Information Aggregation and Allocative Efficiency in Complex Environments

Thesis by

Christoph Brunner

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2010

(Defended September 9, 2009)

To my parents.

Acknowledgements

Primary thanks goes to my principle advisor, Jacob Goeree, for his unending support and encouragement. I also thank my advisors Charlie Plott and Tom Palfrey as well as my co-authors Charles Holt, Colin Camerer and John Ledyard for their comments and assistance. The discussion with seminar participants at Caltech and at various esa-meetings have helped to improve these papers. The financial support of the Gordon and Betty Moore Foundation is gratefully acknowledged.

Abstract

It has been suggested that information cascades might occasionally prevent asset markets from performing efficiently (e.g., Alevy et al. 1997). We run experiments in which private signals about an asset with a common value are released sequentially. That allows us to compare the quality of information aggregation in periods in which an information cascade would occur in the absence of prices to the quality of information aggregation in other periods. We find no significant difference, but we do find evidence that prices are less likely to converge to the fully revealing rational expectations equilibrium when early signals are misleading.

In a second chapter, we focus on information cascades in sequential games, where subjects choose between two options and each subject has a small chance of being perfectly informed about which option is correct. In treatment "sequence," subjects observe the entire sequence of predecessors' choices, while in treatment "no-sequence" they only observe the number of times each option has been chosen. Subjects tend to follow their immediate predecessor in treatment sequence, which is the optimal strategy under common knowledge of rationality. In treatment no-sequence, fully rational agents would follow the minority of their predecessors (Callander and Hörner 2009), but subjects follow the majority more often than the minority. Models that combine heterogeneity in the level of strategic thinking and allow for some degree of trembling (e.g., noisy introspection proposed by Goeree and Holt 2004) fit our data best.

A third chapter evaluates the performance of four different auction formats. We find that bidders are not always bidding on the currently most-profitable combination of available items as often assumed in the literature. Instead, subjects sometimes

submit jumpbids. As a result, a clock auction (Porter et al. 2003) in which prices can only increase incrementally generates particularly high revenues. We also find that subjects are reluctant to risk exposure: when they have a high value for a combination of items but low values for each item separately, they are unwilling to bid high on these single items unless the auction allows them to submit a package bid. Such package bids specify that the bid is only valid if the bidder wins all items included in the package.

In the last chapter, we compare five different stationary concepts: Nash equilibrium, quantal-response equilibrium, action-sampling equilibrium, payoff-sampling equilibrium and impulse-balance equilibrium. Selten and Chmura (2008) run a large number of completely mixed 2×2 games in the laboratory for that purpose. We reanalyze their data and find that there are no significant differences with respect to goodness of fit except that the Nash equilibrium fits worse than all the other models. In a game with a risky and a safe choice (Goeree et al. 2003), impulse-balance equilibrium yields a particularly good fit, which is due to its built-in loss aversion. When other models are augmented with loss aversion, they yield an even slightly superior fit. In games in which losses cannot occur (McKelvey et al. 2000), action sampling and payoff sampling fit better than logit QRE and impulse balance, which fit better than Nash.

Contents

List of Figures x

List of Tables xiv

1 Introduction 1

2 Cascades in Experimental Asset Markets 5

2.1 Introduction . . . 5

2.2 Experimental Design . . . 8

2.3 Conjectures and Definitions . . . 10

2.4 Results . . . 12

2.4.1 Bad Cascade Periods . . . 13

2.4.2 Good Cascade Periods . . . 17

2.5 Alternative Cascade Period Definitions . . . 19

2.5.1 Shorter Information Cascades . . . 19

2.5.2 Who Trades First? . . . 19

2.5.3 Who Submits Orders First? . . . 21

2.6 Alternative Explanations . . . 22

2.6.1 Early Expected Value vs. Late Expected Value . . . 22

2.6.2 Strategic Behavior . . . 23

2.7 Conclusion . . . 24

3 Wise Crowds or Wise Minorities? 26

3.1 Introduction . . . 26

3.2 Theoretical Predictions 31

3.3 Experimental Design 33

3.4 Results 36

3.4.1 Subject Behavior 36

3.4.2 Efficiency 42

3.5 Explaining Subject Behavior 43

3.5.1 Logit-QRE 43

3.5.2 Subject Heterogeneity 45

3.5.3 Level-k and Cognitive Hierarchy 46

3.5.4 Noisy Introspection 49

3.6 Conclusion 53

4 An Experimental Test of Flexible Spectrum Auction Formats 56

4.1 Introduction 56

4.2 Alternative Auction Formats 57

4.3 Experimental Design 62

4.4 Results 66

4.5 Individual Bidding Behavior 74

4.6 Conclusion 79

5 A Correction and Reexamination of "Stationary Concepts for Experimental 2 × 2 Games" 81

5.1 Introduction 81

5.2 Reexamining the Selten and Chmura (2008) Results 84

5.3 Differentiating Stationary Concepts in Other Data Sets 90

5.3.1 A Matching Pennies Game with Safe and Risky Choices . . . 90

5.3.2 Asymmetric Matching Pennies Games 91

5.4 Conclusion 93

A Appendix to Chapter 2 96

A.1 Treatment Sequence No-Cascade Periods 96

A.2 Treatment Baseline No-Cascade Periods 97
A.3 Instructions for Treatment Baseline 97
A.4 Instructions for Treatment Sequence 103

B Appendix to Chapter 3 **109**
B.1 Instructions for Treatment Sequence 109
B.2 Instructions for Treatment No-Sequence 113

C Appendix to Chapter 4 **115**
C.1 Rules for the Simultaneous Multi-Round (SMR) Auctions 115
C.2 Rules for the Resource Allocation Design (RAD) and the Simultaneous Multi-Round Auction with Package Bidding (SMRPB) 117
C.3 Rules for the Combinatorial Clock (CC) 118
C.4 Experimental Instructions 120
C.5 Experimental Instructions for the Activity Rule 139
C.6 Experimental Instructions for the Pricing Rule 140

Bibliography **143**

List of Figures

2.1 Prices in Treatment Sequence in Bad Cascade Periods 14
2.2 Prices in Treatment Baseline in Bad Cascade Periods 15
2.3 Prices in Treatment Sequence in Good Cascade Periods 17
2.4 Prices in Treatment Baseline in Good Cascade Periods 18
2.5 Prices for an Additional Bad Cascade Period 20

3.1 Probability the Minority is Correct in Treatment No-Sequence 34
3.2 Probability the Minority is Correct in Treatment No-Sequence 35
3.3 Subject Behavior in Treatment Sequence. 37
3.4 Subject Behavior in Treatment No-Sequence. 38
3.5 Frequency with Which Unanimous Predecessors Are Followed 39
3.6 Frequency with Which a Deviator's Choice is Followed 40
3.7 Frequency with Which the Minority is Followed 41
3.8 Subject Behavior in a Logit QRE Model 44
3.9 Fraction of Rational Choices 46
3.10 Subject Behavior in a Poisson Level-k Model 49

4.1 Eight-Bidder Design with 3 Regions 63
4.2 Efficiency by Auction Format 68
4.3 Revenue by Auction Format 71
4.4 Bidder's Profits by Auction Format 73

5.1 Mean Squared Distances for the Action-Sampling Equilibrium 86
5.2 Visualization of Equilibrium Predictions and Observed Averages 87
5.3 Mean Squared Distances of the 5 Stationary Concepts 87

