).
Equation (2) entails two main departures from the original prospect theory specification. First, it does not involve nonlinear probability weighting because our goal is to estimate individual-level parameters, and the dimension of the estimation problem is large already. Adding a parameter that is highly collinear with utility curvature in our experimental setup would result in an infeasibly large number of parameters, given the structure of our data. Furthermore, typical probability weighting functionals develop the highest impact at extreme probabilities, which are absent from our experiment. Second, prospect theory’s original utility function is concave for gains and convex for losses. In contrast to this, (2) assumes the same type of curvature on the whole real line. This is primarily motivated by some recent empirical results that call prospect theory’s original utility curvature findings for mixed gambles into question (Baltussen, Thierry Post, and Pim van Vliet 2006). Ideally, we would estimate separate parameters for the gain and loss domains, but our experiment does not have enough variation in negative outcomes to do this. In the Web Appendix, we present estimates based on alternative functional form assumptions that include prospect theory–type preferences. The results suggest that assuming the same type of curvature on the gains and loss domains is preferred to the original prospect theory specification. The second reason behind using (2) is that this specification can much easier be built into an environment capturing preferences for early or late uncertainty resolution.

In order to model such preferences, we adopt the general framework of Kreps and Porteus (1978). In line with our experimental setup we consider a two-period setting. All decisions are made in the first period and payments are made in the second period. The outcome of a gamble is revealed either in period 1, directly after all choices have been made (early resolution), or at the time of the payments in period 2 (late resolution).^6^ Assume that agents first calculate period 2 utility for all outcomes based on a function $v(z, \cdot)$, where $z$ is the payoff and the dot replaces the preference parameters. Thereafter, agents are assumed to use a continuous and strictly increasing weighting function $h(\cdot)$ to calculate their first-period utility, with period 2 utility $v$ as its argument. The period 1 utility of a degenerate lottery that gives a certain outcome in period 2 is then given by $h(v(z, \cdot))$. The evaluation of nondegenerate lotteries hinges on the timing of uncertainty resolution: let $V(\pi)$ denote the period 1 utility of a lottery $\pi$ with payoffs in period 2. Then $V$ is given by

$$
V(\pi) = \begin{cases} 
\mathbb{E}[h(v(z, \cdot))] & \text{for early uncertainty resolution} \\
|h(\mathbb{E}[v(z, \cdot)])| & \text{for late uncertainty resolution.}
\end{cases}
$$

Note that the expectations operator is always applied to the quantity that is known at the end of period 1. If uncertainty resolves early, the decision maker applies the weighting function to the utility of the specific outcomes of $\pi$. If the outcome of $\pi$
remains uncertain until the second period, he applies the weighting function to its expected value. Kreps and Porteus (1978) show that \( h \) is strictly convex (concave) if and only if the decision maker always prefers early to late (late to early) resolution, and is linear if the decision maker is indifferent.

We use a power specification for \( h(\cdot) \) and a slightly modified version of \( (2) \) for period 2 utility. The modifications become necessary to handle risk aversion and risk lovingness, and negative payoffs within the Kreps-Porteus framework. We explain them carefully in the Appendix. The curvature of \( h(\cdot) \) is governed by a parameter \( \rho \), which thus captures preferences toward uncertainty resolution. A value of \( \rho \) smaller (greater) than one indicates a preference for late (early) resolution, and \( \rho = 1 \) indicates indifference.

**B. Econometric Implementation**

Based on this specification of utility, we formulate structural econometric models that can be estimated by maximum likelihood. The models allow for individual heterogeneity in preference parameters and in the tendency to make errors. The heterogeneity can be captured by observed characteristics (“observed heterogeneity”) or not (“unobserved heterogeneity”). Assume that individual \( i \in \{1, \ldots, N\} \) faces \( j \in \{1, \ldots, J_i\} \) dichotomous choices between two binary lotteries \( \pi_j^A = (A_j^{low}, A_j^{high}, A_j^{late}, p_j^{high}) \) and \( \pi_j^B = (B_j^{low}, B_j^{high}, B_j^{late}, p_j^{high}) \). Each lottery is characterized by a low and a high outcome and whether uncertainty resolves late or not. Each pair of lotteries shares a common probability of the high outcome. Let \( y_{ij} = 1 \) if the individual opts for \( \pi_j^B \), and \( y_{ij} = 0 \) otherwise. Define the difference in certainty equivalents of the two lotteries in decision task \( j \) as

\[
\Delta \text{CE}_{ij} = \text{CE}(\pi_j^B, \gamma_i, \lambda_i, \rho_i) - \text{CE}(\pi_j^A, \gamma_i, \lambda_i, \rho_i),
\]

where \( \text{CE}(\pi_j^k, \gamma_i, \lambda_i, \rho_i), \; k = A, B \) is the period 1 certainty equivalent of lottery \( \pi_j^k \) given the utility function defined by (3) and (8)–(10) with the individual-specific parameters \( \gamma_i, \lambda_i, \) and \( \rho_i \). It is straightforward to derive an exact analytical expression for \( \text{CE}(\cdot) \) under our functional form assumptions (see Web Appendix).

