Who Is (More) Rational?*

Syngjoo Choi, Shachar Kariv, Wieland Müller, and Dan Silverman†

November 2, 2010

Abstract

Revealed preference theory offers a criterion for decision-making quality: if decisions are high quality then there exists a utility function that the choices maximize. We conduct a large-scale field experiment that enables us to test for consistency with utility-maximizing behavior and combine the experimental data with the wide range of individual sociodemographic and economic information for the subjects. There is considerable heterogeneity in subjects' consistency scores: high-income and high-education subjects display greater levels of consistency than low-income and low-education subjects, men are more consistent than women, and young subjects are more consistent than older subjects. We also find that consistency with utility maximization is strongly related to wealth differentials: a standard deviation increase in the consistency score is associated with in 15-19 percent more wealth. This is important for understanding the role of decision-making quality in determining why households with similar

*We thank Douglas Gale and Raymond Fisman for detailed comments and suggestions. We are also grateful to James Andreoni, Andrew Caplin, David Card, Stefano DellaVigna, Guillaume Frechette, Steffen Huck, Patrick Kline, John List, Nicola Persico, Imran Rasul, and Hans-Martin von Gaudecker for helpful discussions. This paper has also benefited from suggestions by the participants of seminars at several universities and conferences. Jelmer Ypma provided excellent research assistance. We thank Corrie Vis and Edwin de Vet of CentERdata (Tilburg University) for software development and technical support. We acknowledge financial support from the Economic and Social Research Council (ESRC) Grant No. RES-061-25-0348 (Choi), the Center on the Economics and Demography of Aging (CEDA) and the Coleman Fung Risk Management Research Center (OpenLink Fund) at UC Berkeley (Kariv), and the Netherlands Organisation for Scientific Research (NWO) and the Network for Studies on Pensions, Aging and Retirement (Netspar) (Müller).

†Choi: University College London (email: syngjoo.choi@ucl.ac.uk); Kariv: University of California–Berkeley (email: kariv@berkeley.edu); Müller: University of Vienna and Tilburg University (email: wieland.mueller@univie.ac.at); Silverman: University of Michigan and NBER (email: dansilv@umich.edu).

economic and demographic characteristics accumulate radically different amounts of wealth (Ameriks et al., 2003).
JEL Classification Numbers: C93, D01, D03, D12, D81
Keywords: Decision-making quality, rationality, revealed preference, experiment, CentERpanel.

1 Introduction

Traditional economic analysis assumes that choices are rational; decision-makers choose their preferred alternative from the feasible set given the information available to them. In this standard view, heterogeneity in choices is attributed to heterogeneity in *preferences*, *information*, *beliefs*, or *constraints*. More recently, several strands of empirical research consider heterogeneity in choices driven, instead, by differences in the *quality* of decision-making. Prominent examples of this research include Ameriks et al. (2003), Bernheim and Garrett (2003), and Agarwal et al. (2009).

Whether we treat individuals as high-quality decision-makers has important consequences. If decision-making skills are poor or the costs of making an optimal decision are high, then there are potentially important wedges between the choices that some decision-makers actually make and the choices they would make if they had the skills or knowledge to make higher quality decisions. These wedges matter because then "revealed" preferences may not be "true" underlying preferences. In that case, positive predictions or welfare conclusions based on the revealed preferences may be misleading.

While the possibility of heterogeneity in decision-making quality has important consequences for economic analysis, definitive judgment about the quality of decision-making is generally made difficult by twin problems of *identification* and *measurement*. The identification problem is to distinguish differences in decision-making quality from unobserved differences in preferences, information, beliefs or constraints. Identification is especially important because welfare conclusions and thus (constrained) optimal policy will depend on the sources of any systematic differences in choices. The measurement problem is to define and implement a *portable*, *practical*, *autonomous*, *quantifiable*, and *economically interpretable* measure of decision-making quality.

In this paper, we measure aspects of decision-making quality by calculating how nearly individual choice behavior in an experiment complies with economic rationality in the sense of a consistent (complete and transitive) preference ordering. This criterion for decision-making quality is not as restrictive as might be thought. It simply requires consistent preferences over all possible alternatives, and choices that correspond to the most preferred alternative from the feasible set. Importantly, any consistent preference ordering is admissible. Thus, if there is no utility function that choices maximize then these choices cannot be considered purposeful and, in this way, high quality.

Classical revealed preference theory tells us that choices are consistent with maximizing a (well-behaved) utility function if and only if they satisfy the Generalized Axiom of Revealed Preference (GARP). Since GARP offers an exact test (either the data satisfy GARP or they do not), a variety of goodness-of-fit indices have been proposed for quantifying the extent of violation. The main index is Afriat's (1972) Critical Cost Efficiency Index (CCEI). By definition, the CCEI is bounded between zero and one. The closer it is to one, the closer the data are to satisfying GARP; and the difference between the CCEI and one can be interpreted as an *upper bound* of the fraction of income that an individual is "wasting" by making inconsistent choices.

We test whether choice behavior in an experimental setting in a broad population is consistent with the utility maximization model using revealed preference axioms. Our points of departure from the literature on decision-making quality, cognitive and so-called non-cognitive skills, and economic outcomes derive from two observations:

[1] Consistency with utility maximization is independent of preference type and the experimental task we study makes no special demands of outside knowledge or expertise. This helps to isolate heterogeneity in decision-making quality from heterogeneity in preferences, information, beliefs or constraints (the identification problem).

[2] The CCEI (and other goodness-of-fit indices) has a coherent economic interpretation and is easily adapted for application in a variety of decision domains. The theoretical framework and portability of the measure are valuable for drawing conclusions that go beyond the particular setting of the experiment (the measurement problem).

Within economics there is a vast amount of work on the rationality of decisions, and *laboratory* experiments have provided key empirical guideposts for developments in this area. To connect the insights that we have gained from the experimental study of rational choice under laboratory conditions to practical questions in the broader world, we conducted a *field* experiment utilizing the CentERpanel, a representative sample of over 2,000 Dutch-speaking households in the Netherlands. The advantage of using the CentERpanel is the wide range of individual sociodemographic and economic information that it provides about the panel members.

By combining our experimental setup's capacity with the CentERpanel, we provide three types of analysis:

[1] We begin our analysis with a purely descriptive overview of some important features of the data, concerning the average quality of decisions.

[2] We then move to a regression analysis of the relationship between decision-making quality and demographic and economic characteristics. In this way we address the question: "who is (more) rational?"

[3] Finally, we connect the insights that we gain from the experimental study under laboratory conditions to practical questions concerning wealth differentials in the real world.

In our experiment, we present subjects with a sequence of standard consumer decision problems that can be interpreted either as the selection of a bundle of commodities from a standard budget line or the allocation of individual wealth between *risky* assets. These decision problems are presented using a graphical interface introduced by Choi et al. (2007a) and used by Choi et al. (2007b). Because the design is user-friendly, it is possible to present each subject with *many* choices in the course of a short experimental session, yielding a much larger data set than has been possible in the past. This allows us to analyze the data at the level of the individual subject rather than pooling data or assuming homogeneity across subjects. Because choices are from standard budget lines, we can use revealed preference tests to investigate the extent to which the data comply with utility maximization. Since we observe many choices over a wide range of budget lines, the data allow high-power tests of revealed preference conditions.

If we follow Varian's (1991) suggestion of a threshold of 0.95 for the CCEI, we find that 45.2 percent of our subjects' scores are above this threshold, and of those, 22.8 percent have no violations of GARP. To calibrate the CCEI scores we compare the behavior of our actual subjects to the behavior of simulated subjects whose payoffs are perturbed by small idiosyncratic preference shocks. We conclude that the scores of many subjects essentially satisfy GARP in the sense that their violations are small enough to be attributed to the effect of a "trembling hand." Nevertheless, over all subjects, the CCEI scores averaged 0.881, which implies that subjects on average waste as much as 12.0 percent of their earnings by making inefficient choices. There is also marked heterogeneity in the CCEI scores within and across the demographic characteristics of our subjects. Figure 1 summarizes the mean CCEI scores and 95 percent confidence intervals across standard socioeconomic categories. Alternative measures of GARP violations based on Varian (1990, 1991) and Houtman and Maks (1985) (HM) yield qualitatively similar conclusions.

[Figure 1 here]

We next move to studying, more systematically, the correlations between goodness-of-fit indices and demographic and economic characteristics. Our data are particularly well suited to such investigations, given the heterogeneity in our experimental outcomes and the heterogeneity in our subject pool. Using Heckman's (1979) sample-selection model, we control for the possibility of sample-selection bias. Since the recruitment of CentERpanel members to experiments is random by construction, our instrumental variable is the number of completed CentERpanel questionnaires as a fraction of the total invitations to participate in the three months preceding our experiment.

Our main findings are that high-income and high-education subjects display greater levels of consistency than lower-income and lower-education subjects. Additionally, men are more consistent than women, and young subjects lean more toward utility maximization than those who are old. The magnitudes are large, implying, for example, that low-income subjects on average waste as much as 3.3 percentage points more of their earnings by making inefficient choices relative to high-income subjects. The corresponding numbers for low-education subjects, females, and old subjects are 2.6, 2.4, and 5.1, respectively.

The most basic question to ask about choice data is whether it is consistent with individual utility maximization. Beyond consistency, the next question to ask is whether choices are consistent with a utility function with some normatively appealing structural properties. In decision-making under uncertainty, it is of interest to determine whether choices are also consistent with the *dominance principle* in the sense of Hadar and Russell (1969)–that is, the requirement that an allocation should be preferred to another, regardless of subjects' risk attitudes, if it yields unambiguously higher monetary payoff. The dominance principle is compelling and generally accepted in decision theory.[1] Overall, the choices made by subjects in our experiment show

[1] The dominance principle is compelling and generally accepted in decision theory. As

low rates of stochastic dominance violations, which decrease with education level and increase with age.

Finally, we examine whether consistency scores are useful in explaining wealth differentials. As Ameriks et al. (2003) emphasize, appropriately confronting the pattern of differential wealth outcomes with empirical or experimental evidence will have implications in many areas of economic theory and policy. Furthermore, virtually every realm of individual decision-making enters wealth accumulation, and wealth is likely a critical input into most any measure of economic well-being.