5.4 Theory Specific Squared Distances by Blocks of 100 Periods 88

5.5 A Matching Pennies Game with Safe and Risky Choices 90

5.6 Mean-Squared Distances for Game 4 91

5.7 An Asymmetric Matching Pennies Game. 92

A.1 Prices in Treatment Sequence No-Cascade Periods 96

A.2 Prices in Treatment Baseline No-Cascade Periods 97

A.3 Drawing Signals. 100

A.4 Your Screen. 102

A.5 Submitting Orders. 103

A.6 Your Screen: Updated Info. 104

A.7 Accepting Standing Orders. 105

A.8 Withdrawing Orders I. 106

A.9 Withdrawing Orders II. 106

A.10 Sorting Standing Orders. 107

A.11 Filtering Standing Orders. 107

A.12 End of Period Feedback. 108

A.13 Summary. 108

B.1 Urn Representation. 110

B.2 Information Given to Subjects in Treatment Sequence. 111

B.3 Record Sheet. 113

B.4 Information Given to Subjects in Treatment No-Sequence. 114

C.1 Slide 1. 120

C.2 Slide 2. 121

C.3 Slide 3. 121

C.4 Slide 4. 122

C.5 Slide 5. 122

C.6 Slide 6. 123

C.7 Slide 7. 123

C.8 Slide 8. 124
C.9 Slide 9. 124
C.10 Slide 10. 125
C.11 Slide 11. 125
C.12 Slide 12. 126
C.13 Slide 13. 127
C.14 Slide 14. 127
C.15 Slide 15. 128
C.16 Slide 16. 128
C.17 Slide 17. 129
C.18 Slide 18. 129
C.19 Slide 19. 130
C.20 Slide 20. 130
C.21 Slide 21. 131
C.22 Slide 22. 132
C.23 Slide 23. 133
C.24 Slide 24. 133
C.25 Slide 25. 134
C.26 Slide 26. 134
C.27 Slide 27. 135
C.28 Slide 28. 135
C.29 Slide 29. 136
C.30 Slide 30. 136
C.31 Slide 31. 137
C.32 Slide 32. 137
C.33 Slide 33. 138
C.34 Slide 34. 138
C.35 Activity Rule I. 139
C.36 Activity Rule II. 140
C.37 Pricing for the Combinatorial Clock I. 141

C.38 Pricing for the Combinatorial Clock II. 141

List of Tables

2.1 Experimental Design. 10
2.2 Test Results Bad Cascade Periods 16
2.3 Test Results Good Cascade Periods 18
2.4 Test Results for Cascade Periods Based on the Time of the First Transaction 20
2.5 Test Results for Cascade Periods Defined Based on the Time of the First Order 22
2.6 Using Devalue to Explain Differences in the Quality of Information Aggregation. 24

3.1 Experimental Design Parameters and Subjects' Earnings. 36
3.2 Overview of Different Models 51

4.1 Summary Statistics by Auction Format. 74
4.2 Bidding Behavior in SMR. 76
4.3 Average Bid Characteristics when Complementarities Are Low 77
4.4 Bidding up to Value. 78
4.5 Size of Jumpbid 78

5.1 Estimates for 5 Stationary Concepts 85
5.2 Wilcoxon Matched-Pairs Signed-Rank Tests for the SC Data 89
5.3 5 Stationary Concepts for Experimental Games 93
5.4 Pairwise Tests 94

C.1 Efficiency, Revenue, and Profits by Auction Format. 142

Chapter 1

Introduction

There are many economically relevant situations in which individual actors hold private information. Aggregating this privately held information can often greatly improve the quality of individual choices. Therefore, it is of great interest to examine institutions that perform that function. One example of such an institution is a double auction market. According to the efficient market hypothesis, prices in asset markets aggregate privately held information quickly and without bias (Fama 1970). A large experimental literature (e.g., Plott and Sunder 1982, 1988) examines under which conditions the aggregation of individually held information succeeds in such markets. However, the process of how information is aggregated is still not very well understood. In chapter 2, we will focus on one possible explanation for why privately held information is not always aggregated successfully. Since neither individually held information nor valuations are easily observable in the field, we turn to the laboratory to address this question.

An important part of the aggregation process in asset markets involves individual traders observing the bids and asks that their predecessors submitted. When traders no longer pay attention to their own private signals and instead follow the choices of their predecessors, an "information cascade" occurs and privately held information is no longer being aggregated (Bikhchandani et al. 1992, Banerjee 1992). It has been suggested that such information cascades might occasionally prevent asset markets from performing efficiently (e.g. Alevy et al. 1997). However, Avery and Zemsky (1998) show that price adjustments prevent information cascades in markets under common knowledge of rationality when there is only one dimension of uncertainty. Only when some of these conditions are relaxed is it possible that agents neglect their private information and simply follow the choices of their predecessors (Avery and Zemsky 1998, Lee 1998).

In the experiment that we discuss in chapter 2, we implement environments in which information cascades would occur in the absence of prices and test whether they can also be observed in a market setting. We find evidence that the sequence in which private signals are released to traders has an effect on the quality of information aggregation: when early signals are misleading, prices are less likely to converge to the fully revealing rational expectations equilibrium. At the same time, we find that the quality of information aggregation is not significantly worse in periods in which information cascades would occur in the absence of prices.

While further research is needed to establish how important cascades are to explain price formation in asset markets, previous experiments show that they do occur in the environment originally proposed by Bickhchandani et al. (1992). When agents sequentially choose among two options for which they have common values, rational agents should often ignore their private signal and follow their predecessors instead, and subjects indeed often do so in the laboratory (Anderson and Holt, 1997, for example). However, in many situations, it is impossible to observe the entire sequence of predecessors' choices. Instead, only the number of choices for each option is visible. A tourist can, for example, only see the number of diners seated in the restaurants among which he has to choose. Callander and Hörner (2009) show that in such a situation, it can be optimal to follow the minority when some agents have better information than others. Chapter 3 focuses on this type of sequential decision-making problem. In the laboratory, the prediction that subjects follow the minority fails miserably. In fact, they are more likely to follow the majority than the minority. This observed behavior is not entirely irrational: when agents tremble, the expected payoff of following the minority is lower than the expected payoff of following the majority. Trembles alone cannot explain subject behavior in this game satisfactorily, though. We also observe heterogeneity in the level of strategic thinking and models that combine this feature with trembles (e.g., “noisy introspection” proposed by Goeree and Holt 2004) fit our data best.

Like in markets or in sequential games, individually held information is also aggregated in auctions. These institutions typically aim at maximizing either the sellers revenue or allocative efficiency. In the process, information about what the objects for sale are worth to bidders is aggregated, even though this is not the primary purpose of these mechanisms. Since the true values of bidders are typically unobservable in the field, laboratory experiments are a suitable way to test different auction formats. In chapter 4, we compare three combinatorial auctions as well as the simultaneous multiround (SMR) auction. In a treatment with high value complementarities, all three combinatorial auction formats generate a more-efficient allocation than SMR. In these auctions, bidders can submit package bids, which are only valid if the bidder wins all items included in the package. These package bids help bidders to avoid the “exposure problem” which arises when a bidder has to bid on several items separately and risks incurring losses in case he ends up winning only some of these items. We also find that a clock auction (originally proposed by Porter et al. 2003) in which prices steadily increase from one round to another on items for which there is excess demand, generates the highest revenue. Part of the reason might be that subjects are not always bidding in a straightforward way as typically assumed in the literature (e.g., Milgrom 2004, p. 270). Instead, they sometimes submit jumpbids in auction formats that allow them to do so and that can lead to inefficient allocations.

