Reasoning and Institutions: Do Markets Facilitate Logical Reasoning in the Wason Selection Task?*

David V. Budescu[a], Boris Maciejovsky[b]

February 2003

Abstract

A vast literature shows that individuals frequently fail to identify the normative solutions in logical reasoning tasks. Much attention has been devoted to the study of these deviations at the individual level; less effort was exerted to investigate whether institutional settings might facilitate and improve reasoning. In this paper we address this question by embedding the Wason selection task in a competitive market: each of the four cards of the task was traded over multiple periods in anonymous continuous double auctions, and with real financial incentives. The results of two experiments involving 28 markets, with eight subjects each, indicate that errors in logical reasoning persist, and are present in a wide variety of trading variables, such as prices, volume and liquidity. The market's behavior reflects the normatively correct outcome only when a substantial number of traders know the correct solution.

Keywords: Wason selection task, deductive reasoning, experimental markets, biases in reasoning, double auction

---

* We are grateful for financial support by the Austrian National Bank (Jubiläumsfonds) under Project no. 9293. The second author would like to thank the Society for Judgment and Decision Making (Jane Beattie Memorial Award). We are also indebted to Håkan Fink, Ben Greiner, and Kati Müller for their valuable research assistance. The authors have contributed equally to this paper, and the order of authorship is alphabetically.

a University of Illinois at Urbana-Champaign, Department of Psychology, 603 E. Daniel Street, Champaign, IL 61820, USA. Tel.: 217/333 6758, Fax: 217/244 5876, E-mail: dbudescu@s.psych.uiuc.edu

b Max Planck Institute for Research into Economic Systems, Strategic Interaction Group, Kahlaische Str. 10, D-07745 Jena, Germany. Tel.: +49/3641/686 626, Fax: +49/3641/686 623, E-mail: maciejovsky@mpiew-jena.mpg.de

1 Introduction

Over the last decades, a vast literature has emerged indicating that individuals frequently violate basic normative principles of rational and logical behavior. These studies involve principles of judgment under uncertainty (e.g. Gilovich et al. 2002), rational choice (e.g. Kahneman and Tversky 2000) and logical reasoning (e.g. Oaksford and Chater 1998). One prominent example is the failure to apply the modus tollens in logical reasoning in the Wason selection task, which is widely regarded as "the single most investigated problem in the psychology of reasoning" (Evans and Over 1996, p.356). While considerable effort has been devoted to investigate reasoning at the individual level, little is known about the impact of institutional settings. In this study we address this question by embedding the Wason task in a competitive market.

Markets allow participants to engage in exchange relations, mostly with respect to physical goods and property rights. More importantly, they also allow for dissemination of information. Every bid and ask submitted, as well as every trade executed, reveals valuable information to all the participants in the market. Also the total trading volume and the market's liquidity serve as pivotal sources of information. One example of a market with which most readers are familiar is the labor market. Workers in a given industry, even those who do not necessarily seek new positions, can learn to evaluate their (and their co-workers') skills and "worth" by observing and analyzing cues provided by the employment market. They can observe how many new jobs are being created, advertised, or cut; what are the salaries that are requested by, and offered to, new workers in their organization; how often people change employment; how fast they are promoted, etc. This information can lead them to re-assess their employment situation and take appropriate action, i.e. seek a new position, renegotiate their salary, retire, etc.

The primary research question we address in this paper is whether the information flow experienced by subjects in a competitive market facilitates logical reasoning in the Wason selection task. More specifically, do individuals overcome the well documented biases in reasoning when they are given the opportunity to observe the problem solving strategies of their fellow market participants by learning about trades, buying and selling offers, trading volume, and liquidity of these strategies in a competitive market?

The paper proceeds with a discussion of previous empirical evidence on the persistence of violations of normative principles in markets and with an introduction to the Wason selection task as a means to study logical reasoning. Section 2 covers two experimental studies, and in section 3 we summarize and discuss our main findings.

1.1 Individual biases and the market

Persistence of individual biases in judgment, reasoning and choice has long been of interest to researchers in psychology and economics, but the two disciplines have diametrically-opposed views. Psychologists embrace the view that these systematic fallacies and biases that plague individuals are propagated and affect behavior also at the aggregate level. Some economists object to this view and expect markets to wipe out these individual biases, just as averaging cancels out random noise. Economists use several arguments to justify this prediction: (a) traders on markets may have enough financial incentives and experience to eschew mistakes; (b) a small number of rational traders who control large monetary funds may be sufficient to drive market prices to the rational prediction; (c) non-rational traders may also implicitly learn from the actions of rational ones, or may solicit advice by buying information from market specialists, such as investment brokers. Eventually, non-rational market participants would be driven out of the market by running bankrupt (for a discussion see Camerer 1987). Consequently, when biased individuals compete in a market, the aggregate behaviors are expected to converge to the normative prediction.

Markets, and especially those using double auctions[1], have been shown to be very powerful and effective coordinating mechanisms that can generate outcomes, which are in line with the standard economic prediction of individual rationality. For example, Gode and Sunder (1993) as well as Jamal and Sunder (1996) demonstrate that markets populated by non-rational or boundedly rational computer robots converge to rational expectations equilibria, supporting Becker's conjecture that "households may be irrational and yet markets quite rational" (Becker 1962, p.8).

On the other hand, Russell and Thaler (1985) show analytically, under the assumption of no short selling, that the existence of a market is not sufficient to guarantee the elimination of non-rational behavior. Similarly, the models by De Long et al. (1991) as well as by Hirshleifer and Luo (2001) demonstrate that non-rational behavior might, under certain assumptions, be evolutionarily stable. Knez et al. (1985) compare elicited preferences, as indicated by the willingness to pay for, and the willingness to accept a certain risky security, with the rational expectations equilibrium. Their findings suggest that individual responses predicted market behavior more accurately than standard economic theory. Thus, individual non-rationality might translate into, and affect the, aggregate market behavior.

Previous empirical studies on the persistence of biases in markets have focused primarily on subjective probability judgments. Plott and Wilde (1982), Duh and Sunder

1 A participant in a double auction can engage simultaneously in buying and selling activities by submitting bids and asks to the market or by accepting standing bids and asks which have been submitted by other market participants.

(1986) and Camerer (1987) investigated whether asset prices in competitive markets are in line with Bayesian updating or the representativeness heuristic (Kahneman et al. 1982). In Camerer's study (1987), for example, subjects were informed that assets pay state-dependent dividends, and the prior probabilities of the states were represented by bingo cages, with various compositions of red and black balls. Subjects were publicly informed about the prior distributions, and were confronted with a sample of three balls, which was drawn (with replacement) from one of the bingo cages. Based on the priors and the sample, predictions for asset prices can be derived on a variety of competing hypotheses, e.g. Bayesian updating, representativeness, base-rate neglect, etc. The general conclusion from these studies is that prices converge close to the Bayesian prediction, but that the direction of any deviation from this norm can be best described by the representativeness heuristic (Camerer 1995).

Ganguly et al. (2000) explore the base rate fallacy (Bar-Hillel 1980, 1990, Kahneman and Tversky 1973) in asset markets by investigating whether the format of information presentation affects the subjects' pricing of the assets. They contrast the abstract representation of using balls-and-bingo cages with the method of semantic vignettes used in psychological studies on the representativeness heuristic. The results support the authors' conjecture that asset prices would be more strongly biased when subjects received semantic rather than abstract information.

In a study on the three-door problem (also known as the Monty Hall dilemma), Friedman (1998) showed that while deviations from Bayes' law could be considerably attenuated by providing subjects with ample opportunities to learn, the anomaly did not completely disappear even after many repetitions. Recently, Slembeck and Tyran (2002) examined behavior in this context after introducing communication and competition. They showed that each of the two factors reduced deviations from Bayes' law considerably, and when both factors were combined the three-door anomaly was completely eliminated.