We find an economically large and statistically significant association between the CCEI and household wealth. The point estimates indicate that, conditional on measures of permanent and current income, and household structure, a standard deviation increase in the CCEI of the person who is primarily responsible for household financial matters is associated with 15-19 percent more household wealth. This result is little changed when we add controls for violations of the dominance principle or a summary measure of risk tolerance. The point estimates indicate that the latter two measures are related to wealth in anticipated ways, but neither relationship is statistically distinguishable from zero. We find no evidence that the CCEI is capturing unobserved aspects of education. The measures suggested by Varian (1990, 1991) and HM yield similar conclusions.

There are many important questions that remain to be explored using this data set. In work-in-progress, we use the same data to relate findings on individual-level risk attitudes from the experimental data with economic information and sociodemographic information on individuals. More specifically, the analysis of the experimental and field data analysis consists of a combination of structural and descriptive work and provides [1] corroboration of the earlier work of Choi at al. (2007b) on *risk* and *loss* aversion; [2] a new set of estimates of models, in which data on income and wealth are used to identify the coefficient of risk aversion; [3] an attempt to explain heterogeneity in preferences and in types of behaviors in terms of demographic variables; and [4] an investigation of the correspondence between individual investment and savings decisions and behavior in the laboratory. This will enhance our understanding of important economic decisions such as savings and portfolio allocation but it distracts from our fundamental purpose in this paper.

The rest of the paper is organized as follows. Section 2 further discusses

noted by Quiggin (1990) and Wakker (1993), theories of choice under uncertainty were amended to avoid violations of dominance.

the notion of decision-making quality. Section 3 describes the experimental design and procedures. Section 3 organizes the experimental data. Section 4 contains the analysis of the relationship between decision-making quality and demographic and economic characteristics. Section 5 discusses the correspondence between decision-making qualities and wealth differentials. Section 6 discusses the results and describes the margins along which we extend the previous literature. Section 7 contains some concluding remarks.

2 Decision-making quality

As noted above, standard economic analysis attributes heterogeneity in choices to heterogeneity in preferences, information, beliefs, or constraints. A relatively new empirical literature now considers heterogeneity in choices driven, instead, by differences in the quality of decision-making. This literature allows that the choices that some decision-makers actually make may be different from the choices they would make if they had the skills or knowledge to make better decisions.

Ameriks et al. (2003) is a prominent example. That paper provides evidence that differences in individuals' propensity to plan and budgeting behaviors, rather than more standard sources of heterogeneity, explain important variation in wealth accumulation. In another example, Bernheim and Garrett (2003) find evidence that employer-based financial education increases saving. Agarwal et al. (2009) show a U-shaped age pattern in the frequency of dominated choices regarding the use of credit, with both younger and older consumers more prone to error.[2]

The possibility that planning skills, financial education, experience or cognitive decline substantially affect the quality of decision making is important because it suggests there are circumstances when "revealed" pref-

[2] Restricting attention just to the quality of *financial* decision-making, this literature also includes, among others, Duflo and Saez (2003) who investigate effect of financial education on saving, beyond its effect on lifetime earnings; Lusardi and Mitchell (2007) document very low levels of financial planning, financial literacy, and a positive correlation between literacy, financial planning and wealth; and Cole and Shastry (2009) emphasize the importance of education, cognitive ability and financial literacy on financial market participation. Most recently, published in the same issue of the *Economic Journal*, Banks (2010) summarizes the research on the relationships between cognitive function, financial literacy and financial outcomes at older ages; Smith et al. (2010) and Banks et al. (2010) show that wealth and retirement saving patterns are associated with numerical and other cognitive abilities at middle and older ages; and Van Den Berg et al. (2010) and Jappelli (2010) explore some of the potential causes of the differences in cognitive function and financial literacy in later life.

erences may not be "true" preferences. If so, then positive predictions and welfare conclusions based on the "revealed" preferences may be misleading.

Through the collection of uncommonly high quality data, or the exploitation of natural experiments or instrumental variables, the research in this new literature has provided convincing evidence of important differences in decision-making quality. In general, however, definitive judgement about decision-making quality is made difficult by twin problems of *identification* and *measurement*.

> **Identification** The identification problem is to distinguish differences in decision-making quality from unobserved differences in preferences, information, beliefs or constraints.

In observational data, it is usually unclear whether those with less (financial) education, or lower cognitive abilities, fewer planning skills or less financial literacy are making lower quality decisions as opposed to holding different beliefs, or having different preferences over the same outcomes, or facing different incentives and constraints. The distinction has important policy consequences because positive predictions and welfare conclusions depend critically on the sources of any systematic differences in choices between these groups.

Moreover, the identification problem presupposes a measurable notion of decision-making quality. In some cases the relevant incentives are sufficiently clear and data quality is sufficiently high, so that regarding some decisions as higher quality is straightforward and uncontroversial. More generally, a measure of decision making quality is difficult to formalize, quantify and to make practical and portable for use in a variety of choice environments. These features of a measure are especially important to the extent that decision-making quality is a trait–a characteristic of a person that affects decisions in many different contexts.

> **Measurement** The measurement problem is to define and implement a portable, practical, autonomous, quantifiable, and economically interpretable measure of decision-making quality.

In this paper, we measure aspects of decision-making quality by their compliance with economic rationality. In his Foundations of Economic Analysis (1947), Paul Samuelson offered a natural criterion for decision-making quality based solely on observable behavior. Adopting Samuelson's approach, we will say that choices are lower quality if there is no well-defined (utility) function that the choices maximize. Classical revealed preference theory provides a direct test: choices are consistent with maximizing a utility function if and only if they satisfy the GARP. Since GARP imposes the complete set of conditions implied by utility-maximization, the CCEI and other goodness-of-fit indices provide a stringent test of decision-making quality.

The primary methodological contribution of this work is an experimental technique that allows for the collection of richer data about preferences than has previously been possible and can easily be adapted to a wide range of decision-making experiments in large-scale surveys. As a result, the entire apparatus developed here – analytical and experimental techniques – has a number of distinctive features that make it useful for evaluating the quality of economic decision-making:

- **Portable** The analytical techniques and experimental platforms are applicable to many other types of individual choice problems involving personal and social consumption. They can thus make domain-specific predictions and provide a *unified* measure of decision-making quality across domains.

- **Practical** In the real-world, the changes in income and relative prices are such that budget lines do not cross frequently. This means that market data lack power to test revealed preference conditions (Blundell et al., 2003). Our subjects can be given a large and rich menu of budget sets which leads to high power tests.

- **Autonomous** Consistency with GARP is independent of preference type and the experimental task we study makes no special demands of outside knowledge or expertise, thus helping to isolate heterogeneity in decision-making quality from heterogeneity in preferences, information, beliefs or constraints (the identification problem).

- **Quantifiable** The CCEI (and other goodness-of-fit indices) measures the *extent* of GARP violation. In contrast with hypothetical (and unincentivized) survey data, we can understand the results in terms of economic theory, which help us interpreting (as well as designing) the experiments in several ways.

- **Interpretable** The CCEI has a coherent economic interpretation. Because of our rich data set, we are able to generate fairly tight bounds on the CCEI and use these bounds to judge the welfare effects of decision-making quality.

A last feature of the apparatus developed here is that the method and measure may be evaluated for their ability to predict important behavior in the real world. In this paper we consider, for example, whether the CCEI as measured in the experiment can help explain a persistent puzzle about financial choices: the wealth differentials among households with similar lifetime income paths. Existing research on decision-making quality has largely focused on financial choices. For purposes of evaluating measures of financial decision-making quality, wealth differentials (given income) is a natural object of study. Conditional on income, wealth is the result of innumerable financial decisions (saving rates, investment and insurance portfolios, budgets, product choices, and more) all of which may vary in quality of decision-making.

In addition, Bernheim et al. (2001) and Ameriks et al. (2003) show that heterogeneity in wealth is not well-explained either by standard observables such as income, education or family structure, or by standard unobservables such as intertemporal substitution or risk tolerance. These papers suggest that departures from standard models, or departures from strict rationality, may help to explain the massive disparities in wealth. If consistency with individual utility maximization in the experiment were a good proxy for financial decision-making quality then the degree to which consistency differ across subjects should help explain differential patterns of wealth. We take up this question in section 5.

3 Experimental design

3.1 Sample

The experiment uses the CentERpanel, an online, weekly, stratified survey of a sample of over 2,000 households and 5,000 individual members. The panel is designed to be representative of the Dutch-speaking population in the Netherlands. The CentERpanel thus provides a unique opportunity to combine experimental data with demographic and economic variables from the survey. The subjects in the experiment were randomly recruited from the entire CentERpanel body. The experiment was conducted *online* under the CentERdata protocol with 1,182 CentERpanel adult members, using the experimental technique introduced by Choi et al. (2007a) and used by Choi et al. (2007b).[3] The experimental methodology allows us

[3] CentERdata is an independent research institute affliated with the Tilburg School of Economics and Management (TiSEM) at Tilburg University in the Netherlands. The

to identify individual behaviors that may be related to a wide range of individual characteristics.

Table 1 below provides summary statistics of individual level characteristics. We present the data for *participants* (completed the experiment), *dropouts* (logged in and quit the experiment) and *nonparticipants* (recruited for the experiment but did not log in). We use six standard socioeconomic categories: gender, age, education, income, occupation, and household composition. The groupings of different levels of education are based on the categorization of Statistics Netherlands (Centraal Bureau voor de Statistiek). The low, medium and high education levels correspond to primary education or lower secondary education, secondary education, and higher education, respectively. We use household monthly gross income-level categories such that the proportions of participants in each category are approximately equal. The classification of occupations is based on the categorization of Statistics Netherlands.

[Table 1 here]

3.2 Procedures

Our experimental interface was incorporated into the CentERpanel and the experiment was hosted as part of their survey. In our experiment, we presented subjects with several decision problems under uncertainty. Each decision problem was presented as a choice from a two-dimensional budget line. A choice of the allocation (x, y) from the budget line represents an allocation between accounts x and y (corresponding to the horizontal and vertical axes). The actual payoffs of a particular choice were determined by the allocation to the x_1 and x_2 accounts such that the subject received the points allocated to one of the accounts x or y, determined at random and equally likely. Choices were made by using the computer mouse or the keyboard arrows to move the pointer on the computer screen to the desired point and then clicking or hitting the enter key.[4] The procedures described

CentERdata specializes in online experiments and manages the CentERpanel and several other panels. The panel members complete the questionnaires on the Internet from home. For more information, see http://www.centerdata.nl/en/centerpanel.