A perfectly rational decision maker would choose \( \pi_j^B \) if and only if \( \Delta \text{CE}_{ij} > 0 \). As a first step to allow for stochastic decision making, we add so-called Fechner errors (see, e.g., Loomes 2005) to the CE comparison and model the individual’s choice as

\[
y_{ij} = \mathbb{I}\{\Delta \text{CE}_{ij} + \tau \varepsilon_{ij} > 0\},
\]

where \( \mathbb{I}\{\cdot\} \) denotes the indicator function. We assume that the \( \varepsilon_{ij} \) are independent of each other and of the random coefficients driving the utility function, and follow a standard logistic distribution. The mechanism behind these Fechner errors is that as \( \Delta \text{CE} \) becomes small, the likelihood of choosing the “wrong” lottery increases. The parameter \( \tau \in \mathbb{R}_+ \) governs the individual’s probability to make this type of “mistakes”; the probability of such a mistake increases with \( \tau \).

The use of certainty equivalents in (4) leads to a meaningful interpretation of \( \Delta \text{CE} \) in monetary terms. This means that \( \tau \) has an intuitive interpretation. For example, if the difference in valuations of two lotteries (i.e., the “cost” of making
an error relative to the individual utility function parameters) is $\Delta CE = €10$ and $\tau = 4$, the probability to choose the higher-valued lottery is 0.92. For $\Delta CE = €1$, this probability is only 0.56. Using certainty equivalents facilitates comparisons across subjects—using utility differences directly as, for example, Hey and Orme (1994), implies that the scale of the errors depends on the preference parameters and makes comparisons between subjects more difficult. We can thus estimate $\tau$ along with the preference parameters, contrary to typical applications of random utility models where the scale of utility is identified only by fixing the standard deviation of the error term.

In addition to adding the errors $\tau\varepsilon_{ij}$, we allow for the possibility that subjects choose at random in any given task, following Harless and Camerer (1994). The propensity to do so is governed by the individual-specific “trembling hand” parameter $\omega_i \in [0, 1]$, and the probability of the observed choice $y_{ij}$ of individual $i$ in choice situation $j$, given all the individual-specific parameters, is given by

$$l_{ij}(\pi_j^A, \pi_j^B, Y_{ij}, \tau, \gamma_i, \lambda_i, \rho_i, \omega_i)$$

$$= (1 - \omega_i) \Lambda((2Y_{ij} - 1)\frac{1}{\tau}\Delta CE_{ij}(\pi_j^A, \pi_j^B, \gamma_j, \lambda_j, \rho_j)) + \frac{\omega_i}{2},$$

where $\Lambda(t) = (1 + e^{-t})^{-1}$ stands for the cumulative standard logistic distribution function.

For the sake of a parsimonious and easily interpretable model, we restrict $\tau$ to be the same for all individuals, while allowing subjects to vary in their probability to make random choices. Alternative error specifications are possible in principle (see the references in introduction to this paper) but beyond the scope of the current paper. For example, one might argue that $\tau$ should also be individual specific, but in practice it appears to be difficult to estimate heterogeneity in $\tau$ and $\omega$ separately (although both are identified, in theory).

We use a random coefficients model in order to estimate the distribution of the individual-specific parameters $\gamma_i, \lambda_i, \rho_i$, and $\omega_i$ in the population. This is a natural way of incorporating observed and unobserved heterogeneity directly. An alternative is to first estimate the parameters for each individual separately (Hey and Orme 1994) and then regress the results on sociodemographic characteristics (Dohmen et al. forthcoming). Our approach has a better econometric justification than this alternative approach, and is much easier to do in the case of multiple parameters and so many choices for each respondent. In contrast to the finite mixture model of Harrison and Rutström (2009), we use a continuous distribution of the parameters of interest. The reason is that finite mixture models have difficulty handling a large number of potential values for the parameters and a small set of values seems insufficient to explain the very heterogeneous choice behavior illustrated in Section IIB. Our modeling of unobserved heterogeneity is similar to that of Anna Conte, Hey, and Moffatt (forthcoming); a difference is that we include observed heterogeneity in addition to unobserved heterogeneity.

---

7 See Charles Bellemare, Sabine Kröger, and van Soest (2008) for a similar modelling approach.
In order to work with a concise notation, define

\[ \eta_i = g_\eta(X_i^\eta \beta^\eta + \xi_i^\eta), \quad \eta_i \in \{ \gamma_i, \lambda_i, \rho_i, \omega_i \}, \]

where \( \eta_i \) denotes one of the four individual specific parameters, \( X_i^\eta \) are \( 1 \times K^\eta \) vectors of regressors, \( \beta^\eta \) are \( K^\eta \times 1 \) parameter vectors, and \( \xi_i^\eta \) are the unobserved heterogeneity components of the parameters. The first element of each \( X_i^\eta \) contains 1. The functions \( g_\eta(\cdot) \) are used to impose the theoretical restrictions on the individual specific parameters. For \( \gamma \), this is just the identity function; for \( \lambda \) and \( \rho \), it is the exponential function, guaranteeing that these parameters are positive. For \( \omega \), it is the logistic distribution function, guaranteeing that \( \omega \) is always between zero and one.

We write \( g(X_i \beta + \xi_i) \) for the vector of these four functions.