To summarize, previous empirical studies on the persistence of biases and cognitive errors at the collective level have focused primarily upon Bayesian updating of probability judgments, and the results are inconclusive. In this study we aim to determine whether other individual non-rationalities persist in competitive markets: Instead of studying subjective probability judgments, we focus on logical reasoning. More specifically, we investigate the Wason selection task (Wason 1966), which is widely regarded as the single most investigated reasoning problem, in a market set-up.

1.2 The Wason selection task

The Wason selection task investigates whether individuals employ the normatively correct strategy of formal logic in testing a conditional rule. In the original setting (Wason 1966), participants are shown cards, each with a letter (vowel or consonant) on one side and a number (even or uneven) on the other side. The participants' task is it to verify a conditional rule of the form "if p, then q", more precisely "if there is a vowel (p) on one side of the card, then there is an even number (q) on the other side", by selecting the minimal number of cards that must be turned over to decide whether the rule is true or false. Four cards are presented to the participants who can only see one side of the card; one showing "E" (p-card), one "K" (not p-card), one "2" (q-card), and one "7" (not q-card). The correct solution involves (a) verifying the truthful implication of the rule by turning over "E" (p-card), and (b) examining possible violations of the rule by turning over "7" (not q-card). However, only a small fraction (typically less than 10%) of the participants selects the two required cards (Griggs and Cox 1983). Most people select the "p-card", but fail to select the "not q-card" (failure to apply modus tollens), and frequently select the "q-card" (affirming of the consequent).

Interestingly, individuals solve the task correctly under some circumstances. Griggs and Cox (1982), for instance, show that performance in the Wason selection task depends on the thematic problem contents, and Pollard and Evans (1987) demonstrate that reasoning depends on content and context effects. Moreover, individuals were found to correctly test rules embedded in deontic contexts (such as permissions, obligations, promises, prohibitions, etc.), where the rule is expressed as a social obligation or a right (e.g., Cheng and Holyoak 1985, Griggs and Cox 1983, Johnson-Laird et al. 1972). An example of a socially applied rule is the "drinking-age" rule (Griggs and Cox 1982); "if a person is drinking beer, then this person must be over 19 years of age". Accordingly, four different cards are presented to test the rule: (i) one card stating that the person is drinking beer (p-card), (ii) one stating the person is drinking Coke (not p-card), (iii) one card indicating that the person is 22 years old (q-card), and (iv) one card indicating that the person is 16 years of age (not q-card). In this context the majority of individuals select the correct pair of "p" and "not q". This indicates that crucial aspects of human reasoning lie in the content of a problem, not in its formal structure (Wason and Johnson-Laird 1972).

The failure to apply the modus tollens and the spurious affirmation of the consequent, have been successfully replicated in many different studies. Interestingly, while considerable effort has been devoted to study reasoning at the individual level and the factors that can affect it, little is known about the impact of institutional settings on reasoning. In this study we address this question by taking Wason's selection task to a competitive market.

More specifically, we investigate whether deviations from normative logical reasoning are mitigated when individuals actively and passively experience information flow in the context of competitive markets. In experiment I, we compare individual behavior in the Wason selection task with market behavior, drawing upon a variety of market characteristics, such as trading prices, trading volume, market liquidity, and asset holdings. We also investigate whether the context of the reasoning setting, which was shown to affect individual behavior, has similar effects on the market behavior. In experiment II, we investigate the robustness of these deviations from the normative solution by increasing the information flow, and by allowing some subjects to obtain "inside" information about the correct solution to the Wason selection task.

Is it reasonable to expect that exposure to markets affects significantly people's behavior? There is strong empirical evidence that individuals who are placed in groups charged to solve collectively intellective tasks (i.e., tasks with a single demonstrable solution, such as arithmetic and algebraic problems), ameliorate their individual performance in post-group tasks (e.g. Laughlin and Ellis 1986, Stasson et al. 1991, 1988). In other words, information gained from exposure to the correct solution in a group setting induces better individual performance in subsequent problems of the same type. However, the Wason selection task was never used in this paradigm. Markets are, for all practical purposes, large groups in which participants exchange information without engaging in bilateral encounters, and are governed by more rigid rules and procedures than freely interacting groups. For instance, trading is usually timely restricted; face-to-face interaction is often prevented, offers might only be improved, etc. Therefore, the present study on logical reasoning in a market context, can also be regarded as an investigation of the generalizability of group effects on intellective tasks to institutional settings.

The two views described earlier have distinct predictions: If the market simply reflects the behavior of its (mostly, biased) participants we would expect that the "p-card" and the "q-card" would be valued higher than the other cards, including the correct, but typically ignored "not q-card". Similarly, the differences between the abstract and deontic presentations would be reflected by the market variables. If, on the other hand, the market mitigates individual biases one would predict that the "p" and "not q" cards, that make up the correct solution set, will be valued more than the others. And, one would predict that the bias towards the "q-card" would be attenuated. Similarly, one would expect that in the market there would not be significant differences between the abstract and the deontic framing of the problem.

2 Method

2.1 Experiment I

2.1.1 Experimental design and procedure

One hundred twenty-eight undergraduate students participated in 16 experimental asset markets, which were conducted at the Max Planck Research Lab in Jena, Germany. Seventy-five females and 53 males, aged 19 to 35 ($M = 22.8$, $SD = 2.45$), participated in the study. Experiment I comprised two stages: in the first stage subjects were asked to solve the Wason selection task individually, whereas in the second stage they participated in asset markets involving groups of eight subjects. We used a 2 x 2 design varying (i) the context of the individual Wason selection task (abstract vs. deontic representation) and (ii) the sequence of types of assets traded in the first (and second) half of the trading periods: In sequence *AD*, subjects traded abstract assets in the first half of the markets and deontic assets in the second half. In sequence *DA*, the order was reversed. Each experimental condition consisted of four sessions with eight subjects, for a total of 32 subjects. Table 1 displays the experimental conditions.

[Table 1 about here.]

For the abstract representation of the Wason task we used the standard target rule: “If there is a vowel on one side of the card, then there is an even number on the other side”. Subjects were confronted with four cards, but only one side of the cards (“A”, “K”, “2” and “7”) was shown. For the deontic version we used the rule: “If a salesperson gets a bonus, he/she must have sold more than 125 units”. The four cards displayed showed “bonus”, “no bonus”, “more than 125 units sold” and “less than 125 units sold”. Subjects were asked to select the minimal number of cards, which must be turned over to test the target rule. If subjects picked the correct cards, a payoff of €4 was granted and paid out (at the end of the experiment).

Subjects completed this short task individually, and then participated in the experimental market. After receiving instructions about the double auction and the market, they completed a short test designed to determine their understanding of the trading mechanism. The experiment started only when all the participants answered correctly all the items of the test. Afterwards, subjects participated in two training markets, to become familiar with the trading mechanism. Each of the training markets lasted four minutes. Finally, subjects participated in the actual experimental market that included 16 trading periods. On the market, participants could trade four assets simultaneously in a continuous anonymous double auction, implemented with the software z-Tree (Fischbacher 1999). The four assets represented the corresponding four cards in the Wason selection task. As indicated in Table 1, subjects traded assets representing the abstract and the deontic version of the Wason task. In the abstract version of the task, the assets I, II, III, and IV represent "A", "K", "2", and "7", and in the deontic version they represent "bonus", "no bonus", "sold more than 125 units", and "sold less than 125 units".

At the beginning of each of the 16 periods, subjects were endowed with 120 Experimental Currency Units[2] (ECU), and with four assets of the same type (i.e. either of type-I, type-II, type-III, or type-IV). Two subjects on each market were endowed with each type of asset. Thus, in each trading period there were a total of eight assets of each type available for trade. Asset endowments were counterbalanced across periods, so that each subject was endowed with the same asset type twice in each of the two types of markets (abstract and deontic), in a pre-specified balanced order.