[4] Ahn et al. (2010) extended the work of Choi et al. (2007b) on risk (known probabilities) to settings with ambiguity (unknown probabilities). Fisman et al. (2007, 2010) employ a similar experimental methodology to study distributional preferences and produce very different behaviors. It is of course possible that presenting choice problems graphically biases choice behavior in some particular way—and that is a useful topic for experiment—but there is no evidence that this is the case, as emphasized by Choi et al. (2007b) and Fisman et al. (2007).

below are *identical* to those used by Choi et al. (2007b), with the exception that the experiment described here consisted of 25, rather than 50, decision problems and that there were some minor additional changes resulting from the online experimental setting.[5] Payoffs were calculated in terms of points and then converted into euros. Each point was worth €0.25. Subjects received their payment via the CentERpanel reimbursement system.

Each decision problem started by having the computer select a budget line randomly from the set of budget lines that intersect with at least one of the axes at 50 or more points, but with no intercept exceeding 100 points. This variation in budget lines (prices and incomes) is essential for a thorough test of consistency. The budget lines selected for each subject in different decision problems were independent of each other and of the sets selected for any of the other subjects in their decision problems. Choices were restricted to allocations on the budget constraint, so that subjects could not violate budget balance. During the course of the experiment, subjects were not provided with any information about the account that had been selected in each round. As in Choi et al. (2007b), at the end of the experiment, the computer selected one decision round for each subject, where each round had an equal probability of being chosen, and the subject was paid the amount he had earned in that round.

The resolution compatibility of the budget lines was 0.2 tokens. At the beginning of each decision round, the experimental program dialog window went blank and the entire setup reappeared. The appearance and behavior of the pointer were set to the Windows mouse default and the pointer was automatically repositioned randomly on the budget line at the beginning of each round. We refer the interested reader to Choi et al. (2007a) for an extended description of the experimental interface. Full experimental instructions, including the computer program dialog windows are also available at Online Appendix I.[6]

4 Data description

We next provide an overview of some basic features of the individual-level data. Without essential loss of generality, assume the individual's income is

[5] The number of individual decisions is still higher than it usually is in the literature, and the experiments provide us with a rich data set consisting of enough individual decisions over a wide range of budget lines to provide a powerful test of consistency. The revealed preference analysis presented below shows that the variation in budget lines (prices and incomes) is sufficient for a rigorous test of consistency.

[6] Online Appendix I: (http://emlab.berkeley.edu/~kariv/CKMS_I_A1.pdf).

normalized to 1. The budget set is then $p_1x_1 + p_2x_2 = 1$ and the individual can choose any allocation x that satisfies this constraint. Let $\{(p^i, x^i)\}_{i=1}^{25}$ be the data generated by some individual's choices, where p^i denotes the i-th observation of the price vector and x^i denotes the associated allocation.[7]

4.1 Consistency

Following Afriat's (1967) theorem, we employ the Generalized Axiom of Revealed Preference (GARP) to test whether the finite set of observed price and quantity data that our experiment generated may be rationalized by a utility function $U(x_1, x_2)$. GARP (which is a generalization of various other revealed preference tests) requires that if x^i is indirectly revealed preferred to x^j, then x^j is not *strictly* directly revealed preferred ($p^j x^i \geq p^j x^j$) to x^i. The theory tells us that if the data satisfy GARP, then a utility function that rationalizes the observed allocations exists and, moreover, may be chosen to be *well-behaved* (piecewise linear, continuous, increasing, and concave).[8]

Although testing conformity with GARP is conceptually straightforward, there is an obvious difficulty: GARP provides an exact test of utility maximization – either the data satisfy GARP or they do not – but individual choices may involve errors. Subjects may compute incorrectly, or execute intended choices incorrectly, or err in other less obvious ways. To account for the possibility of errors, we assess how nearly individual choice behavior complies with GARP by using Afriat's (1972) CCEI, which measures the fraction by which each budget constraint must be shifted in order to remove *all* violations of GARP. If the CCEI is close to one, the subject is wasting very little of his earnings. Otherwise, he may be wasting quite a lot. In this sense the CCEI measures the overall "efficiency" of individual behavior.

Put precisely, for any number $0 \leq e \leq 1$, define the direct revealed preference relation

$$x^i R^D(e) x^j \Leftrightarrow e p^i \cdot x^i \geq p^i \cdot x^j,$$

and define $R(e)$ to be the transitive closure of $R^D(e)$. Let e^* be the largest value of e such that the relation $R(e)$ satisfies GARP. The CCEI is the value of e^* associated with the data set $\{(p^i, x^i)\}_{i=1}^{25}$. By definition, the CCEI

[7] More precisely, the data generated by an individual's choices are $\{(\bar{x}_1^i, \bar{x}_2^i, x_1^i, x_2^i)\}_{i=1}^{25}$, where (x_1^i, x_2^i) are the coordinates of the choice made by the subject and $(\bar{x}_1^i, \bar{x}_2^i)$ are the endpoints of the budget line, so we can calculate the budget line $x_1^i/\bar{x}_1^i + x_2^i/\bar{x}_2^i = 1$ for each observation i.

[8] Varian (1982, 1983) modifies Afriat's (1967) results and describes efficient and general techniques for testing the extent to which choices satisfy GARP.

is between zero and one—indices closer to one mean the data are closer to perfect consistency with GARP and hence to perfect consistency with utility maximization—and can be interpreted as saying that the individual is wasting as much as $1 - e^*$ of the income by making inefficient choices. Hence, the CCEI may overstate the extent of inefficiency, but the above procedure is the "least costly" adjustment for removing all violations of GARP.

We provide more details on testing for consistency with GARP and discuss the alternative indices that have been proposed by Varian (1990, 1991) and HM in online Appendix II.[9] In reporting our results, we focus on the CCEI, which offers a straightforward interpretation. In practice, all these measures yield similar conclusions. The tables based on the indices proposed by Varian (1990, 1991) and HM are presented in Online Appendix III.[10,11]

Table 2 below provides a population-level summary of the individual-level CCEI scores. We report the statistics for all subjects, as well as the statistics by socioeconomic categories. The CCEI scores averaged 0.881 over all subjects, and ranged from 0.920 for subjects younger than 35 to 0.843 for subjects age 65 and older. There is also considerable heterogeneity within and across categories. The analysis of the relationship between the differences in consistency scores and demographic differences among experimental subjects is the purpose of our econometric estimation below.[12]

[Table 2 here]

4.2 Power and goodness-of-fit

Revealed preference tests have an important drawback: there is no natural threshold for determining whether subjects are close enough to satisfying GARP that they can be considered utility maximizers. Varian (1991) suggests a threshold of 0.95 for the CCEI, but this is purely subjective. If we follow Varian's (1991) suggestion, we find that out of the 1,182 subjects, 534

[9] Online Appendix II: (http://emlab.berkeley.edu/~kariv/CKMS_I_A2.pdf).

[10] Online Appendix III: (http://emlab.berkeley.edu/~kariv/CKMS_I_A3.pdf).

[11] All indices are computationally intensive for even moderately large data sets. We compute the Houtman-Maks scores using the algorithm developed by Dean and Martin (2010). (The computer program and details of the algorithms are available from the authors upon request.)

[12] To allow for small trembles resulting from the slight imprecision of subjects' handling of the mouse, our consistency results allow for a narrow confidence interval of one token (that is, for any i and $j \neq i$, if $\left|x^i - x^j\right| \leq 1$ then x^i and x^j are treated as the same portfolio).

subjects (45.2 percent) have CCEI scores above this threshold and of those 269 subjects (22.8 percent) have no violations of GARP.[13]

To generate a benchmark against which to compare our CCEI scores, we use the test designed by Bronars (1987), which builds on Becker (1962) and employs the choices of a hypothetical subject who chooses randomly among all allocations on each budget line as a point of comparison. The mean CCEI score across all subjects in our experiment is 0.881 whereas the mean CCEI score for a random sample of 25,000 simulated subjects is only 0.659. Moreover, more than half of actual subjects have CCEI's above 0.925, while only about five percent of simulated subjects have CCEI's that high.

The Bronars' (1987) test has often been applied to experimental data so using it situates the paper in the literature (more below). The setup used in this study has the highest Bronar power of one (all random subjects had violations). Our results show that the experiment is sufficiently powerful to exclude the possibility that consistency is the accidental result of random behavior. Therefore, the consistency of our subjects' behavior under these conditions is not accidental. To provide a more informative metric of the consistency of choices, we follow Choi et al. (2007a) who extend and generalize the Bronars (1987) test. In the interests of brevity, the analysis has been relegated to Online Appendix II.[14]

4.3 Beyond consistency

The theory tells us that if the data satisfy GARP, then a utility function that rationalizes the observed choices exists and, moreover, may be chosen to be well-behaved. Nevertheless, choices can be consistent with GARP but not maximize a utility function that is normatively reasonable for the decision problem at hand. For example, consider choices that always allocate all tokens to x_1. This behavior is consistent with maximizing the utility function $U(x_1, x_2) = x_1$. The broad range of randomly generated budget lines that our experiment involves means that this choice behavior frequently results in allocating all tokens to the more expensive asset, which violates monotonicity with respect to first-order stochastic dominance.

[13] By comparison, Choi et al. (2007b) report that 60 of their 93 subjects (64.5 percent) had CCEI scores above the 0.95 threshold, and of those 16 subjects (17.2 percent) did not violate GARP. The subjects of Choi et al. (2007b) were recruited from undergraduate classes and staff at UC Berkeley. They were given a larger menu of 50 budget lines which provides a more stringent revealed preference test (more below).

[14] Andreoni and Harbaugh (2006) develop power indices for revealed preference tests based on CCEI and discuss the prior indices of Bronars (1987) and Famulari (1995).