We assume that \( \xi_i = (\xi_i^\gamma, \xi_i^\lambda, \xi_i^\rho, \xi_i^\omega)' \) follows a jointly normal distribution independent of the regressors. The regressor matrix contains a dummy for the hypothetical high-incentives treatment to capture the potential effect of giving hypothetical versus real payoffs. Preliminary estimations showed that the difference between low- and high-incentive treatments is better captured by a multiplicative specification than by adding a low-incentive dummy to \( X \). For the low-incentive treatment, we therefore multiply all slope coefficients as well as the standard deviations of the unobserved heterogeneity terms by the same parameter \( \beta^\eta_\text{low incentive} \).

Defining \( \xi^* = (\Sigma')^{-1}\xi \), where \( \Sigma' \Sigma \) is the covariance matrix of \( \xi \), we can express the likelihood contribution of subject \( i \) as

\[ l_i = \int_{\mathbb{R}^4} \prod_{j \in I_i} l_{ij} \left( \pi^A_j, \pi^B_j, Y_{ij}, \tau, g(X_i \beta + \xi) \right) \phi(\xi^*) d\xi^*, \]

where \( l_{ij} \) is the probability given in (5) and \( \phi(\cdot) \) denotes the joint standard normal probability density function with appropriate dimension (4 in this case). The log likelihood is given by the sum of the logarithms of \( l \) over all respondents in the sample and can be maximized by standard methods to obtain the maximum likelihood estimates. The integral in equation (7) does not have an analytical solution, and we approximate it using standard simulation techniques. In particular, we employ Halton sequences of length \( R = 1,000 \) per individual (Kenneth E. Train 2003). We employ the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with numerical derivatives to maximize the log-likelihood function. The variance-covariance matrix of the parameter estimates is based on the outer product of gradients. Standard errors for transformed parameters are calculated using the delta method.

**III. Results**

We present our results in four stages. First, we show that our model’s average parameter estimates can explain the stylized facts in Figure 2. Second, we describe the estimated population distributions of the structural parameters. Third, we
investigate how much of the total heterogeneity is accounted for by observed factors. Fourth, we move to the individual level and illustrate how the choices of respondents R1–R5 translate into preference parameter estimates.

### A. Aggregate Features of the Data

Table 3 contains the estimates for the four parameter vectors \( \beta_\eta, \eta \in \{\gamma, \lambda, \rho, \omega\} \) for models that contain only an intercept and treatment variables. The parameter vector for \( \gamma \) shows a very precisely estimated concave utility function. It is essentially the same in the high-incentive and the hypothetical treatments but substantially larger in the low-incentive treatment. For the low-incentive treatment, we report the multiplicative coefficient \( \beta_{\text{low incentive}} \). Values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter. We therefore find higher absolute risk aversion in the low-incentive treatment (but lower relative risk aversion—see the results with power utility specifications in Tables A33–A38 of the Web Appendix). An increase in the estimate of an exponential utility coefficient when the payoff scale is reduced was also found by Holt and Laury (2002). Whereas they focused on finding a functional form that provides a good approximation to aggregate behavior over varying stakes, we are mainly interested in heterogeneity across subjects at a given level of payoffs. As the comparisons to power utility formulations provided in the Web Appendix show, exponential utility is much better suited for this purpose than the commonly considered alternative of power utility.

For the remaining parameters, it is important to note that the values in Table 3 are all on the original parameter scale—the constant terms are simply \( g_\eta(\beta_0^\gamma) \). The first element in each column, therefore, represents the median (preference or error) parameter in the high-incentive treatment. For the hypothetical treatment, the table shows \( g_\eta(\beta_0^\gamma + \beta_0^\text{hypothetical}^\lambda) - g_\eta(\beta_0^\gamma) \), the (partial) effect of moving from the high-incentive to the hypothetical treatment on the median parameter value.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>( \gamma )</th>
<th>( \lambda )</th>
<th>( \rho )</th>
<th>( \omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0323***</td>
<td>2.38***</td>
<td>1.01</td>
<td>0.0831***</td>
</tr>
<tr>
<td>(0.0010)</td>
<td>(0.159)</td>
<td>(0.0250)</td>
<td>(0.0082)</td>
<td></td>
</tr>
<tr>
<td>Hypothetical treatment</td>
<td>−0.0005</td>
<td>1.31***</td>
<td>−0.0958**</td>
<td>0.0071</td>
</tr>
<tr>
<td>(0.0015)</td>
<td>(0.318)</td>
<td>(0.0405)</td>
<td>(0.0106)</td>
<td></td>
</tr>
<tr>
<td>Low-incentive treatment (^\dagger)</td>
<td>2.77***</td>
<td>0.861***</td>
<td>0.999</td>
<td>1.11</td>
</tr>
<tr>
<td>(0.0868)</td>
<td>(0.0290)</td>
<td>(0.0015)</td>
<td>(0.143)</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Number of observations is 1,422. Estimation follows (7). Regression coefficients are transformed back to the original scale. In other words, the constant is defined by \( g_\eta(\beta_0^\gamma) \) and represents median parameters in the high-incentive treatment. For the hypothetical treatment, the tables show \( g_\eta(\beta_0^\gamma + \beta_0^\text{hypothetical}) - g_\eta(\beta_0^\gamma) \), the (partial) effect of moving from the high-incentive to the hypothetical treatment on the median parameter value.

\( ^\dagger \) The low incentive treatment enters multiplicatively and we report the nontransformed coefficients, i.e., \( \beta_{\text{low incentive}}^\lambda \). Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter.
to the hypothetical treatment on the median parameter value. For the low-incentive treatment, we continue to report $\beta_{\text{low incentive}}$ since it enters multiplicatively.