The subjects' task was to select the minimal number of card(s) to be turned over to test the target rule. Subjects were told that if they believe a certain card is part of the correct solution they should strive to acquire the corresponding asset on the market. The assets that make up the correct solution of the rule paid out dividends of 40 ECU. Assume, for example, that at the end of a trading period a subject holds three assets of type-I, two assets of type-III, and five assets of type-IV. In this case, the subject held three complete sets of the correct assets (type-I and type-IV). Thus the payoff would have been 3 (sets) * 2 (assets) * 40 ECU. The total earnings of our subject in this particular trading period would therefore have been 240 ECU plus his, or her, cash holdings.

Trading periods lasted 180 seconds. Subjects could either submit bids and asks or accept standing bids and asks by other market participants. Only improving offers, i.e. higher bids and lower asks, were allowed. Subjects were not granted credit, nor allowed to sell assets short. After each new entry by one of the participants, the bids and the asks were ranked, and the best offer for potential buyers and sellers was constantly highlighted. Subjects were also provided with a chronological list of all the trades concluded, and were informed about the remaining trading time. Figure 1 displays the schematic screen-shot of the auction.

[Figure 1 about here.]

At the end of each trading period, subjects were informed about their cash and asset holdings. However, they were not informed about their dividend payments. We refrained

2 Ten Experimental Currency Units equal €1.

from providing immediate feedback about dividends because we are interested in seeing if, and how, the market values the solution of the logical reasoning problem, and determine whether the market can overcome individual errors. If we were to disclose the dividend payments, and the correct solution, the market would turn into a simple competition for scarce resources (assets of type-I and type-IV), and would provide no information about our main research questions.

At the conclusion of all 16 trading periods, one of the periods was randomly selected to determine the payoffs. Subjects received their earnings from stage I of the experiment together with their dividend payments and their cash holdings from the randomly selected period of stage II. Experimental sessions lasted about two hours and participants earned an average of €12.31 ($SD =$ €5.37).

2.1.2 Results

Analysis of the individual selection task: Figure 2 displays the frequency of choices of each of the four cards under the two contexts.[3] Clearly the standard results are replicated in both contexts: The most popular solutions selected are either card I alone (26.6% and 18.8% in the abstract and deontic formats), or card I together with card III (37.5% and 18.8%). The correct solution (I and IV) was the third most frequently selected pattern (6% and 14%). All other choice combinations are negligible.

A comparison of the distribution of choices under the two presentations suggests that the context of the task did not matter overall ($\chi^2(3) = 5.74, p > .05$), but it is worth noting that in the deontic representation, card I was identified as part of the solution less often ($\chi^2(1) = 5.95, p < .05$) and card IV more often ($\chi^2(1) = 4.56, p < .05$), than in the abstract version of the task (see Figure 2). Only five subjects picked the correct cards in the abstract treatment, whereas nine subjects identified the correct solution in the deontic representation of the task. A Binomial-test, however, reveals that this difference is not significant.

[Figure 2 about here.]

Analysis of the market variables: We focus on four market variables:

1. Trading prices — the amount (in ECUs) at which a successful trade was completed. These prices simply reflect how the market values the various assets.

2. Trading volume — the number of successful trades.

3 The figure also includes results from experiment II which will be discussed at a later stage.

3. Market liquidity — the sum of bids submitted to the market. This is a measure of how easily and quickly a certain asset can be divested without facing a pronounced market discount.

4. Asset holdings at the conclusion of trading — this measure can be used to infer the subjects' trading objectives.

Since all these measures are affected by the behavior of all eight participants in any given market, the unit of statistical analysis is the market, rather than the individual participant. For the purpose of these analyses the values of the dependent variables were averaged across the eight periods of each market. This decision is justified by (a) the absence of period-by-period feedback and (b) results of preliminary analyses that failed to detect any systematic changes across the eight periods. These means were analyzed in a 4-way mixed ANOVA with two within-subjects factors; context (abstract, deontic assets) and asset type (type-I, II, III, and IV), and two between-subjects factors; context of the individual task prior to the market (abstract, deontic task), and sequence of markets (abstract-deontic = *AD*, deontic-abstract = *DA*). In the analysis of the trading prices we found no significant effects for the two between-subjects factors, and no significant interactions involving these factors, indicating that we were successful in counterbalancing our treatments. The average trading prices in the various within-subject conditions are depicted in Figure 3. There was a significant main effect for context ($F(1;12) = 5.55, p < .05, \eta^2 = .32$) suggesting that the average prices for deontic assets exceeded the prices for abstract assets. There was also a significant difference between the four assets ($F(3;36) = 6.84, p < .05, \eta^2 = .67$). Bonferroni post-hoc tests (at the .05 level) indicate that trading prices for assets of type-I were higher than for assets of type-II and IV.

[Figure 3 about here.]

A similar 4-way ANOVA was performed for the average trading volume. Again, we did not find any significant effects or interaction involving the between-subjects factors. We found only a significant main effect for asset type ($F(3;36) = 9.72, p < .05, \eta^2 = .45$, see also Figure 4). Post-hoc tests (at the .05 level) reveal that the trading volume for assets of type-I exceeded the volume for type-II and type-IV assets.

[Figure 4 about here.]

Next we consider the results of the ANOVA on the sum of bids submitted to the market. Given that subjects could not withdraw bids, this is probably the most direct measure of the subjects' perception of the correct solution. We found a significant main effect for asset type ($F(3;36) = 19.94, p < .05, \eta^2 = .86$, see also Figure 5), as well as a significant two-way interaction between asset type and trading sequence ($F(3;36) = 9.45, p < .05, \eta^2 = .74$), and a three-way interaction between context, asset type and trading sequence ($F(3;36) = 4.33, p < .05, \eta^2 = .57$). Given the three-way interaction, we abstain from interpreting lower order effects. The relevant means for the interaction are given in Table 2. Market liquidity for asset I is higher in the *DA* sequence than in *AD*. When trading asset II in the abstract context, market liquidity is higher in sequence *AD*, whereas if deontic assets are traded, the reverse pattern emerges. Market liquidity for asset III does not appear to vary between sequences and contexts. Finally, market liquidity for asset IV is higher in sequence *AD* than in *DA*, for both versions of the task. However, the difference between *AD* and *DA* is more pronounced for abstract assets.

[Figure 5 about here.]

[Table 2 about here.]

Subjects' valuations of the different assets can also be inferred from a direct examination of the difference between the final asset holdings (at the end of each trading period) and the initial asset endowments. In most cases there was little change in asset holdings. However, the results appear to differ across asset types: for type-I and III, subjects kept their asset holdings in 40.1% and in 40.9% cases, respectively, whereas for type-II and IV the corresponding percentages are 50.4% and 47.8%, respectively. These results are in line with the finding of unequal trading volumes for the four assets. Trading was mainly concentrated on assets of type-I and III (see Figure 4). Higher rates of holding assets simply suggest that it was more difficult to divest assets II and IV (which are not perceived to be in the solution) than it was to trade the remaining two types.

2.1.3 Discussion

In the individual choice task, we replicated the standard results of the classic Wason task; most people turned over the "p-card" and the "q-card" and neglected to examine the diagnostic "not q-card". Therefore, only a small minority of subjects (under 11%) identified the correct solution. Consistent with a vast literature on the topic, we found various indications that performance was slightly better in the deontic representation of the problem, though this difference was not significant.