In chapter 5, we examine games in which players have no private information other than their intentions of how to play the game and their beliefs about what other players will do. Clearly, agents in these games have every interest in learning about these intentions and beliefs held by other players. When playing the same game repeatedly, agents will presumably converge to some stationary equilibrium. Selten and Chmura (2008) run a large number of experiments to test five different equilibrium concepts on completely mixed 2×2 games. We reanalyze their data, correct some errors, and find that there are no significant differences in terms of goodness of fit except that the Nash equilibrium fits worse than any of the other models. In order to differentiate better between the five concepts, we also apply them to previously published data of an experimental 2×2 game with a risky and a safe choice (Goeree et al. 2003). Impulse balance equilibrium, an equilibrium concept with built-in loss aversion, explains subject behavior in this game particularly well. However, all other non-Nash models do even slightly better than impulse balance once they are also augmented with loss aversion. In games in which no losses relative to the max-min payoff are possible (McKelvey et al. 2000), the payoff-sampling and the action-sampling equilibrium fit better than logit QRE or impulse balance, which in turn fit better than Nash.

Chapter 2

Cascades in Experimental Asset Markets

2.1 Introduction

In the mid-1500s, the Netherlands became a center of cultivation and development of new tulip varieties. A market for tulip bulbs was established. From November 1636 to January 1637, prices of rare varieties surged upward and then rapidly collapsed to approximately 10 percent of their peak values (Garber, 1990). This episode, commonly referred to as "tulip mania, is only one example for trade occurring at seemingly irrational prices. Other instances often involve financial markets: in May 1719, the shares of the French Compagnie des Indes sold at approximately 500 livres per share. About 5 months later, the same shares were traded at 10.000 livres. The surge was followed by a crash: in September 1721, the price was down at 500 livres per share again (Garber, 1990). More recent examples of rapidly rising and shortly afterwards even more rapidly collapsing prices include the stock market movements in the United States at the end of the 1920s or the 1980s. In recent years, the valuations of many Internet-related companies exhibited similar patterns.

Since the expected value of these assets given all available information at the time is unknown, it remains unclear whether the according prices deviated from their fundamental value or not. If they did, one possible explanation is that these were instances of information cascades. According to Bikhchandani et al. (1992), "An informational cascade occurs when it is optimal for an individual, having observed the actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information" (p. 994). In an information cascade, individually held information is no longer revealed to others. As a consequence, agents can end up making suboptimal choices given all privately held information at the time.

In this chapter, we will examine whether information cascades really routinely occur in double auctions with endogenous prices. After a brief review of the relevant literature, we will discuss the experimental design and then present the according results.

Even though both Bikhchandani et al. (1992) and Banerjee (1992) mention financial markets as a possible environment in which information cascades might occur, agents in their models sequentially choose among a set of options for which they have common values. There are no prices and no trade occurs. In such environments, information cascades are quite frequently observed in the laboratory (Anderson and Holt, 1997, for example). However, it is unclear whether these results are relevant to financial markets. Avery and Zemsky (1998) show that price adjustments prevent information cascades under common knowledge of rationality when there is only 1 dimension of uncertainty. To test whether the presence of prices really eliminates information cascades, Drehman et al. (2005a) and Cipriani and Guarino (2005) implement markets in which the price of the asset always corresponds to the expected value given the history of previous choices. If all agents were fully rational, subjects should always follow their private information in these markets but they often fail to do so. Nevertheless, neither Drehman et al. (2005a) nor Cipriani and Guarino (2005) find evidence for information cascades. Instead, subjects exhibit contrarian tendencies.

Despite these findings, it is not clear that it is impossible for information cascades to be observed in financial markets. Avery and Zemsky (1998) show that information cascades can occur in a model with an uninformed market maker, noise traders and event uncertainty. Agents in their model are only allowed to trade once in an exogenously determined sequence. Lee (1998) relaxes that latter condition and shows that even under common knowledge of rationality, agents are not necessarily always following their private information. However, these are not information cascades in the sense of Bikhchandani et al. (1992) since agents' actions still depend on their private signal, they just might choose not to follow their signal if it is not strong enough.

In the experimental literature, there is some evidence suggesting that information cascades might occur even when prices are endogenous and in the absence of restrictions on the time and quantity of trade. In an experimental double auction, Barner et al. (2005) find that informed traders tend to trade particularly actively at the beginning of the period except in periods in which information aggregation fails. This result suggests that subjects carefully observe what other traders do and update their valuations accordingly. If the first few actions in a trading period are misleading, prices tend not to reflect the expected value of the asset. Plott and Sunder (1982) also report that informed traders are particularly active in the early stages of trading. Other experimental studies find that subjects tend to rely on their private information too often (Nöth and Weber 1993, for example) or that they purchase too many signals (Krämer et al. 2006). These results would rather suggest that subjects are unwilling to rely too strongly on information revealed by the market and as a consequence, it might be difficult to observe cascades in a market setting.

Hey and Morone (2004) run experimental asset markets in which subjects can purchase signals that indicate the true value of the asset. They report that in 1 period, the first few signals purchased were misleading and as a consequence, prices failed to converge to the true value of the asset. Since they do not report the expected value of the asset given all signals purchased, it is not entirely clear whether information aggregation really failed in this instance. Even if it did, other factors such as a high variance of the signals purchased might have contributed to the failure of prices to converge to the fully revealing rational expectations equilibrium. Moreover, there might have been other periods in which early signals were misleading but information aggregation nevertheless succeeded.

In order to find out whether the sequence in which information is released to the market affects the quality of information aggregation, we design experiments that allow us to control what information is released at which time more closely. Moreover, we run a control treatment in which the same information is released simultaneously. As a result, we can clearly identify periods in which information cascades would occur if agents sequentially chose to buy or sell a unit of the asset at a fixed price. We can then test whether the quality of information aggregation in these periods differs relative to other periods. The treatment in which signals are released simultaneously allows us to control for other possible explanations such as differences in the variance of the signals released.

2.2 Experimental Design

In order to give subjects an incentive to pay attention to what other agents do, all subjects have the same value for the asset that is being traded. This value is determined independently for each trading period and is equally likely to be 0 or 100 units of the experimental currency. Even in such an environment, subjects often place too much weight on their private information, which would prevent cascades from occurring. One possible way to give traders an even higher incentive to infer what signals other traders received is to release complementary signals. For example, Plott and Sunder (1988) run experiments in which there are 3 possible states of the world, x, y and z. Suppose the true state is y. In that case, half of the traders are told that the state is either x or y while the other half knows that the true state is either y or z. Such an information structure clearly encourages subjects to carefully observe what others do. However, it is not suitable to study cascades since traders would never ignore their private information.

Another possibility is to give more accurate signals to some traders. In that case, it should be more obvious to traders with inaccurate signals that it is in their interest to infer what signals other traders received. Moreover, they are probably more likely to ignore their private information given that others know more about the true value of the asset. Plott (2000) reports convergence of prices to the rational expectations equilibrium in an experiment with heterogeneous private signals.

In order to keep the experimental design as simple as possible, there are only 2 types of signals in our experiments: strong signals and weak signals. Subjects with weak signals should have every interest in finding out what information subjects with strong signals have. Therefore, the difference between the accuracy of the strong signal and the accuracy of the weak signal should be large. At the same time, even subjects with strong signals should have an interest in observing what others do and as a result, the strong signal should not be too accurate. Also, if weak signals contain almost no information, it would be trivial to find that subjects' decisions do not depend on their private information. For these reasons, the weak signal reflects the true value of the asset with probability 0.6 while the strong signal corresponds to the true value with probability 0.8. Each one of 8 traders is equally likely to receive a strong or a weak signal.