The estimated median parameter for loss aversion, depicted in the second column of Table 3, is 2.38 in the high-incentive treatment, well in line with previous estimates. It is larger in the hypothetical treatment, probably because of the show-up fee paid in the real treatments—some subjects may have (partly) taken this into account so that they do not see the negative payoff as a loss. The loss aversion coefficient is lower in the low-incentive treatment. The third column of Table 3 reveals that the median coefficients for $\rho$ are precisely estimated and not significantly different from one in the treatments with real incentives and slightly lower in the hypothetical treatment. Hence, the median subject appears to be close to indifferent for the timing of uncertainty resolution.

The magnitudes of the preference parameters are difficult to interpret directly. We follow Choi et al. (2007) and report the risk premia $RP(\pi) = \mathbb{E}[\pi] - CE(\pi)$ of standardized gambles implied by certain parameter constellations. To remain in line with the range of our payoffs, we use the gambles $\pi_1 = (25, 65, 0.5)$ and $\pi_2 = (-15, 25, 0.5)$, with both early and late resolution of uncertainty. Since $RP(\pi) \in [-20, 20]$ for both lotteries, the risk premia are directly comparable and their difference illustrates the impact of the loss aversion coefficient. The magnitudes of $\gamma$ and $\lambda$ are substantial in terms of the implied risk premia: they amount to €5.95 for $\pi_1$ and to €10.59 for $\pi_2$; the impact of $\rho$ is negligible.

Comparing risk premia for $\pi_1$ is also a way to compare our estimates to those found in the literature. Holt and Laury’s (2002) power-expo function estimates based on a wide range of payoffs imply a risk premium of about €2.79. The CRRA interval of their median subject’s choices in their 20x treatment (with payoffs in the same range as ours) leads to an interval between €1.90 and €3.20. The estimates of Choi et al. (2007) for the median subject imply risk premia of €3.34 or €5.50 for the Gul-CRRA and Gul-CARA cases, respectively. In a representative sample of the Danish population, the estimates of Harrison, Lau, and Rutström (2007) using a CRRA functional lead to risk premia of €3.15. Our own estimates of about €6 are slightly higher than the Choi et al. (2007) estimates using an exponential type utility function, and are substantially higher than the other estimates that use power type utility functions (or, in the case of Holt and Laury 2002, more general functions that are close to power utility for the estimated parameter values). The higher-risk premia for the exponential utility seem to be a general characteristic in these settings. We also see the same drop in risk premia as in Choi et al. (2007) when we estimate a prospect theory–type model based on a power utility function (see the Web Appendix, in particular Table A39). The risk premium for the median subject is then €3.43, closely in line with the findings in the other studies. When comparing across studies, one has to bear in mind that stakes and other experimental procedures differ. Compared to the studies cited above, the stakes in our experiment (€−15 to €87) lie in the intermediate range, being somewhat higher than the ones in Choi et al. (2007) ($0 to $50), roughly the same as in Holt and Laury (2002) ($0 to $77), but lower than in Harrison, Lau, and Rutström (2007) (at the time of experiment equivalent to between €7 and €604).

As shown in the last column of Table 3, the median random choice propensity ($\omega$) is about 8.3 percent. This can be related to the estimate for $\tau$ shown in Table 4, which is about 4. Taken together, the parameters imply that if $\Delta CE = €10$, the
probability to choose the higher-valued lottery is 0.88; if \( \Delta CE = €1 \), this probability is only 0.55. These results correspond to the substantial error rates found in other studies using nonstudent samples. See, for example, Nicolas de Roos and Yianis Sarafidis (2010) or Huck and Müller (2007). Comparing lab and Internet data, von Gaudecker, van Soest, and Wengström (2008) conclude that the high error rate is indeed due to the composition of the sample rather than the Internet environment.

The random choice probabilities are the same in the hypothetical treatment and in the high-incentive treatment. They are higher in the low-incentive treatment, but the parameter \( \tau \) is much smaller there. The implication for the overall tendency to make a suboptimal choice therefore varies with the characteristics of the lotteries involved.

### B. Distribution of Preferences and Errors

Figure 4 depicts the estimated population distributions of the four random parameters in our model for the high-incentive treatment. The graphs are based on Table 3, i.e., they account for observed as well as unobserved heterogeneity.

Given the heterogeneity in individual choices that we showed in Section IIB, the large dispersion in the four parameter distributions is hardly surprising. The large mass in the right tail of the distribution of \( \lambda \) implies that relying on a benchmark value between two and three for the loss aversion parameter in model calibrations, something that seems common in the literature (see e.g., Shlomo Benartzi and Richard H. Thaler 1995), is hard to justify. The estimated distribution of \( \omega \) shows that most choices are driven by utility comparisons and therefore provide

