Attention, Revealed Preferences, and Consequentialist Behavior

Peter Hammond[a], Stefan Traub[b,*]

[a]*Department of Economics, University of Warwick, Coventry, UK*
[b]*Department of Economics, University of Bremen, Germany*

January, 2009

Abstract We present an experimental analysis of subjects' individual information collection and choice behavior in a series of multi-stage portfolio-selection problems. We recorded not only subjects' choices but also monitored the process of mouse movement that revealed information about the different portfolios. Our main results are, first, that subjects focussed their attention only on a small subset of consequences and, second, that the conditional probability of being substantively rational in terms of GARP if the previous choice was GARP-consistent was very high. These results suggest that the observed deviations from substantive rationality come from the fact that people deliberately choose to restrict the amount of information that is used to make up their minds to a "satisfactory" minimum.

JEL classification: C91, D83

Keywords: Rationality, Uncertainty, Decision Making, Consequentialism, Attention

*Corresponding author. Institut für Volkswirtschaftslehre, Universität Bremen FB 7, Hochschulring 4, D-28359 Bremen, Germany; phone +49–218–2765, fax +49-218-3155, traub@uni-bremen.de.

1 Introduction

Hammond (1988) showed that the independence axiom which gives expected utility its empirical content is implied by a basic normative principle which he called consequentialism:

> "...the normative principle that behaviour should be ordinal - i.e. maximize a complete and transitive preference ordering - could be justified by requiring dynamic choice in decision trees to depend only upon outcomes and not upon the structure of the decision tree ... it is no misuse of the terminology to call it consequentialism as well ... In single person decision theory, consequentialism is defined informally to mean that behavior is explicable merely by its consequences" (p. 503).

However, violations of the independence axiom and other basic normative principles of rational decision making such as Allais' Paradox (Allais, 1953) seem to occur too frequently and to be too persistent in order to be dismissed as choice anomalies or just to be wiped out by quantifying additional independent variables in the utility function (Machina, 1987) or extending the domain of consequences (Hammond, 1989). Hence, many alternatives to substantive rationality have been formulated in the literature such as Simons (1955) bounded rationality model. He suggested to

> "...replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacity that are actually possessed

> by organisms, including man, in the kinds of environments in which such organisms exist" (p. 99).

Generically, the satisficing approach comprises a binary payoff function, i.e. it distinguishes only between satisfactory and unsatisfactory outcomes, instead of utility function, and an aspiration level. In contrast to substantive rationality people are assumed to have a preference for rules (procedural rationality) that simplify decision problems (heuristics, rules of thumb etc.) rather than for consequences (Simon 1976, 1978; Elster 1979; Sen 1995).

Tversky and Kahneman's (1974) heuristics and biases approach holds that

> "... people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors" (p. 1124).

Abelson and Levi (1985) judged that "this view is thus related to that of bounded rationality, but the boundedness comes not from the impracticality of spending the time and effort to make truly optimal judgments but from genuine failure to appreciate normatively appropriate strategies ... in general, the picture has a pessimistic tone" (p. 233).

What level of rationality applies? Are we just "slaves of motivational forces" (Abelson and Levi, 1985, p. 234), limited rational (as in the heuristics and biases approach), procedurally rational or still substantive rational? Experimental work comparing different theories with EU performed by Harless and Camerer (1994), Hey and Orme (1994), Carbone and Hey (1994, 1995), and Morone and Schmidt (2008) showed that EU performs surprisingly well, in particular, when controlling for the parsimony of the model in terms of parameters and allowing for noise or decision errors. Hey (2005) concluded that "... subjects do appear to have a plan and they do appear to implement it" (p. [redacted]). Morone and Schmidt (2008) emphasized that "EU does not perform substantially worse than its alternatives" (p. 9). Do we, thus, need better descriptive theories, or is consequentialism less inadequate to describe actual behavior than widely held?

The idea of this paper is that answering the above questions for the level of rationality of human decision making obviously requires us to study the *process of decision making* or to include rules within consequences. However, in economic standard theory the process of information collection and decision making remains a black box, because utility maximization implies that a decision maker always chooses that option which yields the highest utility. The famous as-if assumption (Friedman and Savage, 1948) nicely reflects economists' careless handling of decision processes: "Whatever the psychological mechanism whereby the individuals make choices, the choices appear to display some consistency, which can apparently be described by our utility hypothesis" (p. 298).[1] What if we can find evidence for the observed errors or deviations from substantive rationality coming from the fact

[1]A conceptual 4-phase model of judgment was formulated by Simon (1966) and Hogarth (1987). It distinguishes between problem recognition, identification of alternatives, evaluation of alternatives, and selection of an alternative. The task environment (frame) is likely to matter for the first three phases. Evaluation of alternatives, for instance, could results in a recoding of outcomes into gains and losses. In a dynamic framework there might also be feedback effects and learning.

that people deliberately choose to restrict the amount of information that is used to make up their minds to a "satisfactory" minimum? One could call this kind of behavior "rationally bounded" because people still have a preference for consequences (and exhibit maximizing behavior). Obviously, this is different from bounded rationality, where it is assumed that people have a preference for rules (and exhibit satisficing behavior).

The current paper experimentally investigates subjects' individual information collection and choice behavior in a series of 16 three-stage portfolio-selection problems. We recorded not only subjects' choices but also monitored the process of mouse movement that revealed information about the different portfolios. In the first step of our analysis, we test for every subject the predictive power of the supporting set of consumption bundles of GARP (Varian, 1982), both in a strict normative sense and in a nonparametric test version allowing for decision errors. Then, using regression analysis, we investigate the factors that had an impact on our subjects' attention in terms of the time the left the different possible allocations visible on the screen until the final decision was made. Eventually, we combine the data on attention and choices in a logit regession in order to find out whether and how subjects' attention influenced the rationality of their choices. We also test for homogeneity of preferences.

Only a bit more than a fifth of our subjects is classified as substantively rational in terms of revealed-preference theory (GARP) even by our more generous nonparametric test. However, most subjects who made a rational choice on the second stage of a portfolio-selection problem turned out to be rational on the third stage as well. In general, the share of GARP consistent choices was significantly greater for male than for female subjects. Subjects concentrated their attention on a small subset of allocations from which they usually made their final choices. Amongst other things, the regression analysis showed learning effects which were more pronounced for females. The main result of the logit regression is that the more attention was paid to consequences contained in the supporting set, the likelier that a GARP consistent final choice followed.

The paper is organized as follows. In the next section we review the most relevant experimental literature. Section 3 gives a detailed description of our experiment. The results are presented in Section 4. Section 5 concludes.

2 Literature Review

The first experimental investigation of Samuelson's (1938) revealed preferences theory of consumer demand was conducted by Sippel (1997). Basically, the experiment involved two parts and consisted of several different treatments. In the beginning, subjects were informed that they would have to wait for 60 minutes in the second part of the experiment. While waiting, they would be allowed to use only goods acquired in the first part of the experiment, such as magazines, soft drinks or snacks. In the first part, subjects were endowed with a budget of experimental currency units and then were asked to place their orders for these goods given 10 different possible price vectors. Before the second part began, a random lottery determined each subject's personal price vector and the orderings were were actually carried out.

The experiment was run at the University of Bonn's experimental lab with law and economics students. In the basic treatment 92% of subjects violated SARP. Even allowing for piecewise-linear preferences, still 42% violated GARP. Similar numbers were obtained for the other treatments. Investigating his data more closely, Sippel (1997) showed that the majority of subjects did not exhibit more than one inconsistency, given a maximum of 45 possible pairwise comparison. Allowing for a small optimizing error of 5% in terms of Afriat's (1973) efficiency index, only 10% of subjects remained inefficient. The drawback of allowing for optimizing error, however, was a big loss in the power of the test against the null hypothesis of purely random choice. Hence, Sippel (1997) concluded that "... the evidence for the utility maximization hypothesis is at best mixed. While there are subjects who appear to be optimising, the majority of them do not" (p. 1442).

Cox (1997) used data from the "token economy experiment" that Battalio et al. (1973) conducted with 38 female patients in a mental hospital. As a distinctive feature of the experiment, labor supply was endogenous. 24 subjects (63%) did not exhibit any GARP violations. Février and Visser (2004) reported an experimental test of GARP using 120 subjects from the general population of Dijon in France. The consistency test was based on five choice situations with regard to food products. 29% of subjects violated GARP. Age and income of the subject did not exert a significant influence on violation GARP while – in one setup – female subjects were performing better than males. Subjects buying for families were less likely to be inconsistent. The more different consumption goods were demanded, the higher the likelihood of GARP violations. Using a modified version of the dictator game, Andreoni and Miller (2002) showed that 158 of 178 (89%) subjects exhibited rationally altruistic behavior. Bruyneel et al. (2006) provided a revealed preferences test of the collective consumption model (Chiappori, 1988), which allows for consumption externalities and public consumption.

Banerjee and Murphy (2007) first classified their subjects (69 undergraduate economics and graduate MBA students at Georgia State University) either as rational or irrational based on a revealed preferences test using consumption bundles of candy bars. Only 53.6% of the subjects were classified as WARP-consistent. In a second step they estimated and compared demand coefficients from the resulting subsamples. While both WARP-consistent and WARP-inconsistent subjects' demand estimates satisfied the law of demand, WARP-consistent demand estimates differed significantly from WARP-inconsistent ones. Applying a probit-regression model to their data Banerjee and Murphy (2007) showed that gender predicted WARP-consistency rather well: being male increased the likelihood of WARP-consistency by more than 24% as compared to being female. Other variables such as cognitive abilities, being a graduate student and time did not exhibit a significant effect.