However the most interesting results pertain to the performance of the market for the various cards: Clearly, the market valued the four assets differentially. And, most remarkably, there was a close correspondence between the market variables and the standard (and faulty) individual reasoning in this task. Trading prices, trading volume, market liquidity, and asset holdings suggest that most subjects considered asset I to be the core correct solution. This asset was traded more often than all the others, attracted more interest than the others, and was traded at the highest price. In motivating the current study, we suggested that the effectiveness of the market institution in this context would be demonstrated by a reversal in the salience and attention paid to the "q-card" (asset III) and the "not q-card" (asset IV). The results are unequivocal; the two key assets (in fact, all four assets) are ranked identically by individuals solving a reasoning problem (in an abstract or a more concrete context), and by all the indices of market activity (prices, volume, and liquidity). Changing the context of the reasoning task from the abstract to the deontic affected trading prices (of asset IV) in the anticipated direction, but did not affect the trading volume and market liquidity.

This striking persistence indicates that cognitive errors that plague individual reasoning persist in competitive markets. Evidently, merely transplanting the reasoning task into a competitive environment that offers clear and strong financial incentives is not sufficient to improve reasoning. One possible explanation is that no institution, no matter how efficient, can create something out of nothing! A necessary condition for the market to perform its magic is the presence of a meaningful number of rational traders who can drive it, through their behavior and their indirect influence on the others, in the right direction. Our markets did not necessarily meet this test: five of the 16 markets included no participant who solved the task correctly in the original presentation, and eight of them included only one participant who solved the problem[4], so clearly there was no "rational core" that could drive the market towards the desired normative prediction! Another possible concern is that our markets were too short (eight periods). In the second experiment we seek to address these issues by investigating a market for assets in the Wason task in which this minority of normative traders is present. To achieve this goal we informed in each market a subset of subjects about the correct solution which will pay dividends. Hence, we will call these subjects "insiders". In addition, we increase the length of the market by doubling the number of trading periods.

4 Each of the remaining three markets was populated by two subjects who solved the task correctly.

2.2 Experiment II

2.2.1 Experimental design and procedure

In experiment II, 96 undergraduate students participated in 12 experimental asset markets, which were also conducted at the Max Planck Research Lab in Jena, Germany. Sixty-two females and 34 males, aged 19 to 52 ($M = 23.31, SD = 4.12$), participated in the study.

Experiment II was identical to the first in most respects. In the first stage subjects were asked to solve the Wason selection task individually, and in the second stage they traded assets (cards) on the market. There were two major changes in the design: (a) In this study we only used the abstract representation of the task. One implication of this change is that each market consists of 16 periods (instead of eight in the original study). (b) In the market stage we introduced and manipulated the number of market insiders (1, 2 and 4), having one baseline treatment (with no insiders). The experimental conditions as well as the baseline treatment consisted of three sessions with eight subjects. Table 3 displays the experimental design in more detail.

[Table 3 about here.]

The insiders were selected randomly. They were explicitly informed about the correct solution to the Wason selection task at the beginning of each trading period. Naturally, they maintained their informed status throughout the experiment. These participants were also informed that not all subjects might have received identical information. Thus, insiders did not know how many informed subjects were on the market.

The experimental sessions lasted about two hours and subjects earned, on average, €15.04 ($SD =$ €6.92).

2.2.2 Results

Analysis of the individual selection task: The distribution of the individual cards turned over in the Wason selection task prior to the market is presented in Figure 2, next to the results from the first study. Clearly the results are highly similar. In fact, they are not significantly different from the choices in the abstract test in the first study ($\chi^2(3) = 2.06, p > .05$). Figure 6 complements this analysis by presenting the distribution of the subset of cards chosen by the various subjects in both studies. Recall that only five out of 64 subjects picked the correct solution in experiment I. The rate of correct solutions is higher in this study (14 out of 96 = 15%), but the difference is not significant ($\chi^2(1) = 1.67, p > .05$).

[Figure 6 about here.]

Analysis of the market variables: We focus on some of the same market variables analyzed in the first experiment: the average trading prices, market volume, and market liquidity within each market, across all 16 periods. In this study we also examine the pattern of trades among informed "insiders" and the remaining "outsiders" in the various conditions. We analyzed the trading prices in a 2-way mixed ANOVA with one within-subjects factor, asset type (I, II, III, and IV), and one between-subjects factor, number of insiders (0, 1, 2, and 4). Figure 7 displays the average trading prices of the four assets in the various groups. We found a significant main effect for asset type ($F(3; 24) = 10.00, p < .05, \eta^2 = .83$), but neither a main effect for the number of insiders nor an interaction effect. The Bonferroni post-hoc tests (at the $p = .05$ level) show that trading prices for asset I were higher than all other assets and prices for asset IV exceeded those of asset II.

[Figure 7 about here.]

We applied a similar 2-way ANOVA, to analyze trading volume (see Figure 8). We found a significant main effect for asset type ($F(3; 24) = 7.47, p < .05, \eta^2 = .79$), but neither a main effect for the number of insiders nor an interaction effect. The post-hoc tests indicate that the trading volume for asset II was lower than the trading volume for the other three assets (all tests $p < .05$).

[Figure 8 about here.]

The ANOVA on the sum of bids submitted to the market replicates this pattern: a significant main effect for asset type ($F(3; 24) = 92.59, p < .05, \eta^2 = .98$), but neither a main effect for the number of insiders nor an interaction effect. The post-hoc tests show that liquidity for asset I was higher than liquidity for the other three assets, and that liquidity for asset II was lowest (all results at $p < .05$).

It is interesting to point out that although the interaction of asset and condition was not significant in either of the three ANOVAs, in the presence of two or four insiders, the "not q-card" (asset IV) was traded at higher prices and at higher volume than the "q-card" (asset III).

Comparison of insiders and outsiders: The impact of insiders on market behavior becomes most evident if one analyzes the nature and frequency of the transactions (and other market activities) involving the informed "insiders" and the remaining "outsiders". Table 4 summarizes the frequency of trades involving assets that are part of the solution (I and IV), as a function of the information status of the seller and of the buyer. The table presents the ratio of observed and expected number of trades[5] for those assets that are part of the correct solution. Thus, ratios above (below) 1 indicate more (less) trades than expected by purely random matching. The last column of the table presents a likelihood ratio G^2-test of the hypothesis that the observed frequencies match the expected ones. All the tests are significant (compared to a χ^2 distribution with three degrees of freedom), but these values should be interpreted with great caution, since the assumption of independence is not met.

First, note that insiders trade correct cards among themselves ($I-I$) more frequently than expected, while outsiders trade these cards among themselves ($O-O$) less frequently than expected. This behavior on part of the insiders makes perfect sense since subjects were initially asymmetrically endowed with assets, and insiders knew that dividends were only paid out when the correct assets are held in equal proportions. Thus, even if a participant holds part of the correct set cards, he/she nevertheless had to engage in trading to acquire the missing asset. Second, note in those transactions involving the two types of subjects, insiders tend to buy correct cards from outsiders at rates exceeding expectation, but the opposite trend (insiders selling correct cards to outsiders) is very rare. Both trends hold for all experimental conditions. Third, the observed/expected ratio of the number of trades diminishes with the number of insiders on the market, except for the case when insiders sell assets to outsiders.

[Table 4 about here.]

Finally, we investigate whether insiders managed to exploit their informational advantage by earning higher dividends than the uninformed subjects. Market insiders accumulated average dividends of 1,123.81 ECU ($SD = 749$), while the corresponding average dividend for uninformed subjects was only 419.20 ECU ($SD = 569$). This difference is statistically significant ($F(1;94) = 21.78, p < .05, \eta^2 = .53$), and we can attribute it directly to their information advantage, rather than superior trading skills. There was no significant difference between informed and uninformed subjects with respect to trading profits, i.e., market insiders did not profit more from selling their assets than the uninformed outsiders.