Trading occurs in a continuous double auction that was implemented in jTrade. Each subject is given an endowment of 5 units of the asset at the beginning of each one of 7 trading periods. Subjects also receive a cash loan of 500 units of the experimental currency that they have to repay at the end of the period. Even if prices are at 100, subjects would thus be able to purchase at least 5 units. In order to preserve the symmetry between the buy and the sell side, subjects only have values for at most 10 units of the asset. If a subject ends up holding more than 10 units of the asset at the end of a period, his value for these assets would nevertheless at most be 1000 units of the experimental currency.

In treatment baseline, all subjects receive their private information at the beginning of each trading period. They can then trade for 2.5 minutes. In treatment sequence, the first subject also receives his signal at the beginning of the period and 30 seconds after the market period opens, the second subject receives his private information. After another 30 seconds, another signal is released to the next subject until all 8 subjects received their signals. The position of each subject in this sequence is randomly determined for each period and is shown to other traders along with each bid or ask that the subject submits. As a consequence, all subjects always know

Table 2.1. Experimental Design.

Treatment	# Sessions	# Subjects per Session	Average Earnings
Baseline	5	8	\$20
Sequence	5	8	\$20

whether the trader who submitted a certain bid or ask already received his signal. After all signals are released, subjects are given another 2.5 minutes to trade. As a result, a market period in treatment sequence lasts 6 minutes. This duration is quite different compared to treatment baseline but the time of trading available after all information is released is exactly the same in both treatments. Since we use the same signal and value draws for treatment baseline and treatment sequence, we can test whether adding extra time during which information is released sequentially affects the quality of information aggregation.

We run 5 sessions for each treatment using undergraduate and graduate students at Caltech as subjects. Each subject was allowed to participate only once. In each session, there is 1 practice period that does not affect earnings. Subjects receive a \$5 show-up fee as well as \$1 for every 100 units of the experimental currency. As a result, expected earnings are \$22.5 per subject with sessions typically lasting for 1 hour for treatment baseline and 1.5 hours for treatment sequence.

2.3 Conjectures and Definitions

Under common knowledge of rationality, no trade should occur in either treatment sequence or treatment baseline (Milgrom and Stokey 1982). However, we know from previous market experiments that this is rather unlikely to happen. If trade does occur and markets are efficient, prices should always reflect all information available to any trader (Fama 1970). As a consequence, no information cascades should occur. On the other hand, some subjects might fail to use their private information to update their estimate of the value of the asset and instead rely on what they believe is revealed by actions of other market participants. In that case, information cascades could be observed. More specifically, we examine both “good” and “bad” cascade periods and compare the quality of information aggregation in these periods to other periods.

Definition 1 *A bad (good) cascade period is a period which satisfies the following conditions:*

- *If subjects had to sequentially guess whether the value of the asset was 0 or 100 and the sequence of choices was observable, at least half of them would not pay attention to their private signal under common knowledge of rationality.*
- *The majority of subjects would choose the wrong (correct) value.*

For the parameters we use in our experiment, only periods in which the first signal is misleading and both the second and the third signal are either weak or both strong and misleading qualify as bad cascade periods. Similarly, only periods in which the first signal is correct and both the second and the third signal are either weak or both strong and correct are good cascade periods. If information cascades are likely to occur in our markets, we expect that the quality of information aggregation in bad cascade periods is worse than in other periods.

Conjecture 1 *The quality of information aggregation for bad cascade periods is higher in treatment baseline than in treatment sequence.*

Conjecture 2 *The quality of information aggregation for bad cascade periods in treatment sequence is lower than for other periods in treatment sequence.*

To measure the quality of information aggregation, we compute the average of the absolute value of the difference between transactions prices and the expected value of the asset given all private signals. We calculate this average for 4 different sets of transactions:

- All transactions that occurred during the last 2.5 minutes of trading

- All transactions that occurred during the last 1.5 minutes of trading
- The last 5 transactions
- The last 5 units that were traded

The last 2.5 minutes of trading are relevant because all information has been released by that time in treatment sequence. Therefore, prices would be identical in both treatments if markets were efficient. We also take the average over transactions that occurred during the last 1.5 minutes of trading in order to allow for time for traders in treatment baseline to reveal their signals to others. For all measures, the mean is always taken over the number of units that transact. For example, 2 transactions for 1 unit at price x are equivalent to 1 transaction for 2 units at price x. The only exception is the third measure, for which we take the average over the last 5 transactions giving equal weight to each one of them. Transactions at a price of 0 or at prices of 100 or higher are dropped because these are obvious mistakes.

Just like in bad cascade periods, the aggregation of privately held signals stops after some point in good cascade periods if information cascades occur. However, The price at which information aggregation typically stops is almost always quite close to the expected value of the asset given all private signals. Therefore, we do not expect information aggregation in good cascade periods to be substantially worse than in other periods but we nevertheless test whether or not it is. We will also test whether there is a substantial difference in the quality of information aggregation between treatment sequence and treatment baseline using all periods as observations. Since there are only a few bad cascade periods and since we have no reason to expect substantial differences in other periods, we do not expect these differences to be significant.

2.4 Results

In this section, we will first test the conjectures stated above. Since we only find very weak evidence to support them, we then test whether circumstances other than the sequence in which signals are released might be favorable to trigger information cascades. To conclude, we provide some evidence for strategic behavior on the part of subjects, which might explain why we fail to find significant differences between periods that could be favorable to information cascades and other periods in terms of the quality of information aggregation.

2.4.1 Bad Cascade Periods

To test conjecture 1, we compare bad cascade periods in treatment sequence to bad cascade periods in treatment baseline. Figures 2.1 and 2.2 display the according price patterns separately for each one of the 5 periods that qualify as bad cascade periods. The first number at the top of each period-specific graph indicates the session while the second number corresponds to the period. The size of the dots is proportional to the number of units exchanged. The line corresponds to the expected value of the asset given all private signals. In treatment sequence, prices clearly fail to converge to the rational expectations equilibrium in session 3, both in period 4 and period 6. It could very well be that bad information released early in these periods led to an information cascade that prevented prices from effectively aggregating information. However, prices in the corresponding periods of treatment baseline also failed to converge to the expected value of the asset given all privately held information. As a result, none of the tests we run allows us to reject the null hypothesis that the quality of information aggregation in treatment baseline is equivalent to the quality of information aggregation in treatment sequence for bad cascade periods. Table 2.2 contains the p-values of Wilcoxon matched-pairs signed-rank tests for the 4 different measures of the quality of information aggregation.

Conjecture 2 does not fare much better than conjecture 1. Wilcoxon rank-sum tests do not allow us to reject the null that the median of the quality of information aggregation is identical in bad cascade periods compared to other periods within treatment sequence. When comparing all 35 periods of treatment sequence to all 35 periods of treatment baseline, we are getting fairly close to rejecting the null hypoth-