<table>
<thead>
<tr>
<th>Table 4—Standard Deviations of the Random Coefficients, ( \tau ), and Log-Likelihoods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimnal set of covariates</td>
</tr>
<tr>
<td>( \sigma_\gamma )</td>
</tr>
<tr>
<td>( (0.001) )</td>
</tr>
<tr>
<td>( \sigma_\lambda )</td>
</tr>
<tr>
<td>( (0.042) )</td>
</tr>
<tr>
<td>( \sigma_\rho )</td>
</tr>
<tr>
<td>( (0.024) )</td>
</tr>
<tr>
<td>( \sigma_\omega )</td>
</tr>
<tr>
<td>( (0.090) )</td>
</tr>
<tr>
<td>( \tau )</td>
</tr>
<tr>
<td>( (0.068) )</td>
</tr>
<tr>
<td>( \tau_{\text{Low Inc}} )</td>
</tr>
<tr>
<td>( (0.008) )</td>
</tr>
<tr>
<td>Log-Likel</td>
</tr>
</tbody>
</table>

*Notes: Number of observations is 1,422. Estimation follows (7) based on the utility function defined by (A3) and (A1). The entries for \( \sigma_\gamma \) are the standard deviations of the untransformed normal distributions of the random coefficients.

† The low incentive treatment enters multiplicatively. Coefficient values smaller than one indicate a lower value of \( \tau \).

---

Table 4 presents the underlying estimates for \( \Sigma \). We report only the results imposing a diagonal structure on \( \Sigma \); substantive results are very similar for a general variance-covariance matrix, but the estimates of the off-diagonal elements of \( \Sigma \) were very inaccurate; see Web Appendix, Tables A12–A14.
information on preferences. Still, a nonnegligible fraction of the populations often chooses at random.

Since the parameters themselves are difficult to interpret, we again consider the risk premia. Figure 5 plots the risk premia for $\pi_1 = (25, 65, 0.5)$ and $\pi_2 = (-15, 25, 0.5)$ at various quantiles of the parameter distributions, showing that the variation in preference parameters also induces substantial heterogeneity in terms of risk premia. The horizontal lines in panels A–B plot the risk premia for the median preference parameters and for early resolution of uncertainty. The two sets of bars show what happens if each parameter separately is set to its tenth or ninetieth percentile.\footnote{More detailed numbers are given in Table A39 of the Web Appendix.} For $\pi_1$, the median subject demands a risk premium of about €6 Euros (see panel A). Changing $\gamma$ to its tenth percentile gives a negative risk premium of €$-3.28$, while at the ninetieth percentile the risk premium is €11.85. The risk premium of $\pi_1$ does not depend on $\lambda$ or $\rho$. The picture is quite different for $\text{RP}(\pi_2)$, as shown in panel B. The baseline risk premium is now €10.59 and the largest heterogeneity stems from the loss aversion parameter. The risk premium becomes negative when $\lambda$ is set to its tenth percentile. Because of the already high median value of $\lambda$, the effect of increasing it to a higher value than the median is much less. Finally, we see that due to the additive term in the definition of $v(z, \gamma, \lambda, \rho)$ in (8) for $z < 0$, there are slight changes in the risk premium when $\rho$ is changed, in spite of the fact that only early resolution of uncertainty is considered.
Panels C and D depict the case of late resolution. The median subject is almost indifferent with respect to the uncertainty resolution timing: risk premia are only €0.01 lower than for early resolution. The only substantial difference with respect to the early resolution case concerns the impact of changing $\rho$: moving it to the tenth or ninetieth percentile now has an impact of up to €3.74 on the risk premia. This effect is substantial, but smaller than the effects of heterogeneity in other parameters found above.

C. Can Observed Factors Account for Parameter Heterogeneity?

An important question for empirical applications is the extent to which observed covariates can account for heterogeneity in structural parameters. In our case, we can phrase this question in terms of the reduction in variation of $\xi_i$ if we control for typically available covariates in equation (6). We can check this by comparing the distribution implied by the observed covariates only (setting $\xi_i$ to zero) to the overall distribution. From Figure 6, it is evident that our large set of controls accounts for only a small part of the total variation in the preferences. For example, 90 percent of the conditional medians for the risk preference parameter $\gamma$ lie in a range that accounts for less than 30 percent of the distribution of $\gamma$ including the unobserved
heterogeneity component. Only 1 percent of the conditional medians of $\omega$ imply a random choice propensity larger than 0.34, compared to 21 percent of the values of $\omega$ themselves. Furthermore, we note that for the overall distribution of parameters, it hardly makes a difference whether we account for observables or not. The two resulting sets of graphs are virtually identical (see Figure A5 in the Web Appendix). The individual choices thus contain much more information than what is captured by sociodemographic groups.

Notwithstanding the small degree of overall heterogeneity that can be accounted for by observables, the gray lines in Figure 6 also make clear that there are important differences between socioeconomic groups. This is reflected in a large number of significant coefficients in Table 5. The associations broadly reflect those found in the literature and we briefly highlight some of them.