Do outsiders learn from the actions of the insiders? To examine possible effects of information dissemination on the market, we compared the trading behavior (bids and asks),

5 The expected number of trades is a function of the number of insiders, N_i, and outsiders, N_o, and varies across conditions. For example, if $N_i = 2$ (and $N_o = 6$), and if all pairings are equally likely, we expect $^{30}/_{56}$ of the trades to involve only outsiders, $^{2}/_{56}$ of the trades to take place among insiders, $^{12}/_{56}$ to involve insiders selling assets to outsiders, and $^{12}/_{56}$ to involve insiders buying assets from outsiders.

and the outcomes (dividends earned) of the informed and the uninformed participants in the first half (periods 1-8) and the second half (periods 9-16) of the market more closely. It is extremely difficult to conduct complete and proper statistical analyses since not all types of participants (insiders or outsiders) in the various markets engaged in relevant transactions in the different periods. Therefore, there are many instances of missing data and highly asymmetric numbers in the various sessions. For illustrative purposes only we present in Table 5 the number of bids submitted for assets that are part of the solution (Right) and those that are not in the solution (Wrong) in the two halves of the market only for the group with equal number (four) of insiders and outsiders. Obviously, the uninformed subjects in the treatment groups show no sign of systematic change in the second half of the market following their participation in the first eight periods and their exposure to the behavior of the informed insiders. In fact, their results are undistinguishable from those observed in the baseline treatment (with zero insiders).

[Table 5 about here.]

2.2.3 Discussion

The findings of experiment II are very clear: In the groups with zero or one insider market behavior resembles trading from experiment I. More specifically, subjects valued asset I significantly higher than asset III, followed by assets IV and II. This pattern, however, changes in the treatments with two or four market insiders. Here, trading reveals a market valuation that reflects more closely the normative solution; asset I is still valued highest, however, asset IV (not q-card) is valued more than asset III (q-card).

Our findings indicate that these changes can be attributed to the actions of the informed subjects who manage to exploit their informational advantage by earning higher dividend payments than uninformed subjects. They achieve this by selling to the uninformed outsiders assets that are not part of the solution, and by buying from them those that make up the correct solution. Finally, we found no systematic changes over the course of the experiment in the behavior of the uninformed traders. More specifically, there was no evidence that they learned any characteristics of the correct solution from their interaction with the informed subjects in the market.

3 General discussion

The 2002 Nobel Prize in Economic Sciences was awarded jointly to a psychologist and an experimental economist. The prize citations single out Professor Daniel Kahneman: "for having integrated insights from psychological research into economic science, especially concerning human judgment and decision-making under uncertainty", and credit Professor Vernon L. Smith: "for having established laboratory experiments as a tool in empirical economic analysis, especially in the study of alternative market mechanisms".[6] This paper is a modest effort to lay the foundations of another bridge linking the two disciplines by examining the persistence of one of the most intriguing failures of human cognition in the context of one of the most important settings of experimental economics.

The Wason selection task is prototypical of research on logical reasoning. Its starting point is an abstract representation of an intellective problem that has a single, undisputed, solution. The original work documented the systematic fallacies of unaided human reasoning in this context. Subsequent work has demonstrated that there are ways to overcome, at least partially, this shortcoming of intuitive reasoning. Some researchers have shown the positive effects of transplanting the task into concrete (Griggs and Cox 1982, Pollard and Evans 1987) and deontic contexts (Cheng and Holyoak 1985, Griggs and Cox 1983, Johnson-Laird et al. 1972). Others (Hoch and Tschirgi 1983) have shown a positive correlation between the level of education (and, supposedly, experience with logical reasoning) and performance on this task, indicating that training and experience might improve reasoning.

Previous work has focused exclusively on individuals working in isolation, which is, perhaps, not totally surprising given that the task is viewed as a vehicle to study individual cognitive skills. However, reasoning might benefit from social interactions by acquiring new information and insights, or by learning and simple imitation. It is remarkable that despite the empirical evidence regarding the beneficial effect of having groups solve intellective problems with demonstrable solutions (e.g. Laughlin and Ellis 1986, Stasson et al. 1991), there has been no systematic study of Wason's selection task, the most investigated reasoning problem, in interactive social settings.

In the present paper we studied this selection task in the context of experimental markets that are, essentially, (large) groups in which participants exchange information publicly, while protecting their anonymity, and that are governed by strict rules and procedures. In fact, one could argue that in experimental economic research markets play the role that (small) interacting groups have in social psychology and that the differences between the two reflect the discrepancies between the experimental practices in the two disciplines (for a discussion see Hertwig and Ortmann 2001): Experimental markets enforce a strict script over many replications that offer ample opportunity to learn, and use

6 These quotes are taken from the official web site of the Nobel Foundation: http://www.nobel.se/economics/laureates/2002/presentation-speech.html.

real financial incentives contingent on the participants' performance, but many group decision studies do not implement these features. Next we summarize what we consider the most important implications of our results for psychologists and economists.

We have argued that the study of the market behavior in trading the information pieces for this reasoning task is, essentially, an investigation of the generalizability of group effects on intellective tasks to more rigid institutional settings. We are not aware of any studies of the Wason task in "regular" groups that allow for free and unconstrained communication among members, but we assume that the effects would be no different from other intellective tasks; in most groups where at least one member can identify, explain, and demonstrate the correct solution, the group accepts it, and most of the members who originally failed to solve the problem would endorse it. This pattern was not found in the markets we studied: When some of the participants, in some cases as many as half of them, were informed about the correct solution, they realized the value of the information, traded according to the solution, and increased their individual earnings. However, the implications of the insiders' trading behavior were not properly analyzed and understood by the other participants whose market activity was indistinguishable in all respects from a baseline treatment with no informed traders. It appears that mere exposure to actions of others did not cause them to reconsider their own faulty intuitions. In other words, the information available to everyone on the market was not sufficiently transparent, powerful or persuasive to cause people to identify the correct solution. Clearly, most traders require either full disclosure (e.g. the presence of market insiders), or a direct and complete demonstration of the solution before adopting it (see Tindale 1993). Thus the implication for psychologists who study (individual and group) solutions of intellective problems is that not all forms of social interactions improve performance. More systematic work is needed to pinpoint the essential features that are strictly necessary for the group effect to emerge. One way to address this question is to use the same paradigm and examine behavior on this task in a variety of markets with different rules, trading mechanisms and types of information disclosure. An interesting aspect of the present work is the demonstration, which may surprise many psychologists, that it is possible to document individual biases by studying higher level aggregates of the subjects' behaviors.

Many economists argue that competitive markets wipe out individual biases and irrationalities causing idiosyncrasies to disappear at the aggregate. The results of the 28 markets we investigated indicate otherwise. The individual errors in logical reasoning persist, and can be observed in a wide variety of trading variables, such as prices, volume and liquidity. Only when a substantial number of traders know the correct solution the market variables approach the normatively correct outcome. However, there was no evidence that this was due to the diffusion and sharing of information. Rather, the changes in these market variables reflect primarily the behavior of the informed subjects, who exploited their knowledge, to acquire the correct cards and divest the incorrect ones. There was no support for informational spillovers on the market, allowing uninformed traders to benefit from the actions of the informed. Thus our results suggest that not every trading mechanism can eliminate deeply rooted misconceptions and rectify powerful biases. Future work should determine the limits of the market's capability to mitigate individual biases. One way to pursue this goal is to study various market mechanisms in which the number of useless and non-diagnostic pieces of information that can be traded by the market participants together with relevant cues is systematically varied.