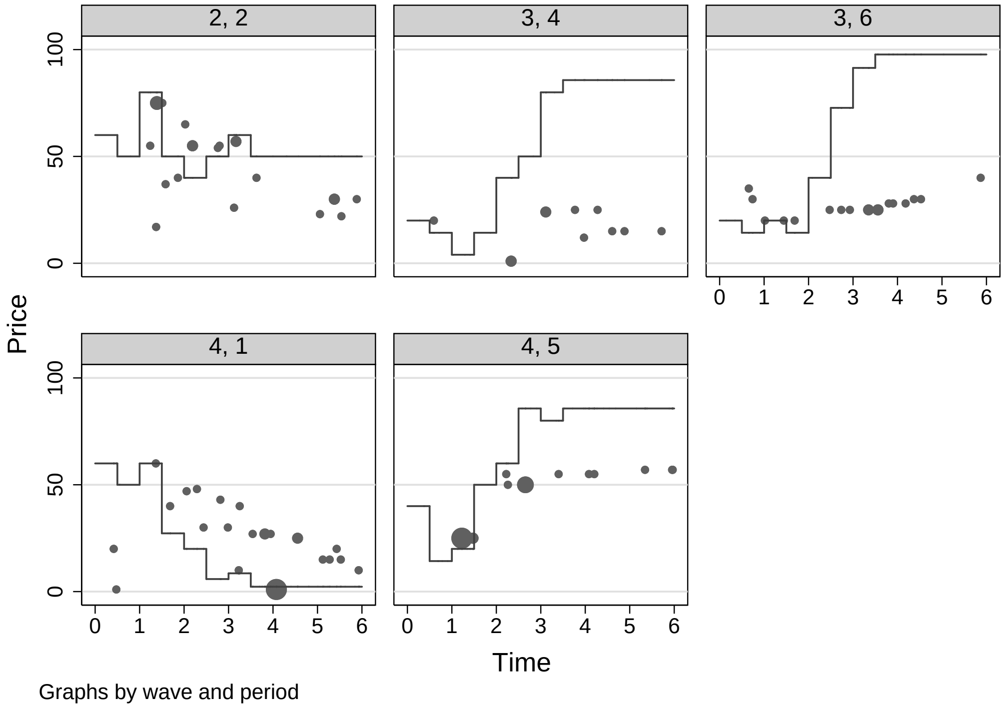


Figure 2.1. Prices in Treatment Sequence in Bad Cascade Periods.

esis of equal quality of information aggregation for some of the measures employed with treatment sequence exhibiting the higher average absolute deviation of prices from the rational expectations equilibrium price.

The fact that neither one of the main conjectures could be confirmed while there is some support for the hypothesis that the quality of information aggregation is higher in treatment baseline compared to treatment sequence might be due to an insufficient number of bad cascade periods. Clearly, information cascades are not guaranteed to occur even when the sequence of private signals would provide favorable conditions. At the same time, information is not always aggregated very efficiently in treatment baseline, either.

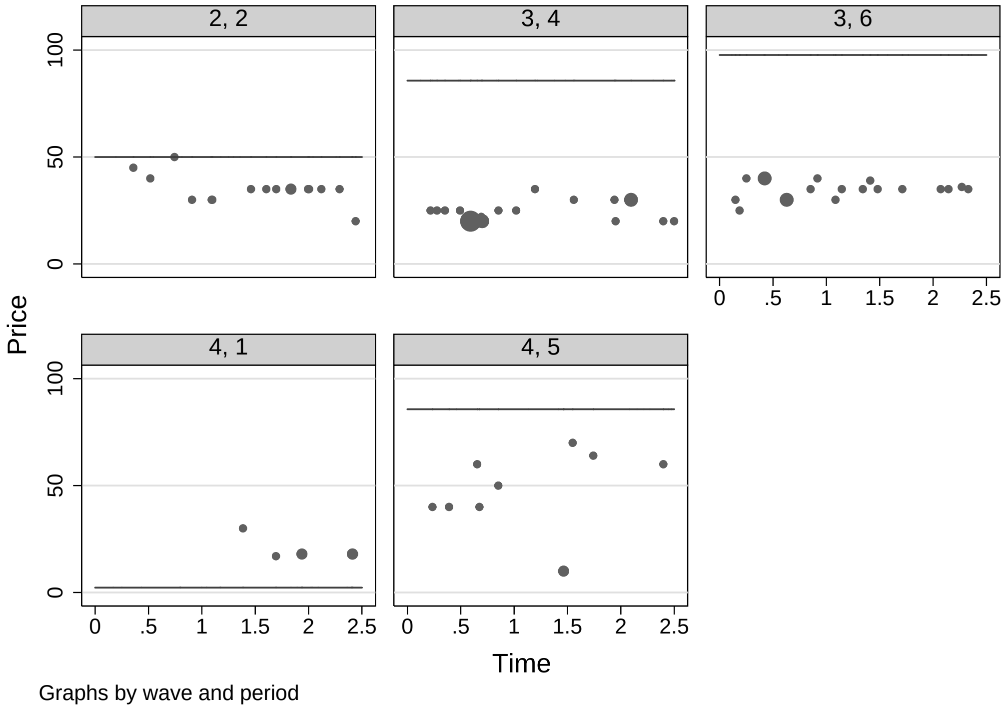


Figure 2.2. Prices in Treatment Baseline in Bad Cascade Periods.

Table 2.2. Test Results Bad Cascade Periods.

Measure	Bad Cascade Periods Sequence vs. Bad Cascade Periods Baseline	Bad Cascade Periods Sequence vs. Other Periods Sequence	All Periods Sequence vs. All Periods Baseline
Mean Absolute Deviation Using Transactions During the Last 90 Seconds	Mean Baseline: 39.7 Mean Sequence: 40.1 p = 1.0	Mean Cascade: 40.1 Mean Other: 31.1 p = 0.39	Mean Baseline: 27.6 Mean Sequence: 32.4 p = 0.18
Mean Absolute Deviation Using Transactions During the Last 150 Seconds	Mean Baseline: 39.7 Mean Sequence: 39.9 p = 0.81	Mean Cascade: 39.9 Mean Other: 30.7 p = 0.57	Mean Baseline: 31.6 Mean Sequence: 32.0 p = 0.59
Mean Absolute Deviation Using the Last 5 Transactions	Mean Baseline: 39.0 Mean Sequence: 39.8 p = 1	Mean Cascade: 39.8 Mean Other: 31.5 p = 0.51	Mean Baseline: 27.5 Mean Sequence: 32.7 p = 0.11
Mean Absolute Deviation Using the Last 5 Units Traded	Mean Baseline: 39.8 Mean Sequence: 40.2 p = 0.81	Mean Cascade: 40.2 Mean Other: 31.4 p = 0.45	Mean Baseline: 27.4 Mean Sequence: 32.7 p = 0.14

We run Wilcoxon matched-pairs signed-rank tests to obtain the p-values for the first and

2.4.2 Good Cascade Periods

Figures 2.3 and 2.4 display the price patterns in good cascade periods for treatment sequence and treatment baseline. The size of the dots is proportional to the number of units exchanged. The line corresponds to the expected value of the asset given all private signals. Table 2.3 contains the according test results. We run Wilcoxon matched-pairs signed-rank tests to obtain the p-values for the first column and Wilcoxon rank-sum tests for the second column. While only 5 out of 35 periods qualify as bad cascade periods, 16 qualify as good cascade periods. This might be part of the reason why some of the comparisons between treatment sequence and treatment baseline almost yield significant results. It appears that the quality of information aggregation is somewhat lower in good cascade periods in treatment sequence compared to good cascade periods in treatment baseline.