The most important precursor study for our own work was conducted by Choi et al. (2007b). Using a newly developed graphical computer interface (for details see Choi et al., 2007a) they investigated subjects' individual choice behavior under uncertainty. 93 subjects, undergraduates and staff members at UC Berkeley, had to split their initial endowment of 100 tokens between two Arrow securities, x and y. In the symmetric treatment both securities were equally likely, in the asymmetric treatment x was twice as likely to be selected by "nature" than y. Each subject was presented a series of 50 such decision problems in which only the budget line was randomly varied by the computer. Subject had to use the mouse pointer on the desired allocation in order to choose it.

Only 16 of 93 subjects (17%) were completely free of GARP violations. However, allowing for small pertubations of the budget lines and an Afriat efficiency index of 0.9 dramatically improved the subjects' performance as 75 of them exhibited an index of 0.9 or higher. Given the large number of choices per subjects, the probability that an Afriat index of at least 0.9 is generated by random choice is distinctly less than 1%. Choi et al. (2007b) also fitted a two-parameter utility function to the individual data. The two parameters of interest were the Arrow-Pratt measure of relative risk aversion and a parameter of loss or disappointment aversion (Dekel 1986, Gul 1991). Subjects were highly heterogenous with respect to their risk attitudes and about 70% of them exhibited significant loss aversion parameters, i.e. "kinked" utility functions at the safe portfolio.

In our experiment, we will study subjects choice behavior in a very similar setup. However, as an important extension we will additionally monitor subjects choice processes. More specifically, we will record not only final choices but also which alternatives were considered by a subject before making her choices, and how much attention in terms of time the respective allocations were kept visible on the screen was paid to each alternative considered. Numerous experiments, in cognitive psychology in particular, have tried to monitor subjects' thought processes while making decisions. Johnson et al. (1986) developed a computer software – Mouselab – which enables the researcher to study how subjects collect information by hiding and disclosing information on the computer screen using the mouse pointer.[2] Mouselab has been used by many experimentalists. For example, Payne et al. (1988) and Bettman et al. (1990) experimentally investigated the relationship of cognitive effort with accuracy in choice (see also Payne et al., 1993). Camerer et al. (1993) studied sequential bargaining process for gains and losses. Costa-Gomes et al. (2001) studied subjects' strategic sophistication with respect to information collection and decision making in a series of 18 two-person normal-form games. The games were displayed on subjects' screens as boxes via Mouselab, where payoffs initially were hidden and could be made visible by moving the mouse cursor into the boxes. In two experiments using Mouselab, Gabaix et al. (2006) analyzed the directed cognition model, which is based on partially myopic cost-benefit calculations. In the first experiment, that admitted rational-search solutions, the directed-cognition model outperformed the fully rational model. The second experiment involved highly complex choice problems without computationally tractable solutions. Here, the directed cognition model successfully predicted the pattern of information acquisition observed in the experiment.

[2]An alternative method of monitoring subjects' though processes is eye-tracking (see, for example, Russo and Rosen, 1975; and Russo and Leclerc, 1994). The relatively new discipline of neuroeconomics (see, for example, Glimcher et al., 2008) tries to understand the "neurobiology of decision making".

3 The Experiment

As in Choi et al. (2007) there were two states of the nature $s = \{A, B\}$ and two associated Arrow securities, each yielding a payoff of one "token" of experimental currency in one state and nothing in the other. Each token was converted into £0.20 of UK currency at the end of the experiment. Subjects had to split their initial endowment of 100 tokens between the two Arrow securities. That is, their choices had to satisfy the budget constraint $p_A x_A + p_B x_B = 100$, where p_s denotes the price of putting one token on account s and x_s is the subject's demand for Arrow security s. Note that subjects could only choose nonnegative integer allocations, i.e., $x_s \in \mathbb{N}$, and prices were rounded off to the first decimal place. In order to give a sensible representation of the allocation problem on the computer screen, any allocation satisfying

$$100 - \max\{p_A, p_B\} < p_A x_A + p_B x_B \leq 100$$

was allowed.

Figure 1 clarifies the basic experimental setup for a scenario where $p_A = 1.5$, $p_B = 1$, and the probability of state A is $\pi = 0.5$. The solid line represents the budget constraint with slope $-p_B/p_A = -1.5$. The dashed 45°-line marks all portfolios for which $x_A = x_B$. In its intersection with the budget line, we located the *safe portfolio* ($x_A = x_B = 40$). It is called safe because it involves no risk of getting less than the expected value of the portfolio in either state of nature. The second dashed line gives the expected value

$$EV(x_A, x_B) = \pi x_A + (1 - \pi) x_B = \frac{\pi}{p_A}(100 - p_B x_B) + (1 - \pi) x_B$$

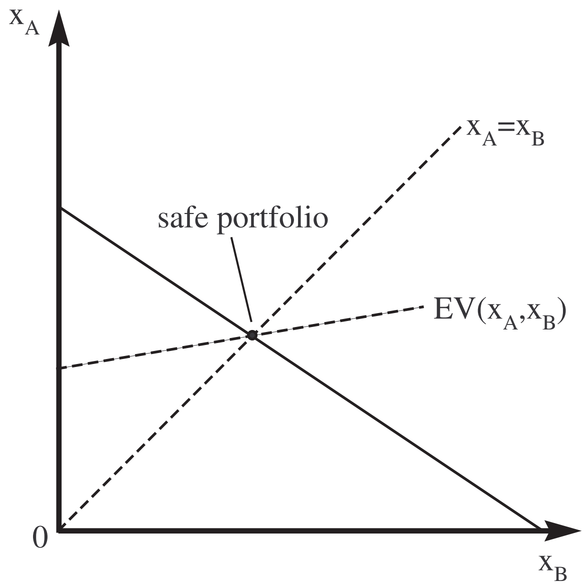

Figure 1: A first-stage choice problem with $p_A = 1.5$, $p_B = 1$, $\pi = 0.5$

of each portfolio as a function of x_B. In figure 1 the slope of the line representing the expected value is the positive fraction 1/6. Hence, portfolios to the left of the safe portfolio are stochastically dominated.

Each subject in the experiment faced 16 rounds of up to three successive choice problems (stages). In the first stage of each round, subjects were graphically presented with a budget constraint $100 = \mathbf{p}^1\mathbf{x}^1$, where $\mathbf{p}^1 = (p_A^1, p_B^1)$ and $\mathbf{x}^1 = (x_A^1, x_B^1)$. The price vector $\mathbf{p}^1$ was taken from the eight-point price set

$$P = \{(1, 1.5), (2, 1), (1, 2.5), (3, 1), (1.5, 2), (2.5, 1.5), (3, 1.5), (2, 3)\}.$$

Furthermore, the probability π of state A being chosen by nature was either 0.5 or 0.67. Note that we used a pseudo-random number generator in order to determine the state of nature.

All subjects were presented the complete set of 16 first-stage choice problems that could be constructed by combining price vectors and probabilities. The order in which they were presented, however, was randomized in order to avoid ordering effects.

Figure 2 shows how each subject's first-stage choice was used to construct the second-stage choice problem. The dashed line represents the first-stage budget line and a dot marks the subject's portfolio choice – e.g., $\mathbf{x}^1 = (22, 67)$ in the figure. The second-stage budget line $\mathbf{p}^2\mathbf{x}^2 = 100$ was determined by first interchanging the two prices, and then replacing the higher new price with a new one chosen at random. Specifically, if the first stage price p_B is lower, as in figure 2, and x_B denotes the first stage allocation to asset B, then the new p_B was randomly chosen from a uniform distribution on the closed interval $[100/x_B, 200/x_B]$, and then rounded to the first decimal place. In the figure, we have $\mathbf{p}^2 = (1, 1.6)$.

Given any sequence $(\mathbf{p}^i, \mathbf{x}^i)_{i=1}^n$ of n pairs of price and quantity vectors satisfying GARP and the normalization $\mathbf{p}^i\mathbf{x}^i = 1$ $(i = 1, \ldots, n)$, Varian (1982, 2006) defines the supporting set $S(\mathbf{p}^{n+1})$ of consumption bundles $\mathbf{x}^{n+1}$ that support any previously unobserved price vector $\mathbf{p}^{n+1}$ so that $\mathbf{x}^{n+1} \in S(\mathbf{p}^{n+1})$ if and only if the extended sequence $(\mathbf{p}^i, \mathbf{x}^i)_{i=1}^{n+1}$ also satisfies GARP and the normalization $\mathbf{p}^i\mathbf{x}^i = 1$ $(i = 1, \ldots, n+1)$. As Varian (1982) notes, the supporting set describes "what choice a consumer will make if his choice is to be consistent with the preferences revealed by his previous behavior" (p. 957).