How general are our results? We showed that the typical reasoning errors in Wason's selection task pertain in double auction markets. Instead of interpreting these results as conclusive proof that biases persist in competitive environments, we prefer to consider them a challenge to identify institutional settings that would minimize these errors. Slembeck and Tyran (2002) showed that the rate of errors in the three-door problem could be markedly reduced by introducing competition in the form of rank-dependent payoffs. Subjects in their study were explicitly informed about the resulting payoffs of each of their two possible choices (remain or switch). We refrained from informing our subjects about their (dividend) earnings after each period, because this would amount to revealing the correct solution and would defeat our goal to study whether the market (group) overcomes the bias. Future research should examine alternative market mechanisms that provide more effective feedback, facilitate information flow associated with trading activities between market participants, and improve the quality of their decisions. Results of such studies would benefit economic research on institutional and market design, as well as cognitive and social psychological research on reasoning.

References

Bar-Hillel, M.: 1980, The base-rate fallacy in probability judgments, *Acta Psychologica* **44**, 211–233.

Bar-Hillel, M.: 1990, Back to base rates, *in* R. M. Hogarth (ed.), *Insights in Decision Making: A Tribute to Hillel J. Einhorn*, Chicago University Press, Chicago, pp. 200–216.

Becker, G. S.: 1962, Irrational behavior and economic theory, *Journal of Political Economy* **70**, 1–13.

Camerer, C. F.: 1987, Do biases in probability judgment matter in markets? Experimental evidence, *American Economic Review* **77**, 981–997.

Camerer, C. F.: 1995, Individual decision making, *in* J. H. Kagel and A. E. Roth (eds), *Handbook of Experimental Economics*, Princeton University Press, Princeton, NJ, pp. 587–703.

Cheng, P. W. and Holyoak, K. J.: 1985, Pragmatic reasoning schemas, *Cognitive Psychology* **17**, 391–416.

De Long, J. B., Shleifer, A., Summers, L. H. and Waldmann, R. J.: 1991, The survival of noise traders in financial markets, *Journal of Business* **64**, 1–20.

Duh, R. R. and Sunder, S.: 1986, Incentives, learning and processing of information in a market environment: An examination of the base-rate fallacy, *in* S. Moriarty (ed.), *Laboratory Market Research*, University of Oklahoma, Norman, OK, pp. 50–79.

Evans, J. S. B. T. and Over, D. E.: 1996, Rationality in the selection task: Epistemic utility versus uncertainty reduction, *Psychological Review* **103**, 356–363.

Fischbacher, U.: 1999, Z-tree: Zurich toolbox for readymade economic experiments: Experimenter's manual, *Working Paper* **21**. Institute for Empirical Research in Economics at the University of Zurich.

Friedman, D.: 1998, Monty Hall's three doors: Construction and deconstruction of a choice anomaly, *American Economic Review* **88**, 933–946.

Ganguly, A. R., Kagel, J. H. and Moser, D. V.: 2000, Do asset market prices reflect traders' judgment biases?, *Journal of Risk and Uncertainty* **20**, 219–245.

Gilovich, T., Griffin, D. and Kahneman, D.: 2002, *Heuristics and Biases: The Psychology of Intuitive Judgement*, Cambridge University Press, Cambridge.

Gode, D. K. and Sunder, S.: 1993, Allocative efficiency of markets with zero intelligence traders: Markets as a partial substitute for individual rationality, *Journal of Political Economy* **101**, 119–137.

Griggs, R. A. and Cox, J. R.: 1982, The elusive thematic-materials effect in Wason's selection task, *British Journal of Psychology* **73**, 407–420.

Griggs, R. A. and Cox, J. R.: 1983, The effects of problem content and negation on Wason's selection task, *Quarterly Journal of Experimental Psychology* **35A**, 519–533.

Hertwig, R. and Ortmann, A.: 2001, Experimental practices in economics: A methodological challenge for psychologists?, *Behavioral and Brain Sciences* **24**, 383–402.

Hirshleifer, D. and Luo, G. Y.: 2001, On the survival of overconfident traders in a competitive securities market, *Journal of Financial Markets* **4**, 489–516.

Hoch, S. J. and Tschirgi, J. E.: 1983, Logical knowledge and cue redundancy in deductive reasoning, *Memory and Cognition* **11**, 200–209.

Jamal, K. and Sunder, S.: 1996, Bayesian equilibrium in double auctions populated by biased heuristic traders, *Journal of Economic Behavior and Organization* **31**, 273–291.

Johnson-Laird, P. N., Legrenzi, P. and Legrenzi, M. S.: 1972, Reasoning in a sense of reality, *British Journal of Psychology* **63**, 395–400.

Kahneman, D., Slovic, P. and Tversky, A.: 1982, *Judgment under Uncertainty: Heuristics and Biases*, Cambridge University Press, Cambridge.

Kahneman, D. and Tversky, A.: 1973, On the psychology of prediction, *Psychological Review* **80**, 237–251.

Kahneman, D. and Tversky, A.: 2000, *Choices, Values and Frames*, Cambridge University Press, Cambridge.

Knez, P., Smith, V. L. and Williams, A. W.: 1985, Individual rationality, market rationality, and value estimation, *American Economic Review, Papers and Proceedings* **75**, 397–402.

Laughlin, P. R. and Ellis, A. L.: 1986, Demonstrability and social combination processes on mathematical intellective tasks, *Journal of Experimental Social Psychology* **22**, 177–189.

Oaksford, M. and Chater, N.: 1998, *Rationality in an Uncertain World: Essays on the Cognitive Science of Human Reasoning*, Psychology Press, Hove, Sussex.

Plott, C. R. and Wilde, L.: 1982, Professional diagnosis vs. self-diagnosis: An experimental examination of some special features of markets with uncertainty, *in* V. L. Smith (ed.), *Research in Experimental Economics*, JAI Press, Greenwhich, pp. 3–20.

Pollard, P. and Evans, J. S. B. T.: 1987, Content and context effects in reasoning, *American Journal of Psychology* **100**, 41–60.

Russell, R. and Thaler, R. H.: 1985, The relevance of quasi-rationality in competitive markets, *American Economic Review* **75**, 1071–1082.

Slembeck, T. and Tyran, J.-R.: 2002, Do institutions promote rationality? An experimental study of the three-door anomaly, *Working Paper* **21**. Department of Economics, University of St. Gallen.

Stasson, M. F., Kameda, T., Parks, C. D., Zimmerman, S. K. and Davis, J. H.: 1991, Effects of assigned group consensus requirement on group problem solving and group members' learning, *Social Psychology Quarterly* **54**, 25–35.

Stasson, M. F., Ono, K., Zimmerman, S. K. and Davis, J. H.: 1988, Group consensus processes on cognitive bias tasks: A social decision scheme approach, *Japanese Psychological Research* **30**, 68–77.

Tindale, S. R.: 1993, Decision errors made by individuals and groups, *in* N. J. C. Jr. (ed.), *Individual and Group Decision-Making*, Erlbaum, Hillsdale, NJ, pp. 109–124.

Wason, P. C.: 1966, Reasoning, *in* B. Foss (ed.), *New Horizons in Psychology*, Penguin Books, Harmondsworth, pp. 135–151.

Wason, P. C. and Johnson-Laird, P. N.: 1972, *Psychology of Reasoning: Structure and Content*, Batsford, London.