Our results support the common finding that women are more risk averse than men (Rachel Croson and Uri Gneezy 2009; Lex Borghans et al. 2009). We further find a positive age and a negative education gradient for risk aversion. The associations with income and wealth do not reveal a clear pattern. Being the household’s financial administrator is associated with lower risk aversion. These findings are in line with those of existing studies, cf. Donkers, Melenberg, and van Soest (2001) or Dohmen et al. (forthcoming); Harrison, Lau, and Rutström (2007) find hardly any significant effects but this may be due to their moderate sample size; Daniel J. Benjamin, Sebastian A. Brown, and Jesse M. Shapiro (2006) find a negative
For the hypothetical treatment, the tables show peaks between 35 and 44 years of age and then declines. It also declines with our findings for education. Loss aversion is higher among women, and its age pattern peaks between 35 and 44 years of age and then declines. It also declines with

Table 5—Estimated Parameters for Model with Full Set of Covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>( \gamma )</th>
<th>( \lambda )</th>
<th>( \rho )</th>
<th>( \omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0316***</td>
<td>2.960***</td>
<td>1.040</td>
<td>0.102***</td>
</tr>
<tr>
<td>Female</td>
<td>0.0079***</td>
<td>0.569*</td>
<td>0.0190</td>
<td>0.0197*</td>
</tr>
<tr>
<td>Age 35–44</td>
<td>-0.0004</td>
<td>1.150**</td>
<td>-0.0056</td>
<td>0.0208</td>
</tr>
<tr>
<td>Age 45–54</td>
<td>0.0046**</td>
<td>-0.726*</td>
<td>0.0040</td>
<td>0.0482**</td>
</tr>
<tr>
<td>Age 55–64</td>
<td>0.0035</td>
<td>-0.0420</td>
<td>-0.136*</td>
<td>0.141***</td>
</tr>
<tr>
<td>Age 65+</td>
<td>0.0150***</td>
<td>-0.667</td>
<td>-0.210***</td>
<td>0.293***</td>
</tr>
<tr>
<td>Hi Sec Educ/Int Voc Train</td>
<td>-0.0065***</td>
<td>0.204</td>
<td>-0.0430</td>
<td>-0.0445***</td>
</tr>
<tr>
<td>Higher Voc Train</td>
<td>-0.0020</td>
<td>0.159</td>
<td>-0.0904</td>
<td>-0.0533***</td>
</tr>
<tr>
<td>University</td>
<td>-0.0165***</td>
<td>0.698</td>
<td>-0.0112</td>
<td>-0.0622***</td>
</tr>
<tr>
<td>Income €22k–40k</td>
<td>-0.0013</td>
<td>-0.409</td>
<td>0.0275</td>
<td>-0.0106</td>
</tr>
<tr>
<td>Income €40k+</td>
<td>0.0006</td>
<td>-1.160***</td>
<td>-0.0618</td>
<td>-0.0216</td>
</tr>
<tr>
<td>Wealth €10k–50k</td>
<td>0.0089***</td>
<td>0.214</td>
<td>0.0911</td>
<td>-0.0360**</td>
</tr>
<tr>
<td>Wealth €51k–200k</td>
<td>-0.0027</td>
<td>-0.0966</td>
<td>0.0409</td>
<td>-0.0249**</td>
</tr>
<tr>
<td>Wealth €201k+</td>
<td>-0.0033</td>
<td>-0.0693</td>
<td>0.0453</td>
<td>-0.0317**</td>
</tr>
<tr>
<td>HH financial administrator</td>
<td>-0.0035**</td>
<td>-0.227</td>
<td>0.0611</td>
<td>0.0001</td>
</tr>
<tr>
<td>Financially knowledgeable</td>
<td>-0.0003</td>
<td>-0.420</td>
<td>-0.0380</td>
<td>-0.0137</td>
</tr>
<tr>
<td>Short duration</td>
<td></td>
<td></td>
<td></td>
<td>0.0820***</td>
</tr>
<tr>
<td>Long duration</td>
<td></td>
<td></td>
<td></td>
<td>0.0206</td>
</tr>
<tr>
<td>Hypothetical treatment</td>
<td>0.0009</td>
<td>1.440***</td>
<td>-0.111**</td>
<td>0.0048</td>
</tr>
<tr>
<td>Low incentive treatment (^1)</td>
<td>2.780***</td>
<td>0.845***</td>
<td>1.000</td>
<td>1.060</td>
</tr>
</tbody>
</table>

Notes: Number of observations is 1,422. Estimation follows (7). Regression coefficients are transformed back to the original scale. In other words, the constant is defined by \( g(\beta) \) and represents median parameters in the high-incentive treatment. For the hypothetical treatment, the tables show \( g(\beta + \beta_{\text{hypothetical}}) - g(\beta) \), the (partial) effect of moving from the high incentive to the hypothetical treatment on the median parameter value. The other values are partial effects of setting the dummy variables to one, given the reference value defined by the left-out categories. These categories are: male, age 18–34, primary/lower secondary education, net annual household income below €22,000, total wealth below €10,000, not being the household’s financial administrator, not being financially knowledgeable (self-rated), completion time between 9 and 18 minutes, high-incentive treatment.

\(^1\)The low-incentive treatment enters multiplicatively and we report the nontransformed coefficients, i.e., \( \beta_{\text{low incentive}} \). Coefficient values smaller than one indicate a negative effect and values greater than one a positive effect on the parameter.

association between risk aversion and cognitive skills, which is consistent with our findings for education. Loss aversion is higher among women, and its age pattern peaks between 35 and 44 years of age and then declines. It also declines with
household income; other variables are not significant. Some of the results for loss aversion contradict those of Gächter, Johnson, and Herrmann (2007). Their experimental design, utility specifications, and sample selection procedure are very different, however, so it is difficult to pin down a precise reason for the divergent results. The elderly are on average less averse to late uncertainty resolution than younger age groups.