After the indicated first stage choice, the supporting set for the subject's

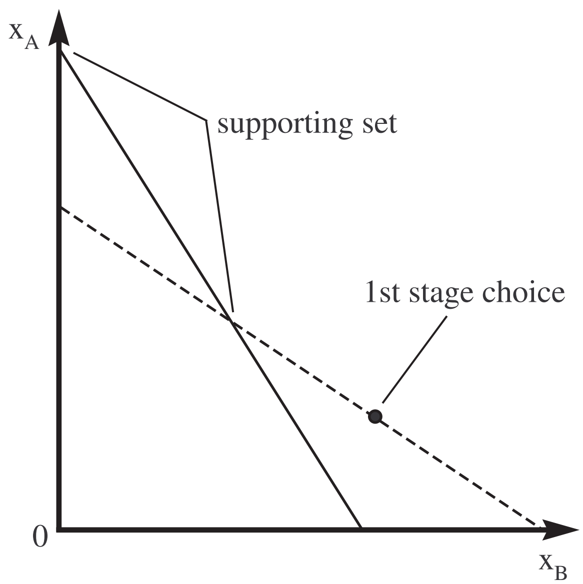

Figure 2: A second-stage choice problem with $p_A = 1$, $p_B = 1.6$, $\pi = 0.5$

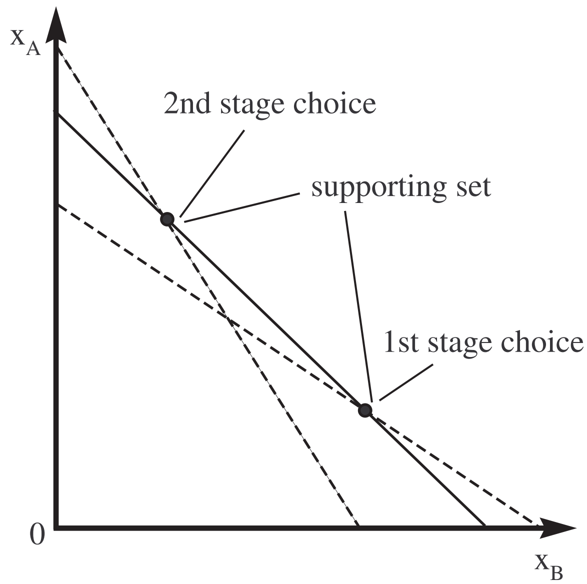

Figure 3: A third-stage choice problem with $p_A = 1.2$, $p_B = 1.1$, $\pi = 0.5$

second stage choice consists of the segment of the budget line between the extreme portfolio with $x_B = 0$ and the intersection of the two budget lines.

If a subject's second-stage choice was from this supporting set, the subject was shown the third-stage problem. Otherwise, the computer program omitted the third stage and proceeded directly to the next of the 16 rounds.

As Figure 3 indicates, the third-stage budget constraint was constructed by first taking the line that passes through the two different allocations that were chosen in the first two stages, then rounding both prices to the first decimal place. For example, assuming that the subject chose $\mathbf{x}^1 = (22, 67)$ at the first stage, followed by $\mathbf{x}^2 = (61, 24)$ at the second stage, at the third stage the price vector would be $\mathbf{p}^3 = (1.2, 1.1)$ as indicated in Figure 3.

Again the supporting set of portfolios can be used to forecast which portfolio a rational subject should chosen given his previous choices. The supporting set is the line segment between the first-stage and the second-stage portfolios.

Figure 4 displays an example screen of the experiment. As soon as a new decision problem appeared, the mouse pointer became visible at its default position at the upper right-hand corner of the screen. When the mouse pointer was moved to a feasible allocation, that allocation was indicated by two numbers and two reference lines marked in red. As long as the mouse pointer was not moved away from this position the information remained visible on the screen, but disappeared as soon as the mouse pointer was moved away from this position. If applicable, the next allocation was displayed.

Subjects could also "fix" and then "release" an allocation by clicking the left mouse button. Once a portfolio was fixed, the numbers and reference lines turned green and stayed visible on the screen until they were released even if the mouse pointer was moved. To choose this portfolio and proceed to the next decision problem, a subject could simply click on an OK button near the lower right-hand corner of the screen.

Some slight time pressure was introduced in order to impose a "cost" on collecting information. The upper right-hand corner of the screen therefore displayed the remaining time out of the original 30 seconds allocated for each choice. When time ran out, if the mouse pointer was over a feasible allocation, or if one had been fixed by an earlier mouse click, that portfolio was recorded as the subject's final choice. Otherwise a missing value was recorded for that choice problem.

An important feature of the experiment is that we recorded not only

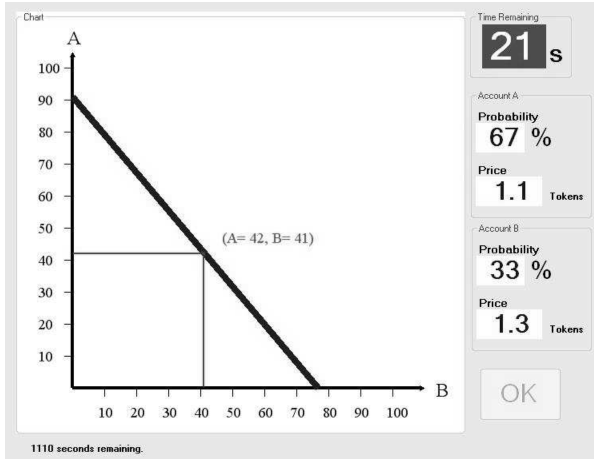

Figure 4: Example screen

subjects' choices but also monitored the process of mouse movement that revealed information about the different portfolios. For each possible allocation near enough to the budget line, we recorded the time (measured in units of about 50 milliseconds) the information remained visible on the screen (except for the time were the allocation was fixed). Later this time data will be used as a proxy for the amount of attention paid to each allocation and to each entire decision problem.

The experiment was conducted on May 20, 2008, in a computer Laborotary at the University of Warwick that had often been used for experiments by other researchers. To avoid bias due to expert knowledge, we recruited 41 non-economics undergraduates (26 male and 15 female students responded to our invitation in time). All had previously agreed to be put included a database of potential recruits for economic laboratory experiments and were contacted by email. Everyone attending and completing the experiment was given £5 of UK currency. Furthermore, subjects were told that one of the choice problems they were going to be presented would be randomly selected for an actual payment at the end of the experiment. The experiment was fully computerized. Since experimental economics and psychology standard software toolboxes such as z-Tree (Fischbacher, 2007) and Mouselab (Johnson et al., 1986) do not offer the graphical displays and the data structure required for our experiment, we programmed the experiment in Visual Basic. First, the subjects were given on-screen instructions (see the Appendix). Then, a training session started where subjects were presented random budget lines and could make choices as often as they wanted. In order to start the experiment, subject had to click a button. After the button was clicked a short countdown was initiated and then the first-stage choice problem of the first round was displayed on the screen. After the last choice, the computer determined the payoff of the subject and the subject was paid off in cash. In total, we paid a total of £461.20 to the subjects, which works out to £11.25 for the average participant.

4 Results

4.1 Data structure

For every subject we recorded up to 48 (16 rounds $\times$ 3 stages) choices, and for every single choice problem a vector of time units allocated to the up to 101 different possible portfolios. Figure 5 shows the data structure for one particular third-stage decision. Along the horizontal axis of the figure we measure the x_B component of the portfolio choice problem. Along the vertical axis we measure the decision time allocated to each option, measured in units of 50 milliseconds. Each small dot represents a feasible allocation

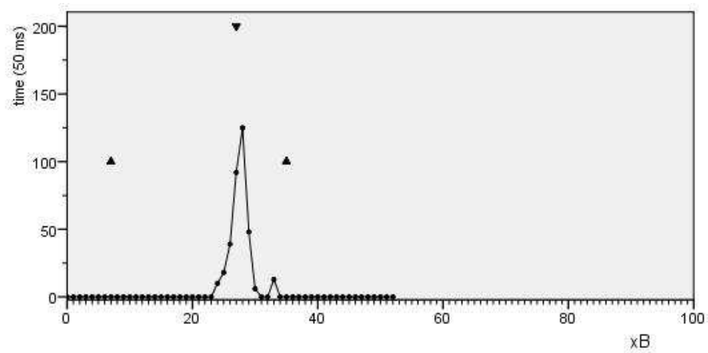

Figure 5: Data structure of a third-stage decision

(there are 53 in the example). The height of the each dot indicates the time that the respective allocation remained visible on the screen. The two upward triangles mark off the decisions in the first two stages corresponding to left and right bounds of the supporting set for the third stage. The downward triangle indicates the chosen portfolio in the third stage. In this example, the subject's attention was focussed only on allocations within the supporting set. Moreover, the selected portfolio was the one which attracted the second largest measured attention time. We should emphasize that most diagrams of attention and three decisions were nowhere near as "well behaved" as the one displayed here.

4.2 Testing the Predictive Power of the Supporting Set

Table 1 gives an overview of our results. Subjects were faced with an identical set of 16 first-stage choice problems, though in different random orders. No subject breached the time constraint of 30 seconds in any choice problem. Hence, in principle, there could have been 16 second-stage choices. However, in cases where subjects chose dominated portfolios such as the dominated extreme portfolio, where the whole budget is allocated to the more costly security, it was not possible to present a sensible second-stage choice problem, because the respective budget line would have had to be very steep (or flat). Our software, therefore, was programmed to abort second-stage choice problems requiring a slope of the budget line greater than 10 (or smaller than 0.1) in absolute value. On average, there were 1.4 such instances per subject, leaving us with 14.6 second-stage choices per subject (91% of the possible 16). About 55% of all second-stage choices (8.1 per subject) were GARP consistent. When the second-stage choice was GARP consistent, a third-stage choice problem was constructed according to the procedure explained in Section 3. Of these third-stage choices, Table 1 reports that 78% were GARP consistent – that is, correctly predicted by the supporting set.