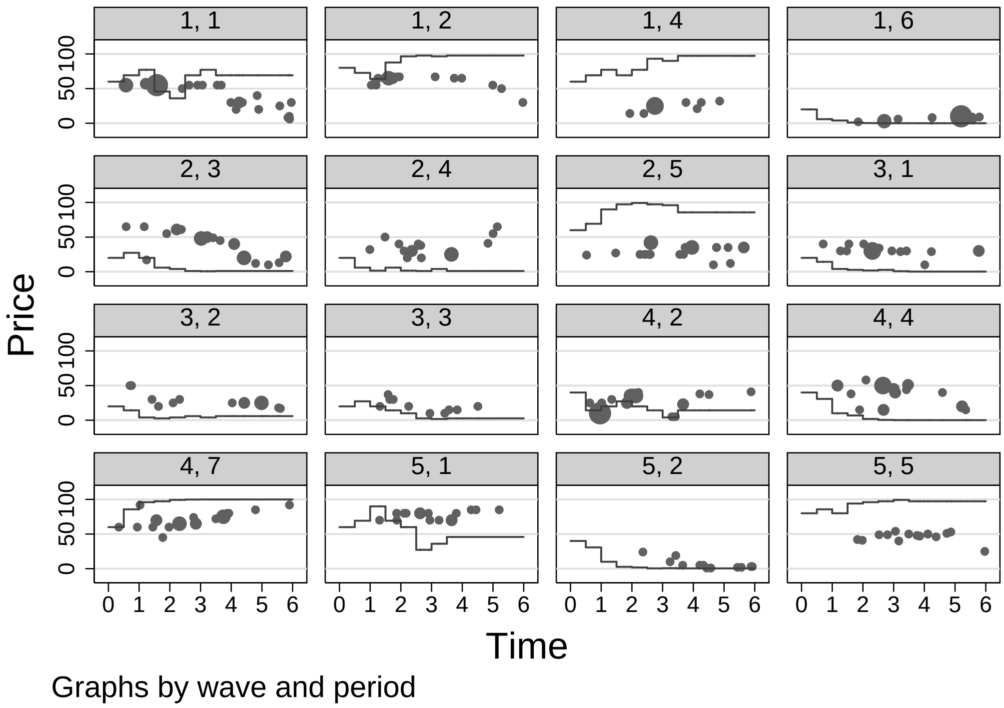


Figure 2.3. Prices in Treatment Sequence in Good Cascade Periods.

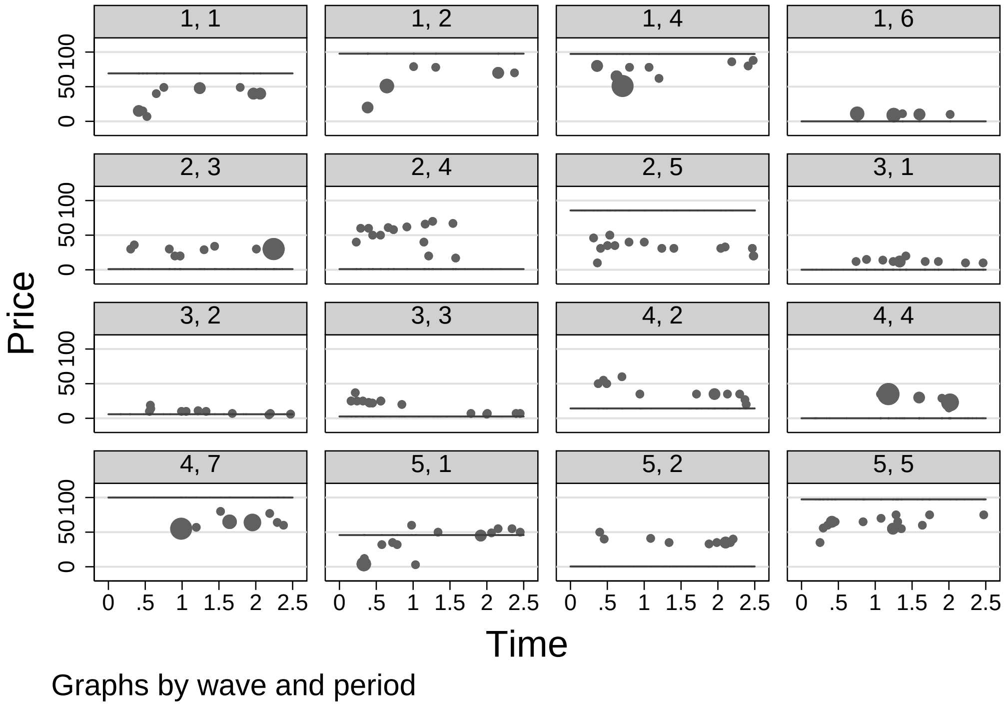


Figure 2.4. Prices in Treatment Baseline in Good Cascade Periods.

Table 2.3. Test Results Good Cascade Periods.

Measure	Good Cascade Periods Sequence vs. Good Cascade Periods Baseline	Good Cascade Periods Sequence vs. Other Periods Sequence
Mean Absolute Deviation Using Transactions During the Last 90 Seconds	Mean Baseline: 23.6 Mean Sequence: 32.4 p = 0.11	Mean Cascade: 32.4 Mean Other: 32.4 p = 0.89
Mean Absolute Deviation Using Transactions During the Last 150 Seconds	Mean Baseline: 28.5 Mean Sequence: 30.3 p = 0.68	Mean Cascade: 30.3 Mean Other: 33.4 p = 0.95
Mean Absolute Deviation Using the Last 5 Transactions	Mean Baseline: 23.7 Mean Sequence: 31.3 p = 0.15	Mean Cascade: 31.3 Mean Other: 33.9 p = 0.79
Mean Absolute Deviation Using the Last 5 Units Traded	Mean Baseline: 23.5 Mean Sequence: 30.9 p = 0.16	Mean Cascade: 30.9 Mean Other: 34.2 p = 0.74

2.5 Alternative Cascade Period Definitions

2.5.1 Shorter Information Cascades

Clearly, the low number of bad cascade periods combined with a relatively high variance in the quality of information aggregation in both treatments contributes to the fact that we did not find much support in favor of conjectures 1 and 2. A possible remedy would be to apply a more liberal definition of cascade periods by relaxing the condition that at least 4 out of 8 traders would ignore their private signal if they chose sequentially whether to buy or sell the asset. However, if only very few traders ignore their private information, it would be difficult to find significant differences with respect to the quality of information aggregation even if an information cascade actually occurred. Therefore, the only extension that we test is one in which at least 3 traders would ignore their private information if they chose sequentially. Unfortunately, this more-liberal definition only yields 1 additional bad cascade period (session 5, period 4) and all differences in the quality of information aggregation remain insignificant. Figure 2.5 displays transaction prices in this additional bad cascade period. The size of the dots is proportional to the number of units exchanged. The line corresponds to the expected value of the asset given all private signals.

2.5.2 Who Trades First?

Another reason why conjectures 1 and 2 could not be confirmed might be that the sequence in which signals are released does not correspond to the sequence in which subjects actually trade. As a consequence, the sequence in which private information is revealed to the market might not correspond to the sequence in which signals are released to traders. To test this hypothesis, we compute the Spearman rank correlation coefficient between the position of a subject in the sequence in which private signals are released in treatment sequence and the time at which that subject first buys or sells at least 1 unit of the asset. We obtain a positive correlation (0.24) and can easily reject the null hypothesis of no correlation (p=0.000).