The largest differences between sociodemographic groups are found for the error parameter. A plausible explanation is differences in numeracy (see James Banks and Zoë Oldfield (2007)). For low-educated persons older than 65 with low income and wealth, we estimate a random choice propensity of 40 percent at the median. Significantly fewer errors are made by the young and highly educated subjects. Point estimates for income and financial literacy are insignificant, but error rates decrease with wealth. We also included dummies for the time used to complete the experiment. They have the expected effect: those who completed the experiment rapidly have higher error rates and those who take a lot of time make fewer errors.

D. Choices and Preferences at the Individual Level

In this section we show how much information the choices in the experiment provide on the subjects’ preference parameters. We do this for the five benchmark individuals described in Section IIB. Our approach is similar to that of David Revelt and Train (2001). The (“posterior”) distribution of preference parameters of each respondent is derived, conditional on observed choices of that respondent and given the estimated (“prior”) distribution of the preference parameters given individual characteristics $X_i$. Conditioning on the subjects’ observed characteristics increases the prior distribution’s chances to provide a good fit to the choices. Panels A–D of Figures 7–11 contain plots of both the prior distribution $F(\eta_i|X_i, \beta, \Sigma, g_\eta)$ and the posterior distribution $F(\eta_i|y_i, X_i, \beta, \Sigma, g_\eta)$. The general picture that emerges is that the posterior estimates of $\gamma_i$ are much more accurate than for the other parameters.

Consider Figure 7 with the posteriors for respondent R1. Her risk aversion parameter is very likely near the mean of the unconditional distribution—the 10 percent and 90 percent quantiles of the marginal posterior distribution imply risk premia for $\pi_1$ of €4.62 and €7.67, respectively.12 The same quantiles for the prior distribution (given the individual characteristics only, and not the observed choices in the experiment) imply risk premia of €−3.14 and €11.59. Hence knowing this subject’s choices helps enormously to determine her risk aversion parameter. For the loss aversion parameter, the distance between the first and ninth deciles shrinks from more than €14 in the prior to less than €6 when conditioning on the subject’s choices. Panel C shows that there is less than a 10 percent chance that individual R1 prefers early to late resolution, while the corresponding prior probability for someone with her socioeconomic characteristics is more than 40 percent. Finally, panel D reveals that her random choice propensity is in the medium range. For all four parameters, conditioning on the choices makes the implied parameter ranges much tighter than conditioning on covariates only.

12 The other parameter values are set to their posterior medians. Tables A40–A43 in the Web Appendix contain the full details, including the risk premia for respondents R1–R4 at various quantiles of their distributions.
As expected from her choices, subject R2’s risk aversion parameter is in the upper range of the distribution. The risk premia for $\pi_1$ implied by the parameters at the posterior first and ninth deciles are €9.37 and €12.52, respectively. The chance that her loss aversion parameter is less than the group median of 4.4 is slightly above 10 percent. The 90 percent quantile is 75, which makes sense since in all gambles with nonzero probability of a negative outcome this individual chose the safe option. The effect of this on the risk premia seems small—those for $\pi_2$ vary between $13.73$ and $14.46$ as $\lambda$ moves from 4.3 to 75—but it should be noted that even with infinite loss aversion, the risk premium will not become larger than 15 (eliminating the potential loss completely). The conditional distribution of $\rho$ almost tracks the population distribution, implying that individual 2’s choices provide very little information on her value of $\rho$. Panel D shows that since her choices are highly consistent, her “trembling hand” error propensity $\omega$ is likely to be low: with 95 percent probability, it is below 7 percent.

Individual R3 (Figure 9) made moderately risk-averse choices if payoffs were nonnegative and resolution of uncertainty was early. For late resolution and potentially negative payoffs, she never made a risky choice. The posterior distributions reflect this choice pattern: her utility curvature parameter is below average with the first and ninth decile at 0.019 and 0.031 and corresponding risk premia for $\pi_1$ of €3.73 and €5.86. The loss aversion parameter is large and more precisely estimated...
than for R2 since some of R3’s choices imply an upper bound on \( \lambda \). The first and ninth deciles are at 6.59 and 28.35. The posterior median of \( \rho \) is 1.8 with the first and ninth decile at 1.34 and 2.46. This implies a preference for early resolution of uncertainty: the risk premium of €4.79 for \( \pi_1 \) in the early resolution case rises to €7.96 if uncertainty resolves late. Even at the first decile of \( \rho \), it is still €6.22.

Figure 10, panel A shows that R4 is risk-loving (\( \gamma < 0 \)). He is very likely not to be loss averse; nor does he prefer early resolution of uncertainty. All risk premia are substantially negative for this subject. He has a low tendency to make suboptimal choices. The posterior distributions of R4’s parameters are quite narrow and directly reflect his choices. On the other extreme, individual R5’s choices are hardly informative about her preferences (Figure 11). The marginal distributions conditioning on her choices are just as dispersed as the unconditional distributions. All we can say is that her behavior is probably driven by a tendency to choose randomly instead of on the basis of the economic model—panel D shows that her \( \omega \) is probably close to one.

### IV. Conclusions

We have described a large-scale experiment on decision making under risk using a representative sample of a broad population. We have analyzed the experimental data using a structural empirical model, disentangling preference parameters for
utility curvature, loss aversion, and preference for the timing of uncertainty resolution, and allowing for several types of errors.