In order to test for substantive rationality at the individual level, we need to consider disaggregated data. After all, the fact that about 55% of all second-stage choices were GARP consistent does not say very much about how consistent any individual might have been. Table 2 lists each subject's ID in the experiment (a six-digit random number), followed by statistics concerning their performance in the second and third-stage choice problems. Columns 2–5 specify respectively the total number of second-stage choices I_2,

Table 1: Consistency of Choices—Aggregated Data

	Outset	Stage		
		1st	2nd	3rd
maximum number of choices	16	16	16	16
actual number of choices	16	14.6	8.1	6.3
actual number (%)	100	91	51	39
% of previous column	—	91	55	78

then the number z_2 and proportion z_2/I_3 of GARP consistent second-stage choices, followed by the significance level $p(z_2)$ of the test for substantive rationality that will be explained below. The last three columns state the same data for the third-stage choices, except that I_3 can be omitted because it is equal to z_2.

Columns 5 and 8 of Table 2 are computed to allow a separate non-parametric exact test for each subject, based on all possible permutations of choice patterns. Let the discretized budget set of the second-round choice problem $i \in \{1, \ldots, I\}$ have K_i elements, of which k_i would satisfy GARP if chosen. Then, $\kappa_i = k_i/K_i$ represents the probability of "accidently" choosing a portfolio that satisfies GARP if a subject randomly picks a portfolio from the budget set. Given a set Γ of I second-stage choice problems (with a maximum of 16), there are 2^I (65536 for $I = 16$) different choice patterns in terms of GARP compliance. Let $\mathfrak{I}$ denote the subset of the I choice problems for which the subject chooses in accordance with GARP. Each choice pattern $\gamma = 1, \ldots, 2^I$ exhibits a probability of $p_\gamma = \prod_{i \in \mathfrak{I}} \kappa_i \times \prod_{i \notin \mathfrak{I}} (1 - \kappa_i)$ to occur randomly. Γ denotes the set of choices patterns, and $\Gamma(\lambda)$ the subset of elements in Γ that shows exactly λ GARP consistent and $(I - \lambda)$ GARP inconsistent choices. Accordingly, the probability of a subject exhibiting exactly λ GARP consistent choices is given by $p_\lambda = p(\lambda = z) = \sum_{\gamma \in \Gamma(\lambda)} p_\gamma$. Finally, the desired significance level of the test for substantive rationality is given by downwards cumulating the p_λ's: $p = p(z) = p(\lambda \geq z) = \sum_{\lambda=z}^{I} p_\lambda$. Note that, since the number of choice pattern is an integer, $p(z)$ in general does not correspond exactly to the conventional significance levels, say 5%. In the following we will therefore look for the smallest possible number $\tilde{z}$ of GARP consistent choices for which $p(\tilde{z}) \leq 0.10$ and call the test significant if the subject exhibits a $z \geq \tilde{z}$.

The test for substantive rationality with regard to third-stage choices was performed on the same lines. However, here the probability or significance level of z_3 has a different interpretation. Remember that third-stage choice problems were shown to subjects only if the respective second-stage choices were GARP consistent. Accordingly, the test gives the conditional probability of choosing in accordance with GARP in the third round if the previous second-stage choice was already GARP consistent.

First, we comment on the second-stage choices. Consider the subject with ID 428534, for example. The subject was presented 15 (of 16 maximum possible) second-stage choice problems. 13 (87%) of her portfolio choices were correctly predicted by the set of supporting bundles for GARP. The individual probability of attaining 13 out of 15 possible successes for this subject is given by 2.4%. Hence our tests rejects the null that the 13 GARP consistent choices were just accidently predicted correctly by the supporting set.

Table 2: Consistency of Choices—Individual Data

Subject	2nd-stage Choices				3rd-stage Choices		
ID	total	GARP consistent		test*	GARP consistent		test*
	I_2	z_2 (I_3)	z_2/I_2	$p(z_2)$	z_3	z_3/I_3	$p(z_3)$
154677	16	8	0.50	0.829	7	0.88	0.007*
198010	16	15	0.94	0.001*	11	0.73	0.030*
208580	14	3	0.21	0.997	3	1.00	0.031*
248676	16	9	0.56	0.538	7	0.78	0.016*
287574	15	7	0.47	0.975	6	0.86	0.043*
304457	16	16	1.00	0.000*	15	0.94	0.000*
328351	15	0	0.00	1.000	—	—	—
329989	13	13	1.00	0.052*	12	0.92	0.000*
366624	15	15	1.00	0.004*	13	0.87	0.054*
379623	16	10	0.63	0.419	9	0.90	0.002*
380900	16	1	0.06	1.000	0	0.00	1.000
412456	12	8	0.67	0.284	5	0.43	0.085*
428534	15	13	0.87	0.024*	12	0.92	0.007*
475061	12	5	0.42	0.801	4	0.80	0.067*
480548	16	2	0.13	1.000	1	0.50	0.130
552212	16	15	0.94	0.002*	11	0.73	0.079*
559071	15	5	0.33	0.975	3	0.60	0.102
563761	13	4	0.31	0.980	2	0.50	0.454
572987	10	1	0.10	0.999	0	0.00	1.000
592468	14	2	0.14	1.000	0	0.00	1.000
624411	16	16	1.00	0.003*	13	0.81	0.267

Table continues.

Continuation of Table 2

Subject	2nd-stage Choices				3rd-stage Choices		
ID	total	GARP consistent		test*	GARP consistent		test*
	I_2	z_2 (I_3)	z_2/I_2	$p(z_2)$	z_3	z_3/I_3	$p(z_3)$
630727	16	7	0.44	0.884	4	0.57	0.080*
666116	13	6	0.46	0.821	4	0.67	0.091*
672588	14	11	0.79	0.134	9	0.82	0.064*
676223	16	13	0.81	0.492	10	0.77	0.106
678750	16	16	1.00	0.001*	14	0.88	0.032*
704318	14	4	0.29	0.988	2	0.50	0.440
734947	12	5	0.42	0.918	3	0.60	0.067*
749595	15	5	0.33	0.987	4	0.80	0.111
750692	15	2	0.13	1.000	0	0.00	1.000
773633	16	11	0.69	0.270	10	0.91	0.000*
799339	15	10	0.67	0.198	9	0.90	0.001*
810682	15	8	0.53	0.693	3	0.38	0.770
816321	16	15	0.94	0.003*	13	0.87	0.001*
881617	11	3	0.27	0.986	2	0.67	0.406
889377	15	10	0.67	0.248	6	0.60	0.065*
923531	16	12	0.75	0.114	11	0.92	0.002*
951284	14	7	0.50	0.724	5	0.71	0.042*
977496	13	8	0.62	0.517	8	1.00	0.001*
990536	15	11	0.73	0.223	5	0.45	0.691
992434	14	2	0.14	1.000	1	0.50	0.563

**Significance level of a non-parametric exact test based on all possible permutations of choice patterns. The null hypothesis is random prediction.*

From the perspective of GARP, the overall results are downright frustrating. Only 9 of 41 subjects (22%) are classified as substantively rational decision makers by our permutation test. This figure is distinctly lower as in previous studies. Note, however, that these studies tested GARP directly using a sequence of choices. Significance levels were then computed by tolerating small changes of the chosen portfolios or budget lines. In contrast to this, our test on the predictive power of the supporting set is based on a sequence of two-stage choice problems of which a certain number has to be fulfilled *exactly*. Nevertheless, the degree of rationality achieved by our subjects is remarkably low. Partly this may be attributed to the fact, that our subjects were non-economics undergraduates. However, it might also point to some degree of confusion among the subjects.

The third-stage choices can partly compensate for the bad second-stage performance. 25 of 40 subjects (one subject did not reach any third-stage task) – that is 62.5% — were classified as being substantively rational by our test conditioned on their rational second-stage choices. Except for one subject, all subjects who performed well in the second stage were successful in the third stage too. Hence, being GARP consistent seemed to be to a great extent decision-task specific.

Table 3 explores the choice data for gender effects. The share of GARP consistent choices was significantly greater for male than for female subjects both on the second and the third stage. Likewise the mean rejection probability of substantive rationality as reported in Table 2 was much higher for female subjects. What was the reason for the relatively bad performance of our female subjects? The last three columns of the table report the rela-

Table 3: Consistency of Individual Choices—Gender Differences

	Gender				Significance
	female		male		level
	mean	std. err.	mean	std. err.	
	share of GARP consistent choices				
2nd stage	0.359	(0.071)	0.668	(0.052)	0.001*
3rd stage	0.474	(0.078)	0.764	(0.046)	0.001*
	rejection probability of substantive rationality				
2nd stage	0.820	(0.083)	0.399	(0.075)	0.001*
3rd stage	0.448	(0.089)	0.110	(0.047)	0.005*
	share of dominated portfolios				
1st stage	0.324	(0.049)	0.153	(0.032)	0.004*
2nd stage	0.354	(0.056)	0.138	(0.039)	0.002*
3rd stage	0.267	(0.087)	0.078	(0.029)	0.056*

**Significant at the 10%-level. Two-tailed independent-sample t test (checked for equality of variances).* $n_{female} = 16$, $n_{male} = 25$. *Except for* d_3: $n_{female} = 15$.

tive share of dominated portfolios chosen. In all three stages female subjects chose about three times more dominated portfolios than their male counterparts. Choosing a dominated portfolio involved a loss of potential payoff for the subjects. So, one might suspect some confusion among a subgroup of female subjects. Remember that our students were non-economic undergraduates. Perhaps, these results reflect the fact that male subjects brought along more outside-university experience with investment problems and the type of graphical computer display we used in our experiment.