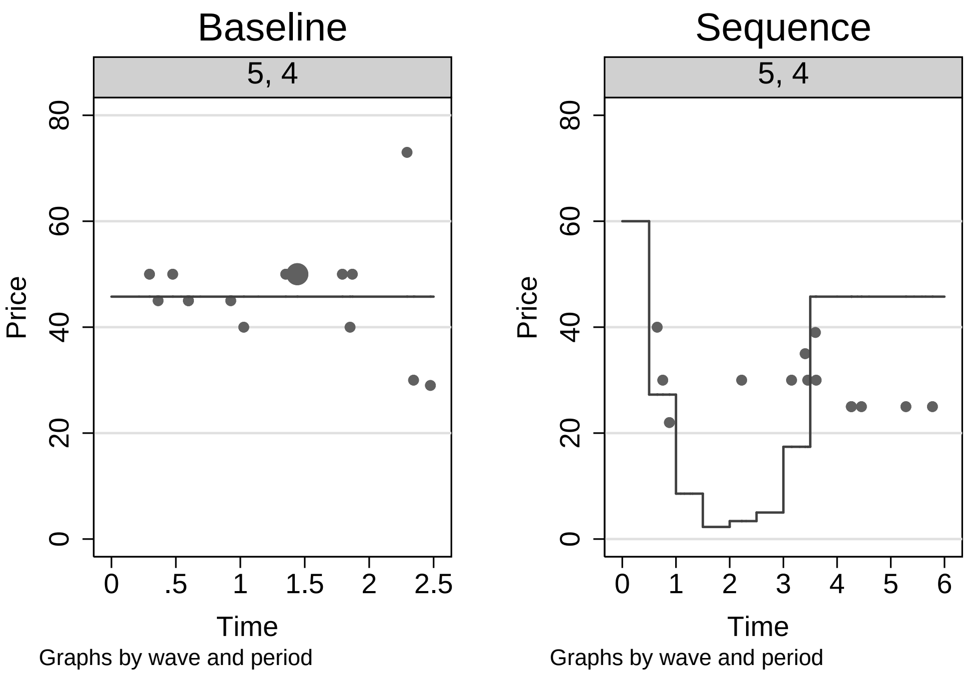


Figure 2.5. Prices in Session 5, Period 4 for Treatments Baseline and Sequence.

Table 2.4. Test Results For Cascade Periods Defined Based on the Time of the First Transaction. We run Wilcoxon rank-sum tests to obtain the p-values.

Measure	Bad Cascade Periods vs. Other Periods Both Treatments	Good Cascade Periods vs. Other Periods Both Treatments
Mean Absolute Deviation Using Transactions During the Last 90 Seconds	Mean Cascade: 31.9 Mean Other: 29.8 p = 0.79	Mean Cascade.: 32.7 Mean Other: 27.4 p = 0.35
Mean Absolute Deviation Using Transactions During the Last 150 Seconds	Mean Cascade: 33.5 Mean Other: 31.6 p = 0.88	Mean Cascade: 35.1 Mean Other: 28.6 p = 0.21
Mean Absolute Deviation Using the Last 5 Transactions	Mean Cascade: 33.8 Mean Other: 29.7 p = 0.57	Mean Cascade: 32.6 Mean Other: 27.7 p = 0.44
Mean Absolute Deviation Using the Last 5 Units Traded	Mean Cascade: 34.9 Mean Other: 29.5 p = 0.45	Mean Cascade: 32.0 Mean Sequence: 28.2 p = 0.55

Since the correlation is not perfect, we reclassify cascade periods based on the sequence in which subjects trade. We apply the same definition of cascade periods that we originally used (Definition 1) but we now use the sequence of signals obtained by the first 3 subjects who are buying or selling at least 1 unit of the asset to classify periods. In treatment sequence, it occasionally happens that some of the these first 3 traders have not yet received their private information at the time at which they first trade. We are not taking these traders into account since their private signal clearly cannot have been revealed to the market at such an early stage. As a result, 7 out of 70 periods qualify as bad cascade periods and 5 of these periods are from treatment sequence and 3 of them were already originally classified as cascade periods. Wilcoxon rank-sum tests do not allow us to reject the null hypothesis that the median of the quality of information aggregation in these bad cascade periods differs from the median in other periods. When redefining good cascade periods based on the time of the first transaction, 34 out of 70 periods qualify but none of the differences are significant (table 2.4).

2.5.3 Who Submits Orders First?

Instead of the time of the first transaction, the time at which a trader first submits a bid or an ask might be more closely related to the time at which he reveals his private information to the market. Therefore, we test whether the time of the first bid or ask is related to the time at which the private signal is received in treatment sequence. The Spearman rank correlation coefficient is 0.29 and we can safely reject the null that the 2 variables are unrelated (p=0.000). Instead of taking all bids and asks into account, we only consider bids above 20 and asks below 80 since any signal would justify lower bids or higher asks.

A reclassification of cascade periods based on the time of the first bid or ask yields 5 bad cascade periods, 4 of these periods already originally qualified as bad cascade periods. We also obtain 35 good cascade periods. Wilcoxon rank-sum tests do not allow us to reject the null hypothesis that the quality of information aggregation does

Table 2.5. Test Results for Cascade Periods Defined Based on the Time of the First Order.

Measure	Bad Cascade Periods vs. Other Periods Both Treatments	Good Cascade Periods vs. Other Periods Both Treatments
Mean Absolute Deviation Using Transactions During the Last 90 Seconds	Mean Cascade: 25.5 Mean Other: 30.0 p = 0.81	Mean Cascade.: 31.4 Mean Other: 28.1 p = 0.82
Mean Absolute Deviation Using Transactions During the Last 150 Seconds	Mean Cascade: 25.9 Mean Other: 31.1 p = 0.74	Mean Cascade: 30.5 Mean Other: 30.9 p = 0.79
Mean Absolute Deviation Using the Last 5 Transactions	Mean Cascade: 30.8 Mean Other: 30.0 p = 0.89	Mean Cascade: 30.8 Mean Other: 29.5 p = 0.95
Mean Absolute Deviation Using the Last 5 Units Traded	Mean Cascade: 31.4 Mean Other: 29.9 p = 0.79	Mean Cascade: 30.2 Mean Sequence: 29.9 p = 0.87

not differ between bad cascade periods and other periods. Good cascade periods also do not seem to yield a different quality of information aggregation (table 2.5).

2.6 Alternative Explanations

2.6.1 Early Expected Value vs. Late Expected Value

Even though subjects might not completely ignore their private information, they might place too much weight on information that the actions of other traders reveal. In that case, misleading early signals would still lead to a lower quality of information aggregation but not necessarily in such a clear-cut way as the cascade model suggests. To measure the extent to which early signals are misleading, we compute the expected value of the asset given the first 4 signals. We then take the absolute value of the difference between this early expected value and the expected value given all private signals. This variable (*devalue*) is then used to explain the quality of information aggregation. Using OLS, we estimate a coefficient for variable *devalue* (β) as well as an intercept (α) separately for treatment sequence and treatment baseline using all 35 periods for each treatment as observations. The results of these regressions are displayed in table 2.6.

In treatment sequence, the coefficient of *devalue* is always significant at the 10% level. The larger the difference between the early expected value of the asset and the late expected value of the asset, the larger the mean absolute deviation of prices from the late expected value. In that sense, misleading early signals do have an effect on the extent to which prices converge to the rational expectations equilibrium price.

Since *devalue* is correlated with the variance of the signals that traders receive, it could be that a higher variance of signals is the true cause for the observed worse quality of information aggregation in periods with high values of *devalue*. In that case, we would expect the coefficient of *devalue* to be significant in treatment baseline as well. Since that is not the case, we conclude that the sequence in which signals are released does indeed affect the extent to which markets can aggregate privately held information. We also test whether including a variable that measures the standard deviation of signals significantly improves the fit of these regressions. Wald tests do not allow us to reject the null hypothesis that the coefficient of the standard deviation of signals is zero.

2.6.2 Strategic Behavior

When we classify periods as cascade periods, we assume that subjects reveal their private information to other traders. If they fail to do so, the information that the market receives might not correspond to the information used to classify periods, which could explain why information aggregation in bad cascade periods is not substantially worse than in other periods. In order to test whether subjects are trying to mislead other traders, we examine the first order submitted in each period. At that time, the only information subjects have is their private signal. We only consider bids above 20 and asks below 80 that were made by traders who had already received their signal. 28% of such first orders are misleading in the sense that subjects are submitting a buy order even though they received a low signal or that they submit a sell order even though they received a high signal. Not all of these orders are in-

Table 2.6. Using Devalue to Explain Differences in the Quality of Information Aggregation.