Violations of first-order stochastic dominance may reasonably be regarded as errors, regardless of risk attitudes—that is, as a failure to recognize that some allocations yield payoff distributions with unambiguously lower returns. A simple violation of dominance is illustrated in Figure 2 below. The budget line is defined by the straight line AE and the axes measure the future value of a possible allocation in each of the two states. The point B, which lies on the 45 degree line, corresponds to an allocation with a certain outcome. The individual chooses allocation x (position along AB), but could have chosen any allocation x' (position along CD) such that $F_{x'} \leq F_x$ where $F_{x'}$ and F_x are the resulting payoff distributions. If this individual only cares about the distribution of monetary payoffs, then he will be willing to pay a positive price for a lottery yielding $F_{x'} - F_x$, which has only nonpositive payoffs (that is, for a lottery in which each asset had an equal probability of being chosen). Notice that any decision to allocate *fewer* tokens to the *cheaper* asset (that is, corresponding to a position along AB) violates dominance but need not involve a violation of GARP, whereas any decision to allocate *more* tokens to the *cheaper* asset (that is, corresponding to a position along BE) never violates dominance.

[Figure 2 here]

If subjects identify an allocation with the resulting probability distribution over payoffs then preferences satisfy the *reduction principle*; that is, $(x_1, x_2) \sim (x_2, x_1)$ because they generate the same payoff distribution. If preferences satisfy the reduction principle then the choice subject to every budget constraint allocates more tokens to the cheaper asset. We would like to test this decomposition by observing choices from linear budget sets. Unfortunately, this is not possible: choices from linear budget sets determine the demand function but the demand function does *not* uniquely determine preferences (Mas-Colell, 1977; 1978). However, *symmetry* provides implications about choices from linear budget sets (that is, about demand functions) that are testable on the basis of observed choices from standard budget sets.

We identify choice behavior as *symmetric* if (x_1^*, x_2^*) is chosen subject to the budget constraint $p_1x_1 + p_2x_2 = 1$ if and only if (x_2^*, x_1^*) is chosen subject to the *mirror-image* budget constraint $p_2x_1 + p_1x_2 = 1$. That is, choice behavior responds symmetrically to inverse price ratios.[15] Clearly, if choice behavior is symmetric then the choice subject to every budget constraint

[15] The reduction principle implies that choice behavior is symmetric only when the derived demand function is single valued. GARP is also compatible with multi-valued demand functions so preferences may not be strictly convex.

allocates more tokens to the cheaper asset. Hence, symmetry imposes restrictive (if convenient) patterns on demand behavior, but it is a natural result of symmetric probabilities (each account had an equal probability of being chosen).

To test whether choice behavior is symmetric (for a given subject), we can combine the actual data from the experiment and the mirror-image data, compute the CCEI for this combined data set, and compare that number to the CCEI for the actual data.[16] By definition, the CCEI for the combined data set consisting of 50 observations can be no bigger than the CCEI for the actual data. Clearly, always allocating all tokens to one of the assets generates severe violations of GARP in the combined data set, but the subset of actual data is perfectly consistent.[17] Similarly, any decision to allocate fewer tokens to the cheaper asset will necessarily generate a simple violation of the Weak Axiom of Revealed Preference (WARP) involving its mirror-image decision.

Thus, we can construct a formal non-parametric test of symmetric behavior by following the strategy above: compute the CCEI for the combined data set and compare that number to the CCEI for the actual data set. The difference reflects an upper bound on the additional income that the subject is wasting by not always allocating more tokens to the cheaper asset. Nevertheless, the combined data set obviously provides a more stringent test of GARP so it can contain new violations of GARP even if actual choices always allocated more tokens to the cheaper asset.

If we again follow Varian's (1991) suggestion of a threshold of 0.95 for the CCEI, we find that in the combined data set the scores of 251 subjects (21.2 percent) are above this threshold and of those only 24 (2.0 percent) have no violations of GARP. Table 3 below reports summary statistics and percentile values of the CCEI scores for the combined data set. We report the statistics for all subjects, as well as the statistics by socioeconomic categories. The last column lists the difference between the mean CCEI's for the actual data set and for the combined data set. The CCEI scores for the combined data set averaged only 0.733 over all subjects, and ranged from 0.786 for subjects younger than 35 to 0.679 for subjects age 65 and older, representing a decrease from the CCEI scores for the actual data set of 0.148, 0.134 and

[16] The data generated by an individual's choices are $\left\{\left(\bar{x}_1^i, \bar{x}_2^i, x_1^i, x_2^i\right)\right\}_{i=1}^{25}$ and the mirror-image data are obtained by reversing the prices and the associated allocation for each observation $\left\{\left(\bar{x}_2^i, \bar{x}_1^i, x_2^i, x_1^i\right)\right\}_{i=1}^{25}$.

[17] Of the 1,182 subjects in the experiment, only 29 subjects (2.5 percent) almost always allocated all tokens to the one of the assets by choosing the same endpoint of the budget line.

0.165, respectively. Overall, a sociodemographic category that had a lower mean actual CCEI score exhibits a larger decrease. In our econometric analysis below, we use both the CCEI scores for the actual data set and for the combined data set.

[Table 3 here]

4.4 Risk attitudes

We summarize attitudes toward risk by a single univariate measure, which we will use as a measure of risk aversion in the regression analysis concerning wealth differentials.[18] Because the experiment is symmetric and budget lines are drawn from a symmetric distribution, we summarize the risk aversion of our subjects by reporting the fraction of tokens allocated to the cheaper asset. The only behavior consistent with infinite risk aversion is always allocating the tokens equally between the two assets. On the other hand, always allocating all tokens to the cheaper asset is the behavior that would be implied by pure risk neutrality. In general, subjects less averse to risk will allocate a larger fraction of tokens to the cheaper asset. Figure 3 summarizes the mean fraction of tokens allocated to the cheaper asset and 95 percent confidence intervals across the socioeconomic categories. Note that there is considerable heterogeneity in risk attitudes across categories, which is characteristic of all these data, and that risk attitudes and CCEI scores are modestly correlated ($r^2 = 0.113$).

[Figure 3 here]

5 Decision-making quality and sociodemographics

The relatively large and heterogeneous CentERpanel sample and accompanying survey data allow us to perform what is, to our knowledge, the first analysis of the correlation between demographic and economic characteristics and GARP violations. Table 4 below presents the results of our individual-level econometric analysis. In column (1), we present estimates with the CCEI scores for the actual data set using ordinary least squares

[18] In work-in-progress, we build on Choi et al. (2007b) to estimate preferences using a two-parameter utility function based on Gul (1991)—one parameter is the familiar coefficient of risk aversion and the other is a measure of loss/disappointment aversion—and we relate the individual-level estimates to individual characteristics and external choices.

(OLS).[19] The results show significant correlations. We obtain statistically significant coefficients in all demographic categories, ranging in absolute values from about 0.025 to just over 0.050. These magnitudes are large, implying that demographic differences can account for significant differential changes in income loss due to inconsistent choice patterns. Most notably, females, low-education, low-income, and old subjects on average waste as much as 2.4, 2.6, 3.3, and 5.1 percentage points more of their earnings, respectively, by making inefficient choices.[20] In columns (2) we repeat the estimation reported in columns (1) using the CCEI scores for the combined data set.[21] As expected, the corresponding estimates are of higher magnitude and statistically significant in the age and education categories.

[Table 4 here]

Our analysis above is based on the nonrandomly selected subsample of participants. The lack of observations on panel members who chose not to participate or did not complete the experiment creates a missing data problem. We correct for the possible sample selection bias in our econometric analysis below, using Heckman's (1979) method.[22] Our *exclusion restriction* rests on the number of completed CentERpanel questionnaires as a fraction of the total invitations to participate in the three months prior to our experiment enters the participation equation but not rationality. Our identifying assumption is that this "participation ratio" influences the participation in our experiment but does not influence the laboratory outcomes of interest (Bellemare et al., 2008).

The estimation results are reported in Table 5 below. In column (1), we omit the nonparticipants, focusing on the subsample of participants and dropouts in the data. In column (2), we repeat the estimation reported in

[19] To test for a potential misspecification, we used Ramsey's (1969) RESET test by adding the squared and cubed fitted values of the regression equation as additional regressors, and found no evidence of misspecification (p-value = 0.3098).

[20] Agarwal et al. (2009) document a U-shaped relationship between age and mistakes in financial decision making, suggesting that although cognitive abilities decline with age, experience in financial markets rises with it. We find that consistency with GARP and hence consistency with utility maximization decline dramatically over the lifecycle.

[21] Since the CCEI is a number between zero and one, we repeat the estimations reported in columns (1) and (2) using a fractional regression model (Papke and Wooldridge, 1996). The two specifications yield similar results.

[22] We also use Heckman's sample selection model to analyze the correlates of the Varian (1990, 1991) measure. For the third measure, proposed by HM, we estimate the sample selection model of Terza (1998). These results are provided in Online Appendix III.

column (1), after adding the nonparticipants. We obtain qualitatively similar results on the reduced sample and the entire sample. Finally, testing the null hypothesis that the correlation coefficient ρ is zero is equivalent to testing for sample selection. In columns (1) and (2), we find that ρ is indistinguishable from zero and thus we find no evidence of bias. We interpret these results to indicate that self-selection is not importantly driving the results. It is also noteworthy that in both specifications the coefficient on the exclusion restriction variable is positive and significant, and that many demographic categories are positively correlated with participation. In columns (3) and (4), we repeat the estimation reported in columns (1) and (2) using the CCEI scores for the combined data set and obtain similar results.

[Table 5 here]

6 Wealth differentials and decision-making quality

6.1 Wealth data

The CentERpanel collects information about wealth on an annual basis. Panel members are asked to identify a financial respondent in the household who "is responsible for paying bills, etc." All members of the household age 16 and older respond to a series of standard questions regarding assets and liabilities they hold alone. The financial respondent also provides information about assets and liabilities that are jointly held by more than one member of the household.[23] Our analysis here focuses on household net worth, calculated simply by summing net worth over household members, as averaged over years 2008 and 2009. Summary statistics of this measure are displayed in Table 8 below. We have 703 households with valid wealth data and one or more CCEI scores for household members. The median household has a net worth of nearly €93,000 ($136,000). As is typical of data on wealth, the distribution is positive-skewed. Mean household wealth (€164,130) is much higher than the median and the highest values are so large (maximum of 15.7 standard deviations above the mean) that they seem likely to reflect reporting or coding errors.

[Table 6 here]

[23] The inventory covers checking and saving accounts, stock, bond and other financial asset holdings, real estate, business assets, mortgages, loans, and extended lines of credit. For a complete description see http://www.centerdata.nl/en/centerpanel.