4.3 Testing Homogeneity of Preferences

Two of the price vectors to be used for constructing the first-round choice were given by $(1, 1.5)$ and $(2, 3)$, respectively. The choice data for these decision tasks provide us with a simple test of homogeneity of the subjects' preferences. In order to perform the test, we computed the ratio $x_{AB} := x_A/x_B$ for each decision task. In cases were $x_B = 0$, we set $x_{AB} = 100$. Homogeneity of preferences implies that $x_{AB}(1, 1.5) \stackrel{!}{=} x_{AB}(2, 3)$. We have only two such pairs of observations per subject, one for $\pi = 0.5$ and one for $\pi = 0.67$, which will be analyzed separately. The following normative test was conducted in order to decide whether homogeneity of preferences applies or not: It is rather unlikely, though not impossible, that a subject chooses exactly the same x_{AB} in both situations (for a given π). Hence, we used the ratio $Z := \frac{x_{AB}(1,1.5)-x_{AB}(2,3)}{x_{AB}(1,1.5)} \times 100$ as a test statistic. If $Z \leq 5$ ($Z > 5$), i.e. if the deviation is less (more) than 5% towards either direction, then the test does not reject (does reject) homogeneity of preferences.

Table 4 shows the results of testing on homogeneity of preferences. As

Table 4: Homogeneity of Preferences

		Mean		Test*	
	n	$x_{AB}(1, 1.5)$	$x_{AB}(2, 3)$	m	%
		$\pi = 0.5$			
all	41	8.50	12.22	6	15
female	16	4.06	7.24	1	6
male	25	11.34	15.41	5	20
		$\pi = 0.67$			
all	41	19.72	7.77	2	5
female	16	11.66	3.97	0	0
male	25	24.88	10.21	2	8

**Number (percentage) of subjects for which the mean deviation is less than absolute 5%, i.e. homogeneity of preferences cannot be rejected.*

noted before, we split the table according to π. A glance at the table shows that we cannot reject homogeneity for only 6 subjects (15%) in $\pi = 0.5$ and 2 subjects (5%) in $\pi = 0.67$. The table also lists the mean ratios of tokens put to account A and B, respectively. With $\pi = 0.5$ the average token-ratio increased in favor of account A if prices were doubled. In contrast to this with $\pi = 0.67$ the already very high token-ratio decreased. We checked for statistical significance of these effects and also for gender effects. However, due to relatively large standard errors, none of the tests was significant and we therefore omit the test details.

4.4 Attention and the Nature of the Decision Task

The average time a subject focussed his or her attention on the different possible allocation until a decision was made (*dectim*) amounted to 138.36 time units, i.e. about 7 seconds (see Table 5). There were no gender effects concerning average decision time. Considering only the 16 first-stage choices which were equal to all subjects reveals that subjects became faster from round to round (see Figure 6).

In order to find out whether the decrease of decision time was significant and what influenced the time needed until a decision was made, we ran the panel regression reported in Table 6. *dectim* entered the regression as the endogenous variable. The following exogenous variables were tested for their impact on decision time: $safepo$ denotes the value of the safe portfolio. Table 5 shows that the average value of the safe portfolio was 26.51 tokens (5.30 pounds). Our hypothesis is that the more money is at stake, the more time is taken by subjects to make up their minds. Female subjects on average were

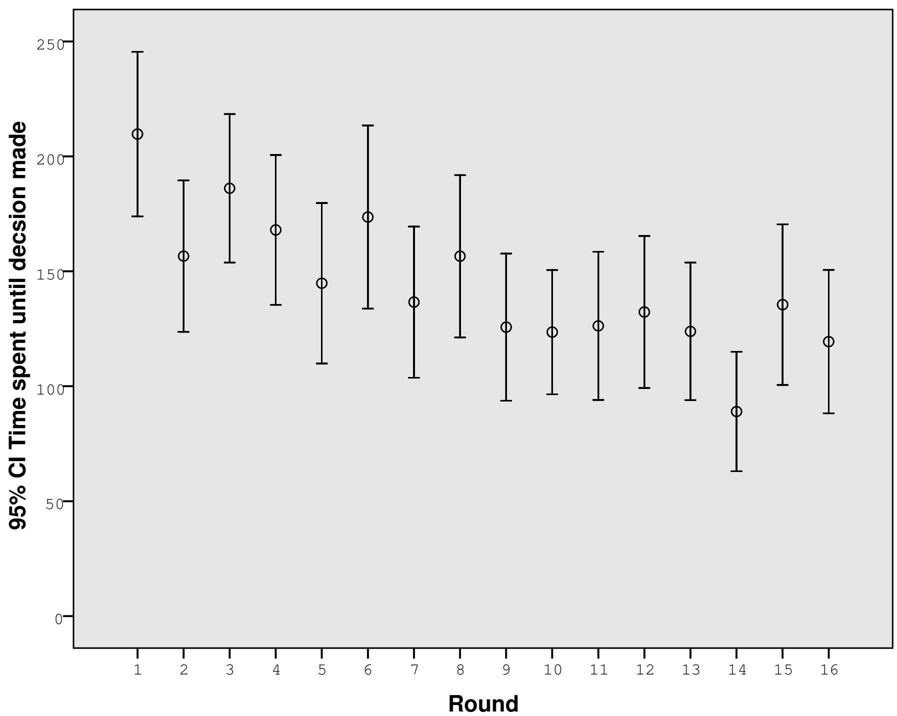

Figure 6: Decision time

presented less valuable decision problems (25.27 vs. 27.19 tokens) because they chose significantly more dominated portfolios (see also Table 3 and the discussion in the previous section).[3] *domina* is a dummy variable with *domina* = 1 (*domina* = 0) if a subject chose (did not choose) a dominated variable. According to Table 5, 19% of all chosen portfolios were dominated. Regarding this variable, we do not have a directional hypothesis. Instead, we want to explore whether subjects choosing dominated portfolios on average spent more or less cumulated time.

The *pricer* variable was included in the regression in order to check whether it becomes easier to solve the decision task when the slope of the budget line gets steeper, i.e. either Arrow security becomes more favorable. It is computed as the absolute value of the normalized price ratio $|p_A/p_B - 1|$. Though insignificant at conventional statistical terms, *pricer* turns out to be a bit higher for female subjects (which is partly due to female subjects choosing more dominated portfolios). *extrem* is a dummy variable for (undominated) extreme choices, i.e. choices where a subject allocated all his or her tokens to either Arrow security. Altogether 4 percent of all choices were extreme in this sense. Note that these portfolios exhibit both the highest expected value and the largest possible payoff variance. Hence, these portfolios might be chosen by risk neutral or risk loving subjects. Choosing the extreme portfolio might also be a sensible strategy for "aggressive" subjects

[3] As explained above, the second- and third-stage choice problems were computed from the subjects' decisions in the first-stage and second-stage choice problems, respectively. If a dominated portfolio was selected in the previous stage, the slope of the budget line had to be very steep (or flat) in the subsequent stage in order to construct a supporting set. A more extreme slope of the budget line implied a less valuable safe portfolio.

who want to maximize expected payoff and minimize time effort at the same time. Male subjects on average exhibited more such choices (6% vs. 2%).

The variable $feasib$ states the number of feasible allocations. We expect the time needed to come to a decision to increase with $feasib$. *watche* denotes the number of actually considered allocations and *relsha* the relative share of feasible allocations considered. Both variables are assumed to increase decision time. Furthermore, *choice*, the number of choice problems already completed, and *probab*, a dummy which takes the value zero (one) if $\pi = 0.5$ ($\pi = 0.67$), entered the regressions. As suggested by Figure 6, we hypothesize that decision time decreases with *choice*. Since $\pi = 0.67$ might make it more difficult to compute the expected value of a chosen portfolio, we expect that subjects needed more decision time if $probab = 1$.

Table 6 presents the results of the regressions. We performed separate OLS regression for the whole sample and broken down by gender. For the latter, we report both pooled sample and fixed-effects panel regressions. Note that we used White's heteroscedasticity robust variance-covariance matrix in order to compute the standard errors. As can be taken from the table, the fit of the regressions is satisfactorily high.

As suggested by Figure 6, experience in terms of the number of choices already completed (*choice*) had a negative impact on decision time. For female subjects this effect was more pronounced. Whether the quickening was due to learning effects or boredom will be investigated in the next section. *probab*, the dummy for $\pi = 0.67$, did not exhibit a significant effect on decision time. If at all, there was a slight tendency among the female subjects of perceiving $\pi = 0.67$ decision problems as being more difficult to solve.