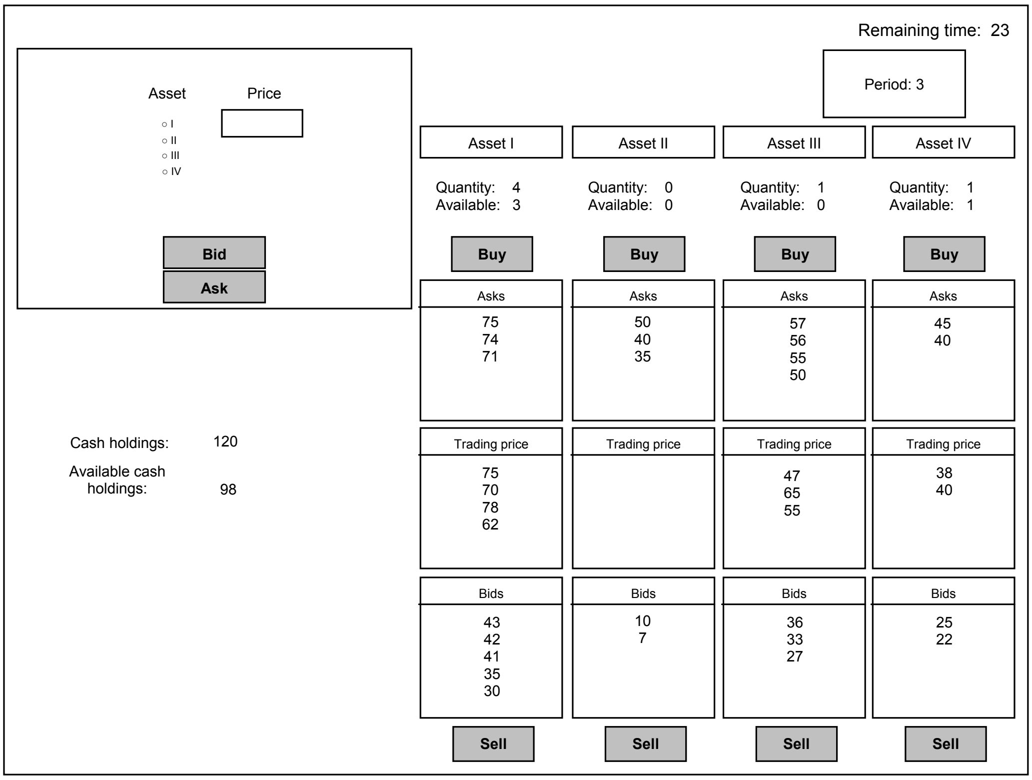


Figure 1: A schematic screen shot of the auction

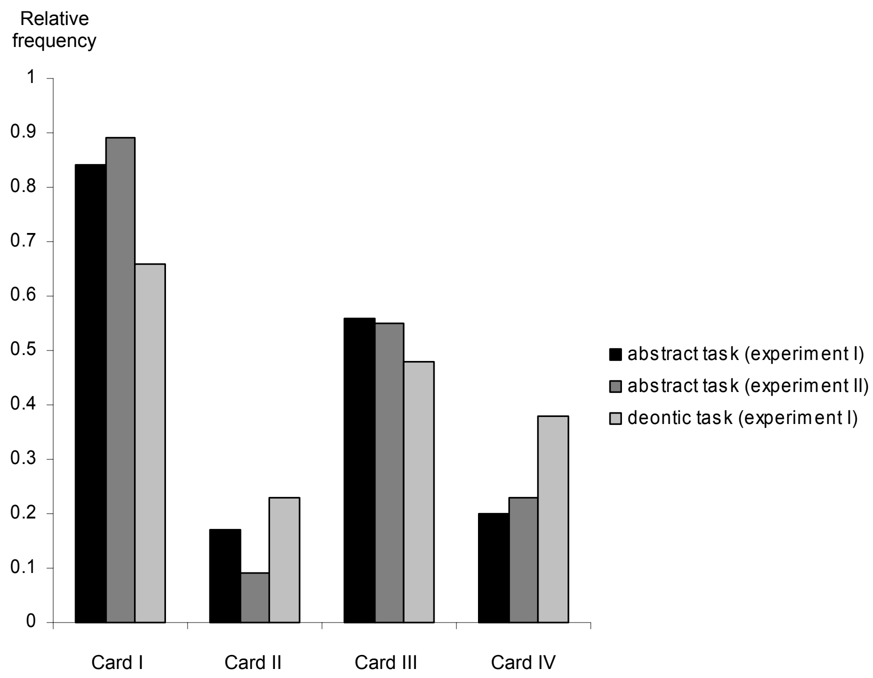


Figure 2: Relative frequency of selected cards in the individual Wason selection task for experiments I and II