6.2 Wealth regressions

There are large wealth differentials among households with similar life-time income paths. Bernheim et al. (2001) and Ameriks et al. (2003) show that these differentials are not well-explained either by standard observables such as income, education or family structure, or by standard unobservables such as intertemporal substitution or risk tolerance. To describe the relationship between decision-making quality and wealth differentials, we estimate regressions of the natural log of household wealth on demographic variables, the natural log of household income, and the CCEI score of the financial respondent in the household.

[Table 7 here]

The estimation results are reported in Table 7 above. In column (1), we present estimates from the entire sample, with no restrictions on age.[24] The point estimate of 1.17 on the CCEI indicates that a standard deviation increase in CCEI score is associated with 15.8 percent more household wealth. As one might expect from a relatively small sample of data on self-reported wealth, the standard error on this point estimate is fairly large. Nevertheless, we can reject a null hypothesis of no relationship at the 5 percent level (p-value=0.029 with standard errors robust to heteroskedasticity).

Standard models often predict that higher net worth would be associated with greater well-being only beyond a certain age. At younger ages, those with better lifetime opportunities may have lower net worth as they borrow in order to invest or to smooth lifetime consumption. With that in mind, in column (2), we repeat the estimation reported in columns (1) with the sample restricted to households with financial respondents who are at least 35 years old. We find that the point estimate on the CCEI is somewhat larger in older ages so a standard deviation increase in CCEI score is associated with 19.2 percent more household wealth. The standard error on this point estimate is fairly large, so while we can reject a null hypothesis of no relationship with considerable confidence (p-value=0.012) we cannot reject a null hypothesis that the point estimates on the CCEI reported in

[24] The sample size drops from 703 to 566 household (80.5 percent). This decline is driven almost exclusively by 74 households (10.5 percent) with negative net worth and thus a missing dependent variable and 54 households with negative household income in 2008 (7.7 percent). In addition, given our relatively small sample and the presence of extreme outliers, we also drop 7 households that represent the union of the top and bottom half a percent of the wealth distribution and the bottom half a percent of the distribution of CCEI scores. Two additional households are dropped due to missing data on education.

columns (1) and (2) are the same. To evaluate the importance of restricting attention to households with strictly positive net worth, in column (3) we estimate the regression in levels (of net worth and income) on the sample ages 35 and older. We again see an economically large association between the CCEI and levels of wealth, though this relationship is not estimated precisely; the coefficient on the CCEI is significant only at the 10 percent level (p-value=0.058).

We interpret our CCEI scores as capturing aspects of decision-making quality. We take this view because choices that are closer to satisfying GARP can be seen as more purposeful; they reflect more consistent treatment of tradeoffs regardless of preferences, information or beliefs. An alternative view is that the CCEI captures unobserved aspects of education or cognitive skill that are correlated with financial outcomes through their correlation with preferences, beliefs, or unobserved constraints. We are conditioning, quite flexibly, on measures of education (the CentERpanel survey does not include measures of IQ and this is an important topic for future work). Nevertheless, we can assess whether *unobserved* aspects of education are driving the relationship between the CCEI and wealth if we assume that these unobserved variables are positively correlated with *observed* education levels. If they are, and if these unobserved variables are important sources of the observed correlation between consistency and wealth, then conditioning on observed education should have a substantial effect on the estimated coefficient on the CCEI. In column (4) of Table 7, we repeat the estimation reported in column (2), after omitting the education level of the financial respondent. Comparing the estimates from columns (2) and (4), we see that removing the controls for education has only a modest effect on the estimated coefficient on the CCEI. In this way, we find little evidence that the relationship between the CCEI and wealth is driven by a correlation between the CCEI and unobserved aspects of education.

Next we turn to evaluate the magnitude of the association between the CCEI and wealth relative to other measures of interest. Restricting attention to the sample ages 35 and older, in column (5), we add our measure of CCEI from the combined data set (combining the actual data from the experiment and the mirror-image data) to the log specification. We find no evidence that, conditional on the CCEI score from the actual data, the CCEI score from the combined data has an independent relationship with wealth. Adding the CCEI from combined data as a regressor has only a modest effect on the point estimate on the CCEI; and the point estimate of the conditional relationship between the CCEI from combined data and wealth is small, but imprecisely estimated. These results are consistent with the idea that the CCEI from the combined data, while adding natural requirements of decision-making quality in our experimental setting, merely represents a noisier measure of the aspects of decision-making quality captured by the CCEI.

In column (6) we add a control for risk attitudes by including the average fraction of tokens the financial respondent allocated to the cheaper asset. The point estimate on this measure of risk attitudes indicates that risk tolerance is negatively associated with wealth, consistent with precautionary saving. The estimate is economically large; a standard deviation increase in the fraction placed in the cheaper asset is associated with 10 percent less wealth. The estimate is fairly imprecise, however; we cannot reject a null hypothesis of no relationship (p-value=0.16). The results are qualitatively similar when, in column (7), we condition on both risk tolerance and the CCEI from the combined data. Finally, the point estimates on the alternative measures of GARP violations based on Varian (1990, 1991) and HM yield qualitatively similar conclusions. In the case of the HM index, the standard errors are relatively small and the opposite is true of the Varian (1990, 1991) index. These results are presented in Online Appendix III.

7 Related literature

Revealed preference tests have been applied to aggregate consumption data. However, real-world data do not provide a particularly rigorous test of consistency because choice sets are such that budget lines do not cross frequently (see Blundell et al., 2003). Furthermore, even a high level of consistency in the individual-level decisions does not imply that aggregate data are consistent. Cox (1987), Sippel (1997), Mattei (2000), Harbaugh et al. (2001), and Andreoni and Miller (2002), among others, ask whether behavior in the laboratory is consistent with utility maximization. The Bronars (1987) test has been widely used, so it allows us to relate our results to this literature. Our study has the highest Bronars power of one (all random subjects had violations). We note that even random behavior can appear consistent if the sample size is small, as it often is in experimental studies.

Our sociodemographic data creates the opportunity to analyze the correlates of experimental outcomes. Our paper thus contributes to the emerging literature on the relation of laboratory behaviors to cognitive ability, typically measured using IQ tests or SAT scores (see, for example, Benjamin et al., 2006, and Dohmen et al., 2010). Different from earlier studies, we use the extent of consistency with utility-maximizing behavior as single measure for "economic cognition" and investigate the correlation between consistency under laboratory conditions and demographic and economic characteristics. The relation of our experimental results to individual characteristics enables us to shed some light on the external validity of our findings, which Levitt and List (2007) and Falk and Heckman (2009) point out is a critical concern for experimental studies.

Also related to the design of our experiment regarding choice under risk, but somewhat further afield, there is a large and growing experimental literature that investigates whether the risk attitudes that arise in the laboratory are connected to attributes that subjects bring to the experiments from outside the lab. von Gaudecker et al. (forthcoming) also conducted risk experiments with CentERpanel members. Our findings in this paper are consistent with their conclusion that "while many people exhibit consistent choice patterns, some have very high error propensities."

8 Concluding remarks

Some decision-makers are better than others. But it is usually hard to tell whether a decision-maker has made a bad choice; he might have uncommon preferences, or face unobserved constraints, or hold (reasonable) beliefs that rationalize his decision. Standard economic analysis takes a libertarian approach; in the absence of data that allow us to identify bad decisions, we assume that all choices are good. The libertarian approach has obvious appeal. We rightly hesitate to evaluate the quality of decisions when we do not have sufficient information to make a definitive judgement.

This study suggests an alternative path. We offered a new field experimental design—employing graphical representations of standard consumer decision problems and using a rich pool of subjects—that enables us to collect richer data than has been possible in the past. These data allow us to say some choices are better than others, in that some choices are more rational than others. Because the data are provided by a relatively large and heterogenous sample, we can thoroughly analyze the correlates of individual levels of rationality and relate rationality in this simple domain to important economic outcomes like wealth. The conclusions of our investigation can be summarized under three headings:

- The first important finding from the experiment is that many subjects reveal nearly perfect consistency with utility maximization in the individual-level decisions. Standard tests suggest that nearly half of our subjects exhibit behavior that appears to be almost optimizing in the sense that their choices nearly satisfy GARP. At the same time, there is important heterogeneity in the consistency of choice.

- The second important finding is that consistency levels are correlated with demographic and economic characteristics. Our study provides, to our knowledge, the first experimental evidence on the question "who is more rational?" This evidence, and the methods developed to gather it, may ultimately prove to be useful for the formulation of economic policy. For example, the relationships between sociodemographics and levels of rationality can be used to inform the design of social programs (Manski, 2001) or (libertarian) paternalistic policies (Thaler and Sunstein, 2003).

- The third important finding is that the differences in the experimental consistency scores help to explain differential patterns of wealth across households. The magnitudes are large, implying that a standard deviation increase in the consistency score is associated with in 15-19 percent more wealth. We view this finding as unexpected given the substantial heterogeneity in our experimental outcomes, and the very brief experimental exposure.

The experimental techniques that we have developed provide some promising tools for future work in these areas, and the results also suggest a number of potential directions. In addition, the experimental platforms and analytical techniques are applicable to many other types of individual choice problems.

References

[1] Ahn, D., S. Choi D. Gale, and S. Kariv (2010) "Estimating Ambiguity Aversion in a Portfolio Choice Experiment." Mimeo.

[2] Afriat, S. (1967) "The Construction of a Utility Function from Expenditure Data." *Econometrica*, *6*, pp. 67-77.

[3] Afriat, S. (1972) "Efficiency Estimates of Production Functions." *International Economic Review*, *8*, pp. 568-598.

[4] Agarwal, S., J. Driscoll, X. Gabaix and D. Laibson (2009) "The Age of Reason: Financial Decisions over the Life-Cycle with Implications for Regulation." *Brookings Papers on Economic Activity*, *2*, pp. 51-117.

[5] Ameriks, J., A. Caplin and J. Leahy (2003) “Wealth Accumulation and the Propensity to Plan.” *Quarterly Journal of Economics*, *118*, pp. 1007-1047.

[6] Andreoni, J. and W. Harbaugh (2006) “Power Indices for Revealed Preference Tests.” Mimeo.

[7] Andreoni, J. and J. Miller (2002) “Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism.” *Econometrica*, *70*, pp. 737-753.

[8] Banks, J. (2010) “Cognitive Function, Financial Literacy and Financial Outcomes at Older Ages: Introduction.” *Economic Journal*, *120*, pp. 357-362.