Table 5: Descriptive Statistics

	All Subjects		Female		Male		Test
	mean	std. err.	mean	std. err.	mean	std. err.	t-value
				all choices			
n	1588		546		1024		
dectim (ms)	138.36	(2.65)	142.41	(4.33)	136.14	(3.35)	1.145
$safepo$ (token)	26.51	(0.22)	25.27	(0.35)	27.19	(0.28)	-4.29***
domina (%)	19.08	(0.99)	31.21	(1.95)	12.40	(1.03)	8.52***
extrem (%)	4.22	(0.05)	1.60	(0.53)	5.66	(0.72)	-4.55***
pricer	1.35	(0.05)	1.45	(0.05)	1.27	(0.08)	1.379
$feasib$ (#)	42.73	(0.46)	39.70	(0.70)	44.39	(0.59)	-5.12***
watche (#)	7.54	(0.14)	7.53	(0.22)	7.55	(0.18)	-0.04
relsha (%)	19.54	(3.46)	20.36	(0.54)	19.09	(0.45)	1.817*
				second-stage choices			
n	598		225		373		
chocin (%)	55.85	(2.03)	36.89	(3.22)	67.29	(2.43)	-7.583***
dectim (ms)	144.57	(4.36)	141.52	(6.92)	146.40	(5.62)	-0.542
timcho (ms)	37.57	(1.72)	38.25	(2.91)	37.16	(2.12)	0.306
relcho (%)	34.72	(1.12)	34.96	(1.83)	34.57	(1.42)	0.169
settim (ms)	68.78	(3.34)	48.33	(4.35)	81.11	(4.55)	-5.204***
relset (%)	51.03	(1.72)	35.88	(2.60)	60.17	(2.14)	7.217***
hindex	0.38	(0.01)	0.39	(0.02)	0.38	(0.01)	-0.861
				third-stage choices			
n	334		83		251		
chocin (%)	76.95	(0.23)	63.86	(5.30)	81.27	(2.50)	-2.977***
dectim (ms)	115.76	(5.42)	128.49	(9.73)	111.55	(6.44)	1.353
timcho (ms)	28.31	(1.71)	30.17	(3.45)	27.70	(1.96)	0.625
relcho (%)	36.40	(1.57)	32.71	(3.20)	37.62	(1.79)	-1.357
settim (ms)	82.85	(4.57)	86.28	(8.90)	81.71	(5.32)	0.431
relset (%)	74.71	(1.87)	65.64	(4.00)	77.71	(2.08)	-2.674***
hindex	0.40	(0.01)	0.38	(0.03)	0.40	(0.01)	-0.695

Table 6: Panel Regression for Decision Time

	All Subjects		Female Subjects				Male Subjects			
	Pooled		Pooled		Fixed Effects		Pooled		Fixed Effects	
	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.
const.	43.97***	(12.26)	46.93**	(22.55)	—		48.18***	(15.21)	—	
choice	-1.01***	(0.18)	-1.55***	(0.34)	-1.43***	(0.30)	-0.82***	(0.21)	-1.06***	(0.18)
probab	1.73	(4.16)	11.11	(7.39)	5.86	(6.31)	-1.76	(5.18)	-1.98	(4.23)
safepo	3.90***	(0.91)	2.81	(1.82)	3.94***	(1.47)	4.30***	(1.16)	4.10***	(0.94)
domina	10.50*	(5.60)	14.06*	(7.79)	-5.78	(7.76)	8.62	(7.82)	4.02	(7.08)
extrem	-22.28**	(8.80)	-55.89**	(27.78)	-41.24*	(23.67)	-15.48	(11.85)	-11.97	(9.92)
pricer	-3.29***	(0.68)	-3.94	(4.00)	-6.95**	(3.06)	-3.28***	(1.22)	-2.17***	(0.71)
feasib	-1.87***	(0.41)	-1.07	(0.95)	-1.47**	(0.72)	-2.19***	(0.54)	-2.02***	(0.43)
watche	2.87***	(0.94)	1.82	(1.62)	0.81	(1.40)	3.38***	(1.05)	3.31***	(0.94)
relsha	356.05***	(37.57)	387.64***	(64.53)	402.12***	(54.84)	340.62***	(43.85)	276.24***	(39.45)
gender	4.67	(4.37)	—		—		—		—	
n	1588		564		564		1024		1024	
$\bar{R}^2$	0.404		0.356		0.484		0.428		0.592	
F	108.49***		35.51***		23.02***		86.88***		45.92***	

Endogenous variable: total time used until decision made (dectim). ***$p \leq 0.01$, **$p \leq 0.05$, *$p \leq 0.10$. *White's heteroscedasticity robust variance-covariance matrix was used to compute standard errors.*

The positive sign of $safepo$ confirms our hypothesis that subjects were responsive to higher incentives in terms of a more valuable safe portfolio by increasing their attention. In the OLS regression the coefficient of 3.9 corresponds to about 0.2 seconds per extra token. Considering the dummy variable for dominated portfolios $domina$ yields an interesting results. $domina$ is positive and significant for the whole sample, but this effect is obviously only due to the female subjects, who needed about 0.7 seconds more time to make up their minds if they chose dominated portfolios. However, the insignificant coefficient (which is even negative) of the fixed-effects regression shows that the effect is entirely subject-specific, i.e. a subgroup of female subjects was generally slower and chose many more dominated portfolios than other subjects. This confirms our hypothesis that some female subjects were confused.

The sign of $extrem$ is negative in all five regression, even though insignificant for male subjects (who made distinctly more extreme portfolio choices). Allocating all tokens to just one account thus speeded up decisions a lot. Female subjects spend about 2 seconds less on a task if they made an extreme choice. Very steep or very flat price ratios ($pricer$) as hypothesized made it easier to reach a decision. Interestingly enough, an increase in the number of feasible allocations ($feasib$), i.e. a larger choice set, decreased decision time. As to be expected, $dectim$ increased with the number of allocations considered (though significantly only for the male subjects). A very strong positive impact on decision time can be announced for the relative share of feasible allocations considered ($relsha$). Investigating a larger share of the budget or information set required more attention.

4.5 Attention and the Supporting Set

In this subsection, we shall analyze the relationship between the subjects' performance in terms of GARP compliance of their choices and attention in terms of decision time. As before, we will separately study GARP compliance for second- and third-stage choices, because third-stage choices are conditioned by having been "successful" in the second-stage choice. We define a dummy variable *chocin*, where $chocin = 1$ ($chocin = 0$) means that the respective choice was (not) predicted by the supporting set of consumption bundles. Since *chocin* is a zero-one variable, we will use a logit-regression model.

Table 5 shows that about 56% of all second-stage choices were consistent with GARP and that there was a strong gender effect (see also Table 3). The table also reports the variables that are assumed to have an impact on GARP compliance. In second-stage choices the average decision time (*dectim*) was 144 ms and there was no gender difference. Hence, decision time alone cannot be made responsible for the observed gender difference with respect to GARP compliance. Likewise, there were no gender differences concerning the absolute amount (*timcho*) and the relative share (*reltim*), respectively, of time allocated to the chosen alternative. On average, subjects held the finally chosen alternative visible on the screen for 38 ms before fixing it. This means that the chosen alternative received about 35% of total attention (*dectim*).

The next two variables indicate how focussed a subject was on allocations contained in the supporting set. *settim* is the aggregated time in milliseconds GARP consistent allocations were held visible on the screen. *relset* is the

Table 7: Logit Regression—Second-stage Choices

	All Subjects		Female Subjects				Male Subjects			
	Pooled		Pooled		Fixed Effects		Pooled		Fixed Effects	
	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.
const.	-4.966***	(1.138)	-2.990	(1.896)	—		-4.535***	(1.466)	—	
gender	0.857***	(0.261)	—		—		—		—	
round	0.063**	(0.028)	0.129**	(0.051)	0.178***	(0.067)	0.034	(0.036)	0.032	(0.042)
probab	0.008	(0.015)	-0.023	(0.027)	-0.040	(0.032)	0.013	(0.020)	-0.012	(0.024)
dectim	0.004	(0.002)	0.001	(0.004)	-0.000	(0.005)	0.005*	(0.003)	0.003	(0.004)
timcho	-0.000	(0.005)	0.000	(0.008)	0.006	(0.011)	0.000	(0.007)	0.003	(0.010)
relcho	-0.005	(0.011)	-0.018	(0.017)	-0.016	(0.024)	0.002	(0.015)	-0.010	(0.019)
settim	-0.005	(0.004)	-0.006	(0.006)	-0.006	(0.008)	-0.005	(0.005)	-0.003	(0.006)
relset	0.062***	(0.007)	0.063***	(0.012)	0.067***	(0.016)	0.062***	(0.009)	0.063***	(0.012)
hindex	2.309**	(0.979)	2.593*	(1.568)	-0.917	(2.097)	2.269*	(1.311)	2.575	(1.687)
n	598		225		225		373		373	
McFadden R^2	0.509		0.471		—		0.491		—	
AIC	0.708		0.776		0.432		0.691		0.380	

Endogenous dummy variable: choice in supporting set (chocin); 'No'= 0, 'Yes'= 1. ***$p \leq 0.01$, **$p \leq 0.05$, *$p \leq 0.10$.