Our model requires a number of specification choices. To save space, we have presented only the results of our preferred specification. In the Web Appendix, we further motivate our specification choices, showing that our model performs better than several alternatives. We also show that our substantive results are generally robust to the specification choices we have made. In particular, running the analysis separately by incentive treatment does not lead to new insights. Not accounting for early or late resolution preferences (setting $\rho$ equal to one) or changing the assumptions on the utility curvature to prospect theory–type preferences does not affect the estimates of the average risk premia for our benchmark lotteries very much. Using power utility rather than an exponential utility function leads to a drop similar to the one observed by Choi et al. (2007).

Our main finding is that risk preferences in the population are very heterogeneous, and only a small part of this heterogeneity can be captured with standard covariates such as age, gender, education, income, and wealth. The structural modelling approach combined with the rich data appears to be a useful tool to handle the heterogeneity. Our four main parameters of interest (the three preference parameters and the tendency to choose purely at random instead of on the basis of utility maximization) are modelled as random coefficients, and we find substantial dispersion in all of them. For example, even though we find that the timing of uncertainty
resolution does not matter much for the median respondent, our estimates imply that there are groups in the population that clearly prefer early resolution and other groups that prefer late resolution. This heterogeneity may well explain the mixed evidence for preferences over temporal lotteries that has been accumulated so far in the literature.

Our structural model is particularly appropriate to analyze the informational content of each subject’s choices for the parameters of that subject. This is shown by comparing the posterior distribution of some subjects given their choices to the prior distribution (given covariates only). We find that the choices are generally very informative about individual preference parameters, except, as expected, in cases where the choice data suggest that the subjects’ choices are probably completely random. Importantly, the model is able to handle intermediate cases and does not require to exclude individuals from the sample a priori. Furthermore, the individual propensity to choose at random in experiments is likely to be informative about the quality of choices in other domains as well.

Finally, the large degree of preference heterogeneity that we have documented implies that controls for individual preferences contain useful information in addition to sociodemographics. Many surveys now contain simple attitudinal questions, which have been shown to correlate with experimental measures similar in spirit to the one used here (Dohmen et al., forthcoming). These measures generally do not

\[ \text{Figure 10. Choices and Preference Parameter Distributions of Respondent R4} \]

\textit{Notes:} Black lines are the estimated parameter distributions for respondent R4’s sociodemographic group (male, age 45–54, primary or lower secondary education, household income above €40,000, wealth between €51,000 and €200,000, not financially knowledgeable, financial administrator, short duration). Gray lines are the marginal distributions of parameters conditional on the choices shown in the first panel. Graphs are based on estimates in Table 5 and the second column of Table 4.
allow an assessment of the precision with which preferences are measured. Our analysis shows such information may be collected at relatively low cost and it would be desirable to make such measures more widely available for empirical work based on micro-data from socioeconomic surveys.

APPENDIX: PREFERENCES TOWARDS THE TIMING OF UNCERTAINTY RESOLUTION

We model the second-period utility function as a slightly modified version of (2):

\[
\nu(z, \gamma, \lambda, \rho) = \begin{cases} 
\max \{ -\frac{\lambda}{\gamma}, 0 \} - \frac{1}{\gamma} e^{-\gamma \rho z} & \text{for } z \geq 0 \\
\max \{ -\frac{\lambda}{\gamma}, 0 \} + \frac{\lambda - 1}{\gamma} - \frac{\lambda}{\gamma} e^{-\gamma \rho z} & \text{for } z < 0,
\end{cases}
\]

where \( S \) is the sign operator given by

\[
S = \begin{cases} 
1 & \text{for } \gamma \geq 0 \\
-1 & \text{for } \gamma < 0.
\end{cases}
\]
We choose the following parsimonious “power function” specification of the weighting function $h$:

$$h(v(z, \cdot)) = -S(-v(z, \cdot))^{\rho - s},$$

where $\rho \in \mathbb{R}_+$. For $\rho > 1$, $h(\cdot)$ is convex and early resolution is preferred to late resolution. Indifference is obtained for $\rho = 1$, and late resolution is preferred for $\rho < 1$.

The building blocks of (8) and (10) seem complicated because of the necessity to accommodate both types of utility curvature. The term $-\lambda/\gamma$ is added for risk lovers to assure that the weighting function $h(\cdot)$ can be applied, i.e., it guarantees that $v(z, \gamma, \lambda, \rho)$ is always greater than zero for $\gamma < 0$. Including $\rho^s$ in the exponent serves to retain the interpretation of $\gamma > 0$ as the coefficient of absolute risk aversion for early resolving lotteries on the positive domain. For such lotteries, $V(\pi)$ collapses to $E[u(\pi)]$ given in (2) if the subject is not risk loving. This implies that the distinction between risk aversion and uncertainty resolution timing preferences is identified for risk-averse subjects if there are gambles on the positive domain and the timing of uncertainty resolution is varied. Note that this distinction is only approximately true for gambles with negative outcomes because of the additive term $(\lambda - 1)/\gamma$ in (8). For risk lovers, the inclusion of $-\lambda/\gamma$ distorts the interpretation of $\gamma$ by the same token. For the parameter values that we estimate, the magnitudes of the distortions are small. The Köbberling and Wakker (2005) definition of loss aversion remains valid for period 2 utility.

REFERENCES