[9] Banks, J., C. O ’Dea and Z. Oldfield (2010) “Cognitive Function, Numeracy and Retirement Saving Trajectories.” *Economic Journal*, *120*, pp. 381-410.

[10] Becker, G. (1962) “Irrational Behavior and Economic Theory.” *Journal of Political Economy*, *70*, pp. 1-13.

[11] Bellemare, C. S. Kröger, and A. van Soest (2008) “Measuring inequity aversion in a heterogeneous population using experimental decisions and subjective probabilities.” *Econometrica*, *76*, pp. 815-839.

[12] Benjamin, D., S. Brown and J. Shapiro (2006) “Who is “Behavioral”? Cognitive Ability and Anomalous Preferences.” Mimeo.

[13] Bernheim, B. D., J. Skinner, and S. Weinberg (2001) “What Accounts for the Variation in Retirement Wealth among U.S. Households?” *American Economic Review, 91*, pp. 832-857.

[14] Blundell, R., M. Browning and I. Crawford (2003) “Nonparametric engel curves and revealed preference.” *Econometrica*, *71*, pp. 205-240.

[15] Bronars, S. (1987) “The power of nonparametric tests of preference maximization.” *Econometrica*, *55*, pp. 693-698.

[16] Charness, G., E. Karni and D. Levin (2007) “Individual and Group Decision Making under risk: An Experimental Study of Bayesian Updating and Violations of First-Order Stochastic Dominance.” *Journal of Risk and Uncertainty*, *35*, pp. 129-48.

[17] Choi S., R. Fisman, D. Gale, and S. Kariv (2007a) "Revealing Preferences Graphically: An Old Method Gets a New Tool Kit." *American Economic Review*, Papers & Proceedings, *97*, pp. 153-158.

[18] Choi S., R. Fisman, D. Gale, and S. Kariv (2007b) "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review*, *97*, pp. 1921-1938.

[19] Cox, J. (1997) "On Testing the Utility Hypothesis." *Economic Journal*, *107*, pp. 1054-1078.

[20] Dean, M. and D. Martin (2010) "How Rational are your Choice Data?" Mimeo.

[21] Dohmen, T., A. Falk, D. Huffman and U. Sunde (2010) "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, *100*, pp. 1238-1260.

[22] Falk, A. and J. Heckman (2009) "Laboratory Experiments are a Major Source of Knowledge in Social Sciences." *Science*, *326*, pp. 535-38.

[23] Famulari, M. (1995) "A Household-Based Nonparametric Test of Demand Theory." *Review of Economics and Statistics*, *2*, pp. 371-382.

[24] Fisman, R., S. Kariv and D. Markovits (2007) "Individual Preferences for Giving." *American Economic Review*, *97*, pp. 1858-1876.

[25] Fisman, R., S. Kariv and D. Markovits (2010) "Exposure to Ideology and Distributional Preferences." Mimeo.

[26] Gul, F. (1991) "A Theory of Disappointment in Decision Making under Uncertainty." *Econometrica*, *59*, pp. 667–686.

[27] Hadar, J. and W. Russell (1969) "Rules for Ordering Uncertain Prospects." *American Economic Review*, *59*, pp. 25-34.

[28] Heckman, J. (1979) "Sample Selection Bias as a Specification Error." *Econometrica*, *47*, pp. 153-161.

[29] Harbaugh, W., K. Krause and T. Berry (2001) "GARP for Kids: On the Development of Rational Choice Behavior." *American Economic Review*, *91*, pp. 1539-1545.

[30] Houtman, M. and J. Maks (1985) "Determining all Maximial Data Subsets Consistent with Revealed Preference." *Kwantitatieve Methoden*, *19*, pp. 89-104.

[31] Jappelli, T. (2010) "Economic Literacy: An International Comparison." *Economic Journal*, *120*, pp. 429-451.

[32] Levitt, S. and J. List (2007) "What do Laboratory Experiments Tell Us About the Real World?" *Journal of Economic Perspectives*, *21*, pp. 153-174.

[33] Manski, C. (2001) "Designing Programs for Heterogeneous Populations: The Value of Covariate Information." *American Economic Review*, Papers & Proceedings, *91*, pp. 103-106.

[34] Mas-Colell, A. (1977) "The Recoverability of Consumer's Preferences from Market Demand Behavior." *Econometrica*, *45*, pp. 1409-1430.

[35] Mas-Colell, A. (1978) "On Revealed Preferences Analysis." *Review of Economic Studies*, *45*, pp. 121-131.

[36] Mattei, A. (2000) "Full-Scale Real Tests Of Consumer Behavior Using Experimental Data." *Journal of Economic Behavior & Organization*, *43*, pp. 487-497.

[37] Papke, L. and J. Wooldridge (1996), "Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates." *Journal of Applied Econometrics*, *11*, pp. 619-632.

[38] Quiggin, J. (1990) "Stochastic Dominance in Regret Theory." *Review of Economic Studies*, *57*, pp. 503-511.

[39] Ramsey, J. (1969) "Tests for specification errors in classical linear least-squares regression analysis." J*ournal of the Royal Statistical Society*, *31*, pp. 350-371.

[40] Sippel, R. (1997) "An Experiment on the Pure Theory of Consumer's Behavior." *Economic Journal*, *107*, pp. 1431-1444.

[41] Smith, J., J. McArdle and R. Willis (2010) "Financial Decision Making and Cognition in a Family Context." *Economic Journal*, *120*, pp. 363-380.

[42] Terza, J. (1998) “Estimating Count Data Models with Endogenous Switching: Sample Selection and Endogenous Treatment Effects.” *Journal of Econometrics*, *84*, pp. 129-154.

[43] Thaler, R. and C. Sunstein (2003) “Source Libertarian Paternalism.” *American Economic Review*, Papers & Proceedings, *93*, pp. 175-179.

[44] Varian, H. (1982) “The Nonparametric Approach to Demand Analysis.” *Econometrica*, *50*, pp. 945-972.

[45] Varian, H. (1983) “Non-Parametric Tests of Consumer Behaviour.” *Review of Economic Studies*, *50*, pp. 99-110.

[46] Varian, H. (1990): “Goodness-of-Fit in Optimizing Models.” *Journal of Econometrics*, *46*, pp. 125-140.

[47] Varian, H. (1991) “Goodness-of-Fit for Revealed Preference Tests.” Mimeo.

[48] Van Den Berg, G., D. Deeg, M. Lindeboom and F. Portrait (2010) “The Role of Early-Life Conditions in the Cognitive Decline due to Adverse Events Later in Life.” *Economic Journal*, *120*, pp. 411-428.

[49] von Gaudecker, H-M., A. van Soest and E. Wengström (forthcoming) “Heterogeneity in Risky Choice Behaviour in a Broad Population.” *American Economic Review.*

[50] Wakker, P. (1993) “Savage’s Axioms Usually Imply Violation of Strict Stochastic Dominance.” *Review of Economic Studies*, *60*, pp. 487-493.

Table 1. Sociodemographic information

	Participants	Dropouts	Non-participants
Female	45.43	37.89	50.00
Age			
16-34	18.53	3.16	26.14
35-49	26.14	12.11	32.13
50-64	35.62	38.42	27.58
65+	19.71	46.32	14.15
Education			
Low	33.59	42.63	30.99
Medium	29.70	22.63	31.61
High	36.72	34.74	37.40
Household monthly income			
€0-2500	22.42	34.73	21.28
€2500-3499	25.13	26.32	18.90
€3500-4999	28.85	16.32	28.93
€5000+	23.60	22.63	30.89
Occupation			
Paid work	53.13	39.47	62.91
House work	11.59	7.89	8.78
Retired	20.90	42.63	13.95
Others	14.38	10.00	14.36
Household composition			
Partner	80.88	67.89	82.64
# of children	0.84	0.32	1.09
# of obs.	1182	190	968

Table 2. CCEI scores

	Mean	Std. Dev.	Percentiles 10	25	50	75	90	# of obs.
All	0.881	0.141	0.676	0.808	0.930	0.998	1.000	1182
Female	0.874	0.147	0.666	0.796	0.928	0.998	1.000	537
Age								
16-34	0.920	0.119	0.734	0.881	0.979	1.000	1.000	219
35-49	0.906	0.123	0.708	0.853	0.966	1.000	1.000	309
50-64	0.863	0.142	0.666	0.784	0.901	0.985	1.000	421
65+	0.843	0.164	0.595	0.770	0.882	0.981	1.000	233
Education								
Low	0.863	0.143	0.665	0.782	0.906	0.987	1.000	397
Medium	0.881	0.140	0.689	0.814	0.926	0.998	1.000	351
High	0.899	0.137	0.686	0.842	0.963	1.000	1.000	430
Household monthly income								
€0-2500	0.856	0.154	0.617	0.769	0.911	0.983	1.000	269
€2500-3499	0.885	0.133	0.705	0.809	0.925	0.999	1.000	302
€3500-4999	0.882	0.141	0.649	0.817	0.932	0.999	1.000	345
€5000+	0.901	0.131	0.729	0.836	0.968	1.000	1.000	266
Occupation								
Paid work	0.896	0.131	0.705	0.833	0.950	1.000	1.000	628
House work	0.873	0.151	0.649	0.795	0.937	0.999	1.000	137
Retired	0.839	0.158	0.597	0.767	0.876	0.971	1.000	247
Others	0.891	0.129	0.712	0.809	0.936	0.998	1.000	170
Household composition								
Partner	0.878	0.142	0.673	0.802	0.927	0.998	1.000	956
Children	0.899	0.128	0.704	0.835	0.959	1.000	1.000	490