Table 8: Logit Regression—Third-stage Choices

	All Subjects		Female Subjects				Male Subjects			
	Pooled		Pooled		Fixed Effects		Pooled		Fixed Effects	
	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.	coeff.	std.err.
const.	-4.657***	(1.718)	-8.674**	(3.791)	—		-2.979	(2.096)	—	
gender	0.781*	(0.426)	—		—		—		—	
round	0.110	(0.044)	-0.095	(0.094)	0.053	(0.104)	0.045	(0.054)	0.032	(0.064)
probab	0.017	(0.024)	0.076	(0.051)	0.084	(0.061)	0.001	(0.030)	0.000	(0.037)
dectim	0.003	(0.005)	0.008	(0.011)	0.010	(0.013)	0.003	(0.006)	0.005	(0.006)
timcho	0.007	(0.009)	0.027	(0.024)	0.007	(0.022)	0.001	(0.011)	-0.006	(0.013)
relcho	-0.023	(0.018)	-0.050	(0.038)	-0.033	(0.044)	-0.008	(0.022)	0.014	(0.028)
settim	-0.010*	(0.006)	-0.023	(0.015)	-0.019	(0.016)	-0.008	(0.068)	-0.005	(0.008)
relset	0.074***	(0.011)	0.097***	(0.030)	0.078**	(0.032)	0.072***	(0.013)	0.068***	(0.016)
hindex	1.716	(1.496)	4.065	(2.574)	2.844	(3.302)	0.121	(1.930)	-1.660	(2.453)
n	334		83		83		373		373	
McFadden R^2	0.508		0.507		—		0.491		—	
AIC	0.591		0.861		0.541		0.691		0.306	

Endogenous dummy variable: choice in supporting set (chocin); 'No'= 0, 'Yes'= 1. ***$p \leq 0.01$, **$p \leq 0.05$, *$p \leq 0.10$.

share of time allocated to GARP consistent allocations with respect to total decision time. Female subjects spent less attention on GARP consistent allocations both in absolute and relative terms. Finally, *hindex* is the Herfindahl concentration index of decision time. For $hindex = 1$, a subject focussed all her attention only on one feasible alternative; for $hindex = 1/feasib$, a subject allocated his or her attention equally to all alternatives.

Our working hypothesis is that all these variables exert a positive influence on GARP compliance. That is, it is assumed to be more likely that a choice is rationalizable if a subject dedicates more time to his or her decision and focusses most of his or her attention to allocations contained in the supporting set and the finally chosen allocation.

In the regression presented in the previous subsection, *choice* – the number of choices already completed – had a significant negative impact on decision time. This could point to learning effects (subjects needed less time to complete a choice for a given quality of choices) or boredom (subjects reduced both decision time and the quality of choices). In order to control for learning effects, we included *round* ($= 1, \ldots, 16$) as an exogenous variable. Furthermore, we again included *probab* in the regression (which turned out to be insignificant for decision time).

Table 7 contains the results of running the logit regression for the second-stage choices. As in the previous linear regression, we present the results not only for the pooled sample but also broken down by gender. For the latter regressions we report both pooled and fixed-effects estimators.

The pooled regression for the whole sample of 598 second-stage choices exhibits, as expected, a significant positive gender dummy. Remember that *choice* decreased decision time for all subjects (see Table 6), where the effect was a bit more pronounced for female subjects. Here, *round* significantly increased the quality of the choices, but only for female subjects. Hence, we can conclude, that routine or learning effects played a role in both groups but in the group of female subjects cumulating "experience" had a much stronger effect on decision speed and quality of choices. It might be not too farfetched to assume that after some additional training sessions the gender difference would disappear.

Among the other variables only *relset* and *hindex* were significant. The larger the share of time spent keeping allocations belonging to the supporting set visible, the likelier that a GARP-consistent allocation was finally chosen. It is surprising though that neither the absolute amount of time nor the time allocated to the chosen allocation had a significant effect. As it seems, all subjects had their own pace in dealing with their decision tasks. They focussed on a relatively small subset of feasible allocations (about 20% as indicated by the *relsha* variable in Table 5) that contained the final decision (but which was not necessarily the allocation receiving most attention).

Table 8 reports the results of the logit regressions for the third-stage choices. The same set of exogenous variables entered the regressions. Note, however, as indicated by the bottom part of Table 5 that there were some important differences. 77% of all third-stage choices were GARP consistent and the gender effect was less pronounced though still significant. Subjects who reached the third stage of a decision round were on average much quicker (116 vs. 145 ms), but they paid the same share of attention to the chosen alternative. However, we already know from the second-stage regressions that *relset* – the share of time allocated to supporting set alternatives – was decisive for the quality of choices rather than decision time. This is also reflected in Table 5, where *relset* increased from 51% to 71% (and the gender difference shrunk).

Table 8 shows that there were no further learning effects among those who had already been rational in the second-stage choices. Gender is still significant at the 10% level. Apart from these variables only *relset* is significant.

5 Conclusion

In this paper, we presented an experimental analysis of subjects' individual information collection and choice behavior in a series of 16 three-stage portfolio-selection problems. An important feature of the experiment is that we recorded not only subjects' choices but also monitored the process of mouse movement that revealed information about the different portfolios. For each possible allocation near enough to the budget line, we recorded the time the information remained visible on the screen. This time data was used as a proxy for the amount of attention paid to each allocation and to each entire decision problem.

Only 22% of subjects were classified as substantively rational decision makers in terms of revealed-preference theory (GARP). However, 62.5% of all subjects who made a rational choice on the second stage of a portfolio-selection problem turned out to be rational on the third stage as well. Hence, we concluded that GARP consistency seems to be to a great extent decision-task specific. The share of GARP consistent choices was significantly greater for male than for female subjects both on the second and the third stage. In all three stages female subjects chose about three times more dominated portfolios than their male counterparts. These results may reflect that male subjects brought along more outside-university experience with investment problems and the type of graphical computer display we used in our experiment, while some female subjects got confused. A regression analysis of the number of choices already completed on the time that was needed to execute the current decision problem supported this presumption as learning effects were more pronounced among female subjects. In general, subjects concentrated their attention on a small subset of allocations from which they usually made their final choices. We ran logit regressions of several variables with potential influence on GARP-consistency of the final choices. Here, our main results was that the larger the share of time spent keeping allocations belonging to the supporting set visible, the likelier that a GARP-consistent allocation was finally chosen.

Two results are really striking: First, that subjects focussed their attention only on a small subset of consequences and, second, that the conditional probability of being GARP-consistent (on the third-stage choice) if the previous choice was GARP-consistent was very high. In our opinion, these results suggest that the observed errors or deviations from substantive rationality come from the fact that people deliberately choose to restrict the amount of information that is used to make up their minds to a "satisfactory" minimum. One could call this kind of behavior "rationally bounded" because people still have a preference for consequences (and exhibit maximizing behavior). Sometimes it leads to substantively rational behavior in the strict sense; sometimes people start from the "wrong" or "too narrow" subset of information and therefore miss the global utility maximum. Obviously, this is different from bounded rationality, where it is assumed that people have a preference for rules (and exhibit satisficing behavior). However, our current experiment is not able to (and was not designed to) statistically differentiate between the two bounded decision models. One might think of heuristics that would give rise to similar results. Hence, further research, both theoretical (developing a model of rationally bounded choice behavior) and experimental (testing and differentiating between the models), is needed.

To summarize, we think that a direct test of bounded decision models requires to monitor evidence regarding subjects thought processes. We believe that "...satisficing behavior should occur, not within a given decision model [(a heuristic)], but in how much detail to include within the model" (Hammond, 2007, p. 185). In other words, such an approach would assume people to exhibit maximizing behavior within a deliberately simplified decision problem. Suppose that an agent uses some kind of rationally bounded model, perhaps after suitable advice. Then in principle one can infer revealed preferences concerning the modeled consequences of whatever decisions the agent seriously contemplated.

Economics, including behavioral economics, focuses on the decisions economic agents actually make. Psychologists tend to focus on the mental states agents may have during a decision making process. We hope that our study shows that the same experiment can yield useful data about both the decision made, and the mental process leading up to that decision. Moreover, the extra data one obtains from this kind of experiment comes rather cheaply (unlike neuro-economics). The main additional cost is programming time. Why not always try to observe as much as possible of the thinking process in any economic experiment?

References

Abelson, R.P., and Levi, A. (1985): "Decision Making and Decision Theory", in: G. Lindzey and E. Aronson (eds.): *Handbook in Social Psychology: Theory and Method*, 3rd ed., Vol. 1, 231–309, Random House, New York.

Afriat, S.N. (1973): "On a System of Inequalities in Demand Analysis: An Extension of the Classical Method", *International Economic Review* 14, 460–472.

Allais, M. (1953): "Le Comportement de l'Homme Rationel devant le Risque, Critique des Postulates et Axiomes de l'Ecole Americaine", *Econometrica* 21, 503–546.

Andreoni, J., and Miller, J. (2002): "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism", *Econometrica* 70, 737–753.

Battalio, R.C., Kagel, J.H., Winkler, R.C., Fisher, E., Jr., Basmann, R.L., and Krasner, L. (1973): "A Test of Consumer Demand Theory Using Observations of Individual Consumption Purchases", *Western Economic Journal* 2, 411–428.

Banerjee, S., and Murhpy, J.H. (2007): *Do Rational Demand Estimates Differ From Irrational Ones? Evidence from an Induced Budget Experiment*, Working Paper.

Bettman, J.R., Johnson, E.J., and Payne, J.W. (1990): "A Componential Analysis of Effort in Choice", *Organizational Behavior and Human Decision Processes* 45, 111–139.

Bruyneel, S., Cherchye, L., and de Rock, B. (2006): *Modeling Collective Rationality: A Nonparametric Test on Experimental Data*, SSRN Working Paper Series.

Camerer, C., Johnson, E.J., Rymon, T., and Sen, S. (1993): "Cognition and Framing in Sequential Bargaining for Gains and Losses", in: K. Binmore, A.P. Kirman, and P. Tani (eds.): *Frontiers of Game Theory*, MIT Press, Cambridge, Mass., 27–47.