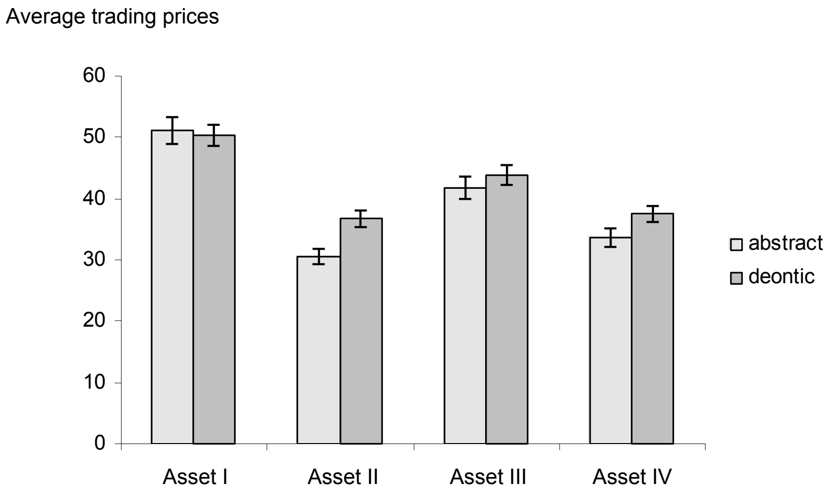


Figure 3: Average trading prices and standard errors in bars

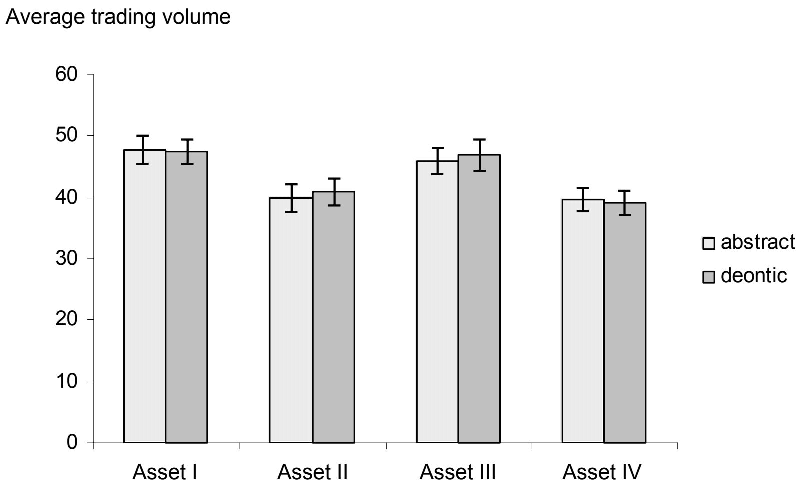


Figure 4: Average trading volume and standard errors in bars

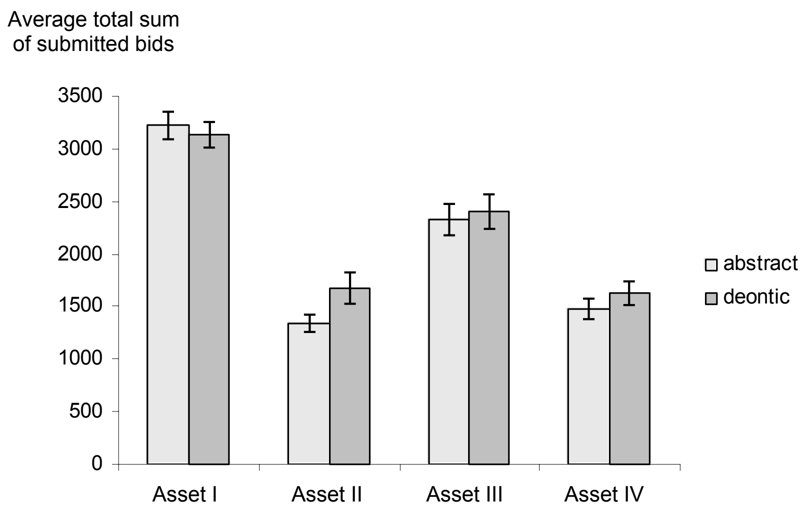
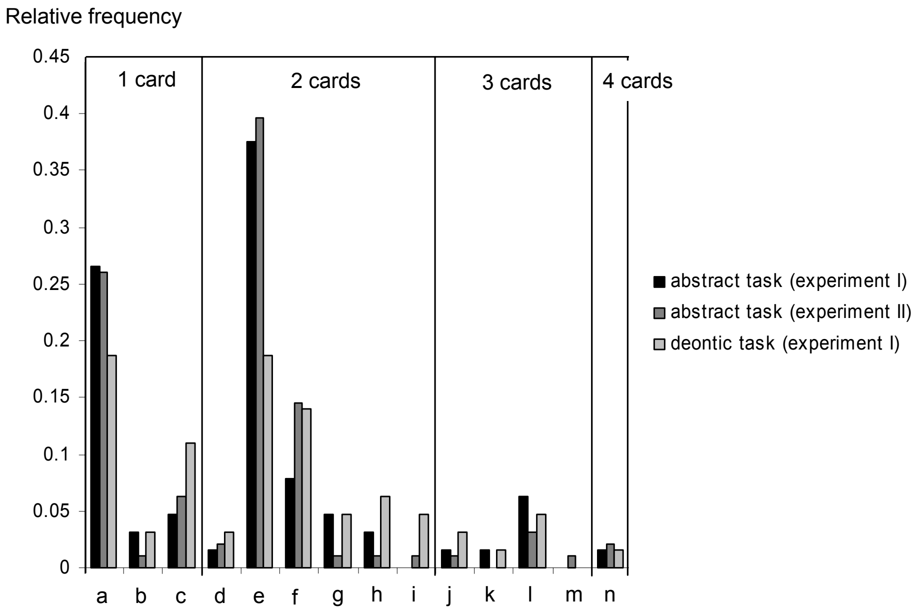


Figure 5: Average total sum of submitted bids and standard errors in bars



a	Card I	f	Cards I, IV *	k	Cards I, II, IV
b	Card II	g	Cards II, III	l	Cards I, III, IV
c	Card III	h	Cards II, IV	m	Cards II, III, IV
d	Cards I, II	i	Cards III, IV	n	Cards I, II, III. IV
e	Cards I, III	j	Cards I, II, III		

* Correct choice

Figure 6: Relative frequency of individual selection patterns in the Wason selection task prior to the market for experiments I and II

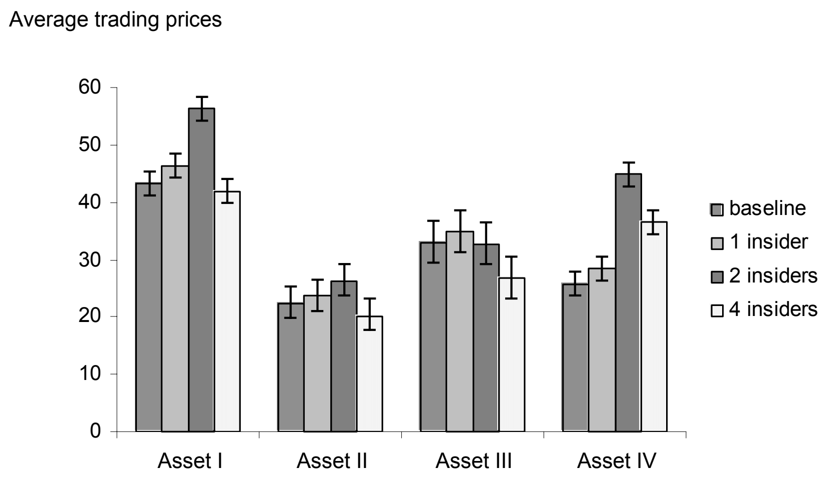


Figure 7: Average trading prices and standard errors in bars

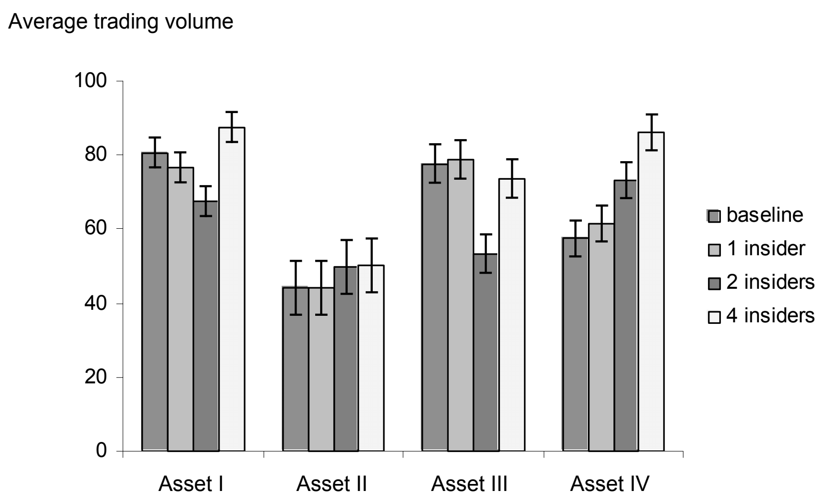


Figure 8: Average trading volume and standard errors in bars

Table 1: Experimental conditions for experiment I

	Market		
Individual task	Periods 1-8	Periods 9-16	Subjects
Abstract	Abstract	Deontic	4 x 8 = 32
Abstract	Deontic	Abstract	4 x 8 = 32
Deontic	Abstract	Deontic	4 x 8 = 32
Deontic	Deontic	Abstract	4 x 8 = 32

Table 2: Mean and standard deviation of the sum of bids submitted to the market with respect to context, asset type and asset sequence

	Abstract		Deontic	
	AD	DA	AD	DA
Type-I	2819.38	3646.13	2830.50	3444.25
	(691.19)	(1193.00)	(777.25)	(1020.67)
Type-II	1674.75	996.50	1492.13	1856.88
	(528.06)	(705.15)	(391.05)	(1479.43)
Type-III	2301.88	2359.88	2362.50	2432.25
	(897.26)	(1223.51)	(726.88)	(1553.25)
Type-IV	1872.25	1087.38	1721.00	1539.25
	(689.15)	(784.95)	(606.75)	(970.41)

Note: Standard deviations are displayed in parentheses.

Table 3: Experimental conditions of experiment II

	Market		
Individual task	No. of insiders	Periods 1-16	Subjects
Abstract	0	Abstract	3 x 8 = 24
Abstract	1	Abstract	3 x 8 = 24
Abstract	2	Abstract	3 x 8 = 24
Abstract	4	Abstract	3 x 8 = 24

Table 4: Proportion of observed to expected trades of assets that correspond to the correct solution

			Seller and buyer pairs				
No. of insiders	No. of outsiders	No. of outsiders	I-I	I-O	O-I	O-O	G^2
1	7	783	-	0.86	0.54	2.30	83.20
2	6	732	2.06	0.78	0.88	1.49	78.90
4	4	892	1.33	0.59	0.77	1.31	56.95

Note: "I" denotes insider and "O" denotes outsider.

Table 5: Number of bids submitted for assets that are part of the solution (R) and those that are not in the solution (W) for the first and second half of trading periods

	Insiders			Outsiders		
	Periods 1-8	Periods 9-16	$\sum$	Periods 1-8	Periods 9-16	$\sum$
R	282	244	526	161	144	305
W	59	45	104	183	163	346
$\sum$	341	289	630	344	307	651