Table 3. CCEI scores for the combined data set

	Mean	Std. Dev.	Percentiles 10	25	50	75	90	Δ Mean
All	0.733	0.229	0.394	0.584	0.775	0.943	0.985	0.148
Female	0.733	0.224	0.409	0.588	0.767	0.941	0.984	0.141
Age								
16-34	0.786	0.228	0.442	0.637	0.881	0.976	0.995	0.134
35-49	0.782	0.206	0.481	0.652	0.845	0.962	0.991	0.124
50-64	0.700	0.225	0.371	0.552	0.735	0.898	0.973	0.163
65+	0.679	0.242	0.334	0.489	0.703	0.902	0.968	0.165
Education								
Low	0.699	0.226	0.374	0.535	0.732	0.902	0.967	0.163
Medium	0.733	0.226	0.394	0.595	0.768	0.941	0.986	0.148
High	0.767	0.227	0.428	0.625	0.849	0.968	0.992	0.131
Household monthly income								
€0-2500	0.706	0.218	0.382	0.535	0.737	0.902	0.977	0.150
€2500-3499	0.741	0.220	0.439	0.612	0.768	0.946	0.986	0.143
€3500-4999	0.730	0.236	0.388	0.556	0.782	0.950	0.984	0.151
€5000+	0.755	0.238	0.383	0.627	0.841	0.952	0.992	0.146
Occupation								
Paid work	0.758	0.222	0.428	0.615	0.817	0.955	0.991	0.139
House work	0.719	0.233	0.380	0.548	0.765	0.928	0.986	0.154
Retired	0.675	0.231	0.334	0.502	0.698	0.872	0.964	0.164
Others	0.738	0.231	0.406	0.599	0.793	0.951	0.983	0.153
Household composition								
Partner	0.729	0.229	0.389	0.583	0.771	0.938	0.984	0.149
Children	0.760	0.216	0.443	0.614	0.815	0.952	0.987	0.139

Table 4. The correlation between CCEI scores and subjects' individual characteristics (OLS)

	(1)	(2)
Constant	.887***	.735***
	(.022)	(.037)
Female	-.024***	-.011
	(.009)	(.015)
Age		
35-49	-.016	-.007
	(.011)	(.020)
50-64	-.052***	-.077***
	(.011)	(.020)
65+	-.051**	-.081**
	(.020)	(.032)
Education		
Medium	.009	.021
	(.011)	(.017)
High	.026**	.060***
	(.011)	(.018)
Income		
€2500-3499	.026**	.026
	(.012)	(.019)
€3500-4999	.020	.006
	(.013)	(.020)
€5000+	.033**	.017
	(.014)	(.022)
Occupation		
Paid work	.028	.030
	(.018)	(.026)
House work	.047**	.039
	(.021)	(.030)
Others	.037*	.035
	(.019)	(.030)
Household composition		
Partner	-.026**	-.023
	(.011)	(.018)
# of children	.001	.001
	(.004)	(.007)
R^2	.068	.058
# of obs.	1182	1182

Omitted categories: male, age under 35, low education (primary and lower secondary education), household gross monthly income under €2500, retired, and not having a partner. Standard errors in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance levels, respectively.

Table 5. The correlation between CCEI scores and subjects' individual characteristics
(sample-selection)

	(1)		(2)	
	Outcome	Selection	Outcome	Selection
Constant	.888***	.544*	.891***	-2.077***
	(.022)	(.311)	(.023)	(.209)
Female	-.024***	.084	-.024***	-.031
	(.009)	(.103)	(.009)	(.068)
Age				
35-49	-.016	-.556**	-.016	-.133
	(.011)	(.230)	(.011)	(.102)
50-64	-.051***	-1.024***	-.052***	-.393***
	(.011)	(.220)	(.011)	(.102)
65+	-.050**	-1.556***	-.051**	-.824***
	(.021)	(.263)	(.020)	(.154)
Education				
Medium	.009	.191	.009	-.036
	(.011)	(.122)	(.011)	(.081)
High	.026**	.168	.026**	.006
	(.011)	(.117)	(.011)	(.084)
Income				
€2500-3499	.025**	.303**	.025**	.281***
	(.012)	(.125)	(.012)	(.094)
€3500-4999	.019	.426***	.019	.186**
	(.013)	(.141)	(.014)	(.094)
€5000+	.033**	.064	.033**	.080
	(.014)	(.147)	(.014)	(.106)
Occupation				
Paid work	.028	-.202	.029	-.040
	(.018)	(.172)	(.018)	(.131)
House work	.046**	.108	.046**	.083
	(.020)	(.200)	(.020)	(.148)
Others	.037**	.081	.037*	.110
	(.019)	(.196)	(.019)	(.147)
Household composition				
Partner	-.026**	.262**	-.027**	.123
	(.011)	(.119)	(.011)	(.092)
# of children	.001	.145**	.001	.031
	(.004)	(.068)	(.004)	(.036)
Participation ratio		1.231***		3.387***
		(.205)		(.125)
ρ	-.028		-.047	
	(.083)		(.063)	
Log peudolikelihood	210.856		-371.973	
# of obs.	1372		2340	

Table 5
(continued)

	(3)		(4)	
	Outcome	Selection	Outcome	Selection
Constant	.759***	.545*	.757***	-2.067***
	(.043)	(.314)	(.038)	(.208)
Female	-.013	.084	-.011	-.032
	(.015)	(.104)	(.015)	(.068)
Age				
35-49	-.001	-.554**	-.009	-.135
	(.022)	(.223)	(.020)	(.101)
50-64	-.062**	-1.023***	-.079***	-.397***
	(.024)	(.212)	(.020)	(.102)
65+	-.049	1.557***	-.078**	-.822***
	(.042)	(.258)	(.032)	(.154)
Education				
Medium	.016	.191	.021	-.036
	(.018)	(.120)	(.017)	(.081)
High	.054***	.169	.059***	.007
	(.018)	(.117)	(.018)	(.084)
Income				
€2500-3499	.017	.304**	.022	.276***
	(.021)	(.127)	(.019)	(.093)
€3500-4999	-.006	.428***	.003	.174*
	(.022)	(.138)	(.020)	(.094)
€5000+	.015	.065	.018	.075
	(.022)	(.145)	(.022)	(.106)
Occupation				
Paid work	.034	-.203	.031	-.035
	(.027)	(.173)	(.026)	(.131)
House work	.036	.109	.038	.075
	(.030)	(.205)	(.030)	(.148)
Others	.032	.081	.034	.110
	(.030)	(.193)	(.030)	(.146)
Household composition				
Partner	-.032	.261**	-.026	.126
	(.020)	(.115)	(.018)	(.091)
# of children	-.000	.145**	.002	.028
	(.007)	(.062)	(.007)	(.036)
Participation ratio		1.230***		3.378***
		(.234)		(.125)
ρ	-.396		-.155	
			(.075)	
Log peudolikelihood			-949.787	
# of obs.	1372		2340	

Omitted categories: male, age under 35, low education (primary and lower secondary education), household gross monthly income under €2500, retired, and not having a partner. Standard errors in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance levels, respectively.

Table 6. Household 2008-2009 net worth summary statistics

(2008 Euros)

Mean		164,130
Std. Dev.		243,548
Max		3,984,151
Min		-180,700
Percentiles	1	-68,237
	5	-4,810
	10	0
	25	10,780
	50	92,979
	75	242,054
	90	412,494
	95	523,839
	99	955,599
# of obs.		703

Table 7. The relationship between households' net worth and CCEI scores

	(1)	(2)	(3)
CCEI	1.170**	1.425**	99933.2*
	(0.535)	(0.565)	(52656.0)
CCEI (combined dataset)			
Risk aversion			
Log 2008 household income	0.623***	0.601***	
	(0.123)	(0.127)	
2008 household income			1.74***
			(0.3)
Female	-0.275*	-0.228	-28223.9*
	(0.154)	(0.164)	(15906.3)
Age	0.004	-0.286	-33974.7
	(0.205)	(0.316)	(27100.3)
Age^2	0.002	0.006	726.5
	(0.004)	(0.005)	(471.1)
Age^3	0.000	0.000	-4.3
	(0.000)	(0.000)	(2.7)
Partner	0.623***	0.682***	48106.5***
	(0.173)	(0.183)	(16995.7)
# of children	0.125	0.103	14472.9*
	(0.086)	(0.092)	(8291.6)
Education			
Pre-vocational	0.242	0.267	13056.4
	(0.459)	(0.459)	(43981.0)
Pre-university	0.528	0.600	57288.3
	(0.474)	(0.476)	(45189.8)
Senior vocational training	0.407	0.403	27365.6
	(0.465)	(0.469)	(42967.4)
Vocational college	0.485	0.448	27964.9
	(0.450)	(0.452)	(42704.3)
University	0.637	0.679	73733.5
	(0.463)	(0.470)	(48008.2)
Constant	0.229	5.932	335793.4
	(3.554)	(5.862)	(5.0E+05)
R^2	0.217	0.179	0.191
# of obs.	566	517	568

Table7.
(Continued)

	(4)	(5)	(6)	(7)
CCEI	1.490***	1.348*	1.545***	1.563**
	(0.574)	(0.714)	(0.591)	(0.735)
CCEI (combined dataset)		0.078		-0.018
		(0.381)		(0.373)
Risk aversion			-1.166	-1.165
			(0.828)	(0.829)
Log 2008 household income	0.629***	0.602***	0.595***	0.595***
	(0.124)	(0.127)	(0.128)	(0.129)
2008 household income				
Female	-0.258	-0.229	-0.232	-0.232
	(0.162)	(0.164)	(0.166)	(0.166)
Age	-0.277	-0.284	-0.307	-0.308
	(0.318)	(0.316)	(0.313)	(0.315)
Age^2	0.006	0.006	0.007	0.007
	(0.005)	(0.005)	(0.005)	(0.005)
Age^3	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Partner	0.683***	0.682***	0.726***	0.725***
	(0.184)	(0.183)	(0.187)	(0.188)
# of children	0.106	0.103	0.092	0.092
	(0.093)	(0.093)	(0.094)	(0.095)
Education				
Pre-vocational		0.264	0.331	0.331
		(0.461)	(0.483)	(0.484)
Pre-university		0.596	0.676	0.677
		(0.486)	(0.498)	(0.498)
Senior vocational training		0.403	0.480	0.481
		(0.469)	(0.493)	(0.494)
Vocational college		0.443	0.549	0.550
		(0.452)	(0.475)	(0.480)
University		0.672	0.745	0.746
		(0.474)	(0.498)	(0.502)
Constant	5.451	5.888	6.938	6.947
	(6.110)	(5.879)	(5.786)	(5.812)
R^2	0.170	0.178	0.186	0.184
# of obs.	517	517	507	507

The CCEI scores for the combined dataset is computed after combining the actual data from the experiment and the mirror-image data. Risk aversion measured by the average fraction of tokens allocated to the cheaper asset. The groupings of different levels of education are based on the categorization of Statistics Netherlands (Centraal Bureau voor de Statistiek). For a complete description see http://www.centerdata.nl/en/centerpanel. Standard errors in parentheses. *, **, and *** indicate 10, 5, and 1 percent significance levels, respectively.