Carbone, E., and Hey, J. D. (1994): "Estimation of Expected Utility and Non-Expected Utility Preference Functionals Using Complete Ranking Data", in: B. Munier, and M.J. Machina (eds.): *Models and Experiments on Risk and Rationality*, Kluwer, Boston, 119–39.

Carbone, E., and Hey, J. D. (1995): "A Comparison of the Estimates of EU and non-EU Preference Functionals Using Data from Pairwise Choice and Complete Ranking Experiments", *Geneva Papers on Risk and Insurance Theory* 20, 111–133.

Chiappori, P.-A. (1988): "Rational Household Labor Supply", *Econometrica* 56, 63–89.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007a): "Revealing Preferences Graphically: An Old Method Gets a New Tool Kit", *American Economic Review* 97, 153–158.

Choi, S., Fisman, R., Gale, D., and Kariv, S. (2007b): "Consistency and Heterogeneity of Individual Behavior under Uncertainty", *American Economic Review* 97, 1858–1876.

Costa-Gomes, M, Crawford, V., and Broseta, B. (2001): "Cognition and Behavior in Normal-form Games: An Experimental Study", *Econometrica* 69, 1193–1235.

Cox, J.C. (1997): "On Testing the Utility Hypothesis", *Economic Journal* 107, 1054–1078.

Dekel, E. (1986): "An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom", *Journal of Economic Theory* 40, 304–318.

Elster, J. (1979): *Ulysses and the Sirens: Studies in Rationality and Irrationality*, Cambridge University Press, Cambridge.

Février, P.F., and Visser, M. (2004): "A Study of Consumer Behavior Using Laboratory Data', *Experimental Economics* 7, 93–114.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-made Economic Experiments", *Experimental Economics* 10, 171–178.

Friedman, M., and Savage, J. (1948): "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy* 56, 279–304.

Gabaix, X., Laibson, D., Moloche, G., Weinberg, S. (2006): "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model", *American Economic Review* 96, 1043–1068.

Glimcher, P.W., Camerer, C.F., Fehr, E., Poldrack, R.A. (2008): *Neuroeconomics: Decision Making and the Brain*, Elsevier/Academic Press, Amsterdam.

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

Hammond, P.J. (1988): "Consequentialism and the Independence Axiom", in: B.R. Munier (ed.): *Risk, Decision and Rationality*, Reidel Publishing Company, Dordrecht, 503–516.

Hammond, P.J. (1989): "Consistent Plans, Consequentialism, and Expected Utility", *Econometrica* 57, 1445–1449.

Hammond, P.J. (2007): "Schumpeterian Innovation in Modelling Decisions, Games, and Economic Behaviour", History of Economic Ideas 15, 179-195.

Harless, D.W., and Camerer, C.F. (1994): "The Predictive Utility of Generalized Expected Utility Theories", *Econometrica* 62, 1251–1289.

Hey, J.D., and Orme, C. (1994): "Investigating Generalizations of Expected Utility Theory Using Experimental Data", *Econometrica* 62, 1291–1326.

Hey, J.D. (1999): “Do People (Want to) Plan?”, *Scottish Journal of Political Economy* 52, 122–138.

Hogarth. R. M. (1987): *Judgement and Choice*, 2nd ed., Wiley, Chichester.

Johnson, E.J., Payne, J.W., Schkade, D.A, and Bettman, J.R. (1986): *Monitoring Information Processing and Decisions: The Mouselab System*. Unpublished Manuscript, Center for Decision Studies, Fuqua School of Business, Duke University.

Machina, M.J. (1987): Choice under Uncertainty: Problems Solved and Unsolved’, *Economic Perspectives* 1, 121–154.

Morone, A., and Schmidt, U. (2008): “An Experimental Investigation of Alternatives to Expected Utility Using Pricing Data”, *Economics Bulletin* 4, 1–12.

Payne, J.W., Bettman, J.R, and Johnson, E.J. (1988): “Adaptive Strategy Selection in Decision Making”, *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 534–552.

Payne, J.W., Bettman, J.R, and Johnson, E.J. (1993): *The Adaptive Decision Maker*. Cambridge University Press, Cambridge, U.K.

Russo, J.E., and Leclerc, F. (1994): “An Eye-fixation Analysis of Choice Process for Consumer Nondurables”, *Journal of Consumer Research* 21, 274–290.

Russo, J.E., and Rosen, L.D. (1975): “An Eye-fixation Analysis of Multialternative Choice”, *Memory and Cognition* 3, 267–276.

Samuelson, P.A. (1938): "A Note on the Pure Theory of Consumer's Behavior", *Economica* 5, 61–71.

Sen, A.K. (1995): "Rationality and Social Choice", *American Economic Review* 85, 1–24

Simon, H.A. (1955): "A Behavioral Model of Rational Choice", *Quarterly Journal of Economics* 69, 99–118.

Simon, H.A. (1966): *New Science of Management Decisions*, Harper, New York.

Simon, H.A. (1976): "From Substantive to Procedural Rationality", in: S.J. Latsis (ed.): *Method and Appraisal in Economics*, Cambridge University Press, Cambridge.

Simon, H.A. (1978): "Rationality as Process and as Product of Thought", *American Economic Review, Papers and Proceedings* 68, 1–16.

Sippel, R. (1997): "An Experiment on the Pure Theory of Consumer's Behavior", *Economic Journal* 107, 1431–1444.

Tversky, A., and Kahneman, D. (1974): "Judgment Under Uncertainty: Heuristics and Biases", *Science* 185, 1124–1131.

Varian, H.R. (1982): "The Nonparametric Approach to Demand Analysis", *Econometrica* 50, 945–973.

Varian, H.R. (2006): "Revealed preference", in: M. Szenberg, L. Ramrattan, A.A. Gottesman (eds.): *Samuelsonian Economics and the Twenty-first Century*, Oxford University Press, Oxford, 99–115.

Appendix

Instructions

Experimental Instructions (Please, read carefully)

This is an experiment in decision-making. The entire experiment should be complete within about 30 minutes. Research foundations have provided funds for conducting this research. Please, pay careful attention to the instructions as a considerable amount of money is at stake. At the end of the experiment, you will be paid privately. Your payoffs will depend partly on your decisions and partly on chance, but not on the decisions of the other participants in the experiment. You will receive 5 pounds as a participation fee. In addition you will receive a payment whose calculation will be explained in the following.

During the experiment, we will speak in terms of experimental "tokens" instead of pounds. At the end of the experiment your payoff will be calculated in tokens and translated into pounds. The exchange rate between tokens and pounds is stated on a note at your workplace.

In this experiment, you will participate in at most 48 decision problems that share a common form. This text describes in detail the process that will be repeated in all decision problems and the computer program that you will use to make your decisions.

In each decision problem, you will be asked to allocate an initial endowment of 100 tokens between two accounts labeled A and B. The A account corresponds to the vertical and the B account to the horizontal axis in a two-dimensional graph. Each choice will involve choosing with the mouse pointer a point on a blue line representing possible token allocations. In each choice, *you may choose any A and B pair that is on the blue line.*

Each decision problem will start by having the computer select such a line randomly, where each line permits a minimum of 10 and a maximum of 100 tokens on each account. The "prices" for the two accounts are stated on the right side of the screen. An example: the blue line runs from 50 on the vertical axis (account A) to 33 on the horizontal axis (account B). Hence, the price for allocating a token to account A is two tokens, and for a token on account B you have to give up three tokens of your initial endowment.

You have exactly 30 seconds for choosing one point on the blue line. The time remaining is stated on the screen. Furthermore, you will receive an acoustic signal during the last five seconds.

To choose an allocation, use the mouse to move the pointer over the blue line. You will be shown the token allocations that belong to the respective points on the blue line. Once you have found the allocation that you like best, click with the left mouse button somewhere on the screen, and the most recent allocation will be fixed. If you want to revise your decision, click the left mouse button again and the line will be released. If you are satisfied with your decision, click the "OK" button with the mouse pointer.

As noted above, you can choose only allocations that are located on the blue line. You have 30 seconds for each choice. If you run out of time before you fixed an allocation, the computer will automatically move on to the next decision problem. If you did not touch the blue line at least once within the 30 seconds in order to display an allocation, the computer will record that you did not make a decision; if you displayed an allocation but did not fix it by mouse click, the computer will record the most recent allocation as your *choice. You cannot revise your decision after having clicked the "OK" button or the 30 seconds have elapsed.*

Afterwards you are asked for your next decision. At the end you will be informed that the experiment has ended and the computer determines you payoff.

Your payoff is determined as follows: at the end of the experiment the computer will randomly select one decision round. It is equally likely that any round will be chosen. Afterwards the computer will decide whether account A or B will be paid off. The probability of an account to be selected is stated on the screen for each decision problem. The probability is either 50:50 or 67:33. Pay attention to the probabilities shown on the screen while making your choice. At the beginning of each decision problem, the probabilities briefly flash up in red color. Be careful: if the computer selects a decision task in which you did not make a choice, your payoff will be zero.

Your payoff in tokens, your choice, and the account that has been selected, will be shown in a popup window. Please, let our assistant know that you have finished.

Your participation in the experiment, your choices, and your payoff will be kept confidential. Only on the payoff receipt will we have to record your name. In order to keep your privacy you should not talk to anyone about the experiment and your choices (at least until the complete experiment has ended). We would like to ask you not to talk during the experiment and to remain silent until the end of the last round.

If you are ready for a trial run, click the "OK" button. If there are open questions, please, contact one of our assistants.