Figure 1. Mean CCEI scores

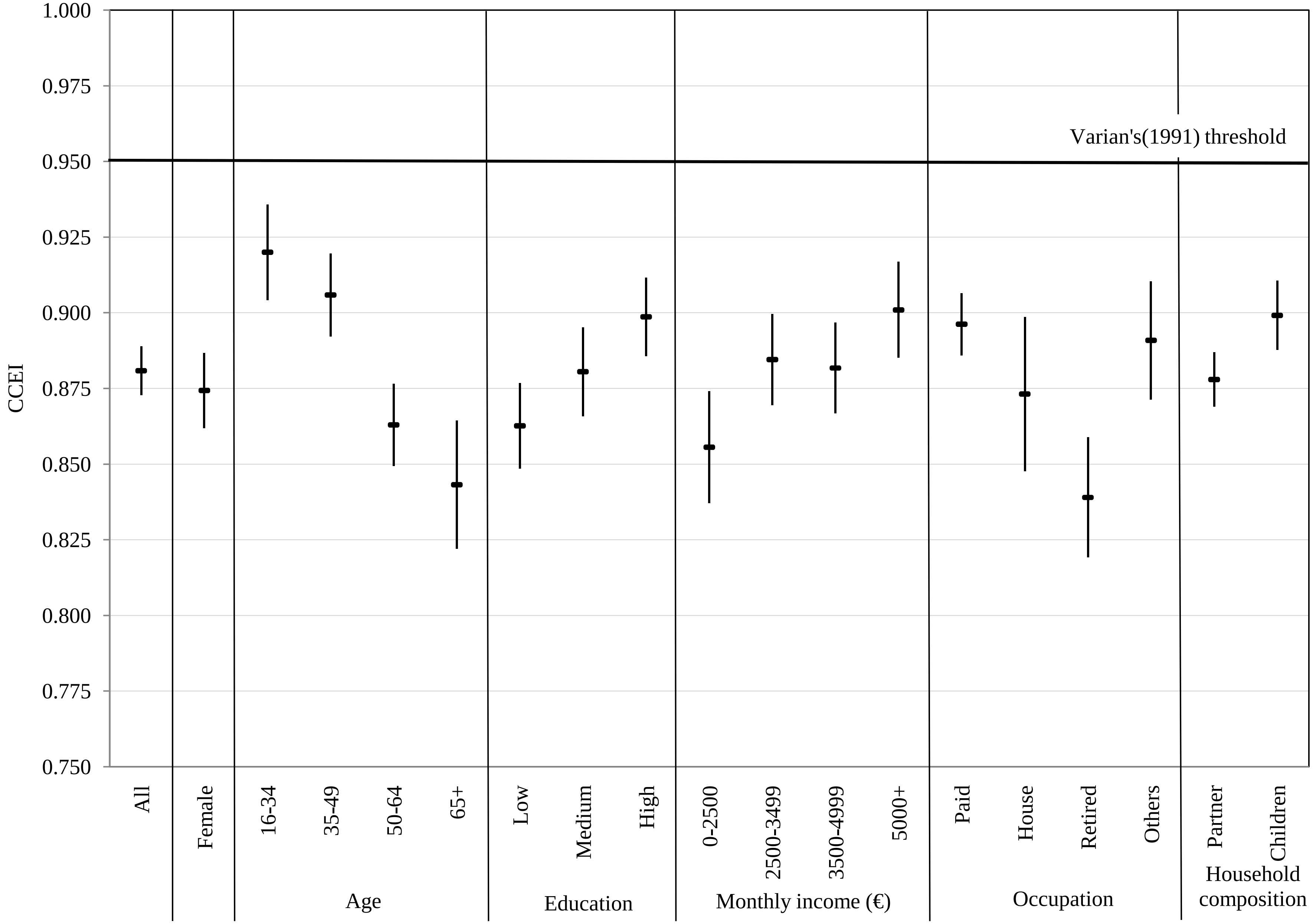
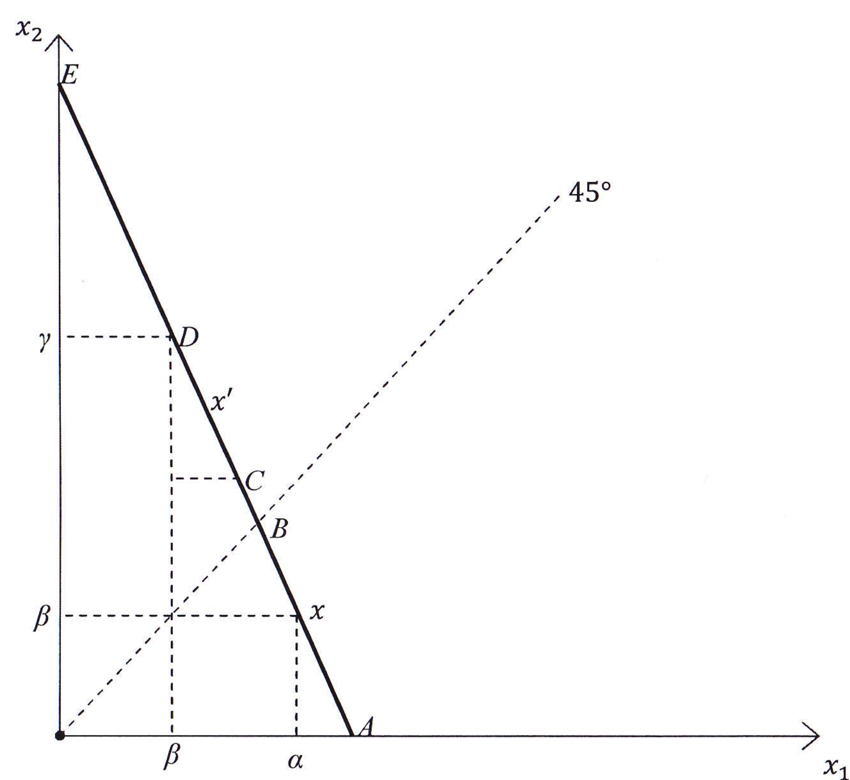


Figure 2. A violation of stochastic dominance



The individual can choose any allocation x' (position along CD) but prefers allocation x (position along AB) such that $F_{x'} < F_x$ where $F_{x'}$ and F_x are the resulting payoff distributions.

Figure 3. The average fraction of tokens allocated to the cheaper asset

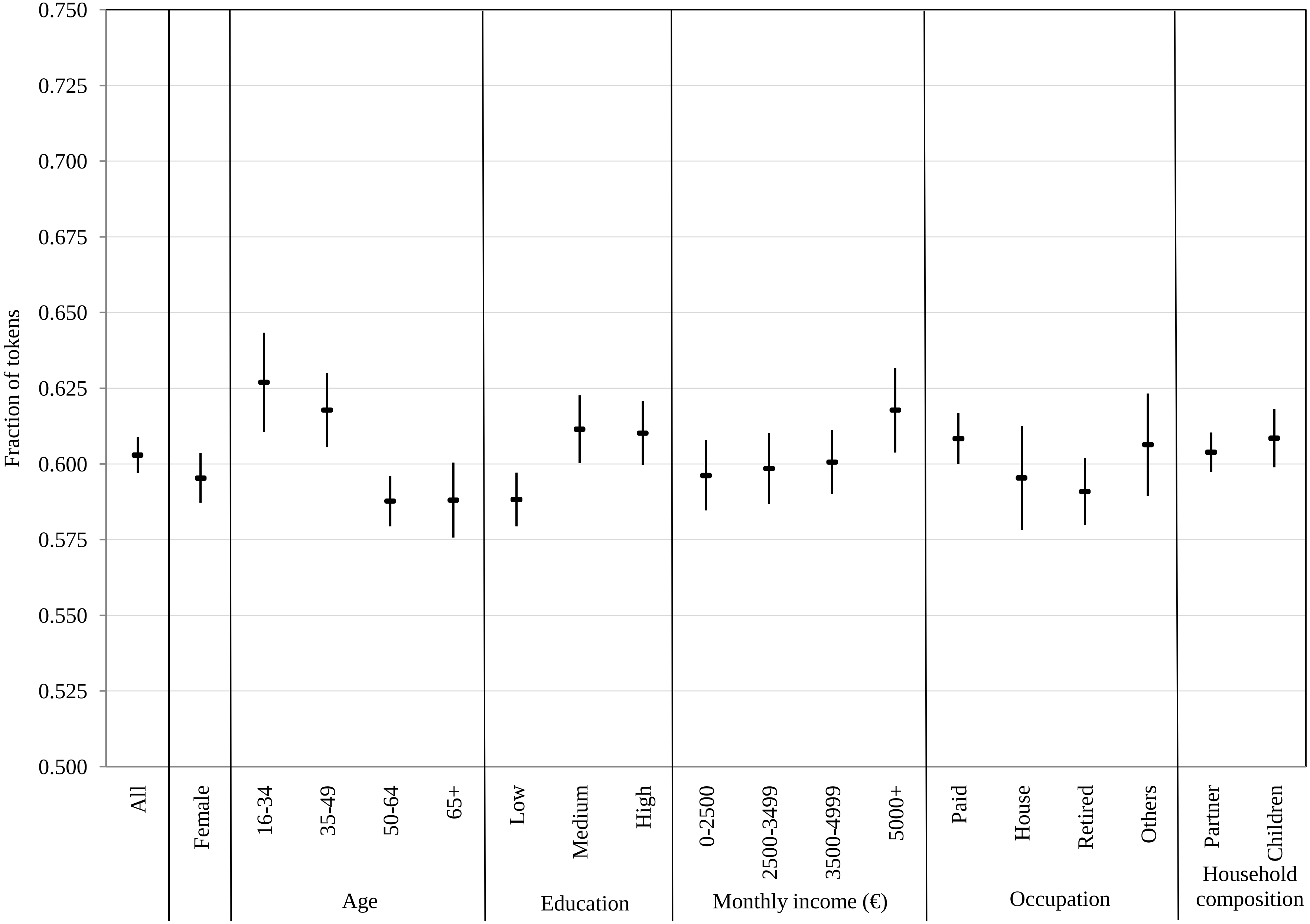


Figure 4. The average fraction of tokens allocated to the cheaper asset