Measure	Treatment	R^2	α (sd)	β (sd)	p-value Wald
Mean Absolute Deviation Using Transactions During the Last 90 Seconds	Baseline	0.06	23.21** (7.735)	0.25 (0.326)	0.30
	Sequence	0.14	24.66** (6.804)	0.44* (0.186)	0.48
Mean Absolute Deviation Using Transactions During the Last 150 Seconds	Baseline	0.07	27.10** (6.754)	0.26 (0.302)	0.27
	Sequence	0.17	23.70** (6.519)	0.47* (0.193)	0.62
Mean Absolute Deviation Using the Last 5 Transactions	Baseline	0.06	22.93** (7.985)	0.26 (0.335)	0.27
	Sequence	0.15	24.61** (6.337)	0.46* (0.191)	0.65
Mean Absolute Deviation Using the Last 5 Units Traded	Baseline	0.07	22.61** (8.103)	0.27 (0.324)	0.29
	Sequence	0.16	24.30** (6.248)	0.48* (0.186)	0.61

Coefficients marked by (* / **) are significant at the (10 / 5) percent level. Robust standard errors clustered by session are shown in parentheses.

consistent with the private signal received. For example, if a subject received a weak high signal, the expected value of the asset is 60. It is then perfectly reasonable to submit a sell order at a price of 70. However, 18% of all first orders are either sell orders at a price below the expected value given the trader's private signal or buy orders at a price above the expected value given the trader's signal. Clearly, other traders will find it hard to figure out what signals these traders had based on the bids or asks that they submitted. No matter whether these are intentional attempts to mislead other traders or simply mistakes, the fact that bids and asks do not always reflect the private information that a trader holds makes it difficult do identify the effect of the sequence of signals on the quality of information aggregation.

2.7 Conclusion

While we find evidence that the sequence in which signals are released to traders affects the quality of information aggregation, there is not much support for conjectures

1 and 2. Bad cascade periods in treatment sequence do not seem to fare substantially worse than bad cascade periods in treatment baseline or other periods in treatment sequence. A possible reason could be that we simply do not have enough observations. Another reason could be the difficulty involved in identifying the sequence in which information is released to the market. In fact, the sequence in which signals are released to traders does not always correspond to the sequence in which subjects actively trade or submit orders. Moreover, when they submit orders, they do not always reveal their private signal but might instead try to mislead other traders.

An alternative experimental design would eliminate these 2 sources of complexity while still preserving an endogenous price. Instead of allowing subjects to trade at any point of time, we could require them to trade in a predetermined sequence. Each subject would be able to submit as many sell and buy orders as desired but only once. As a consequence, subjects would no longer have an interest in misleading other traders since it would be impossible to capitalize on flawed prices by submitting further orders at a later point of time. At the same time, the sequence in which subjects trade would always correspond to the sequence in which they receive their private information. A control treatment would simply correspond to a call market with identical signals and value draws. By eliminating much of the complexity of a continuous double auction while preserving an endogenous price without a market maker, such a design should allow us to establish whether information cascades routinely occur in markets with an endogenous price.

Chapter 3

Wise Crowds or Wise Minorities?

This chapter is based on a paper written jointly with Jacob K. Goeree.

3.1 Introduction

When people are imperfectly informed they may try to learn from others' choices. Prospective graduate students, for instance, often inquire which schools other students chose to apply to. Teenagers consult the charts before deciding which CD to buy and tourists tend to prefer well-occupied restaurants to half-empty competitors. In many of these situations, only others' choices can be observed, not the exact information they had when making their choices. And while in some instances the exact sequence of predecessors' choices is observed (as in the US primary elections), more often only aggregate statistics based on those choices are available (e.g., the number of diners in a restaurant).

When do others' decisions contain relevant information and what course of action do they suggest? Obviously, predecessors' choices matter only when their payoffs are correlated to some extent. For this reason, most of the social-learning literature makes the simplifying assumption that agents have identical values for the available options, as is the case, for instance, when buying stocks.[1] In this common-value environment,

[1]Exceptions include Smith and Sørensen (2000) and Goeree et al. (2006, 2007) who study social learning with heterogeneous payoffs. Others' choices are also important when network externalities are present, e.g., using the same technology standard. See Hung and Plott (2000) and Drehmann et al. (2005b) for sequential decision-making experiments with informational and payoff externalities. Cipriani and Guarino (2005), Drehmann et al. (2005a), and Bose et al. (2009) consider situations

folk wisdom suggests it would be best to follow the majority, an intuition that is formalized by theoretical models of social learning. Bikhchandani et al. (1992), for example, consider a model where agents are privately informed about which of 2 options is better and the quality of information is the same across agents. They demonstrate that after a few decisions, information cascades occur and all agents herd on the majority choice regardless of their private information.[2] Following the majority is also optimal in Banerjee's (1992) model where uninformed agents have the ability to signal that they have no information.

In contrast, Callander and Hörner (2009) consider a situation where it can be optimal to follow the *minority*. In their model, agents differ in terms of the quality of information they possess and observe only the number of decisions for each option. In this paper, we consider the following simplified version of their model: each agent has a small chance of being perfectly informed about which of 2 options is correct or gets no information at all (besides the prior information that puts equal weight on both options). The information agents receive and the order in which they move are exogenously determined.[3] Finally, the probability of being informed is low enough such that, under common knowledge of rationality, it is always optimal to follow the minority. This result is explained in more detail below, but to glean some intuition consider an uninformed agent who learns that 2 predecessors have chosen restaurant A and one has chosen restaurant B. Such an outcome can only occur when the first agent was uninformed and chose the worse of the 2 restaurants. If both the second and the third agent were informed, it would be best to follow the majority but such a case is relatively unlikely when the probability of being informed is low. When only the second agent was informed, the third agent faces a tie and chooses randomly, in which case following the minority is no worse than following the majority. As opposed

where predecessors' choices affect payoffs by changing the prices of the alternatives.

[2]Laboratory evidence provides partial support for these predictions in the sense that information cascades do occur but are often broken. In addition, subjects tend to overweigh their private information vis-à-vis that contained in publicly observable predecessors' choices (see, e.g., Anderson and Holt, 1997; Çelen and Kariv, 2004a, 2004b; Goeree et al., 2006, 2007).

[3]Kübler and Weizsäcker (2004) present a study where the information available to decision makers is endogenous, while Ottaviani and Sørensen (2001) allow the sequence of decision makers to be endogenous.

to a sequence with 2 informed agents, it is quite likely that only the third agent was informed, in which case it is strictly better to follow the minority.

This simple logic extends to more general minority-majority divisions if common knowledge of rationality can be subsumed (see Callander and Hörner, 2009). But once we introduce the possibility of "trembles" or mistakes, it breaks down. Goeree et al. (2007), for example, find that in standard social learning experiments (based on Bikhchandani et al., 1992), cascades do form but almost never last as subjects frequently opt to follow their contrary information and break the cascade.[4] While uninformed agents in our experiment do not possess any private information,[5] trembles may still occur especially because the information conveyed by predecessors' choices may be of low quality and value. Intuitively, the possibility of trembles greatly alters equilibrium predictions in the Callander–Hörner setup. In the example above, for instance, the 2–1 division between restaurants is more likely caused by a trembling uninformed agent than by a deviating informed agent when the probability of being informed is very low.

More generally, whether it is optimal to follow the majority or the "deviant minority" therefore depends on the likelihood of mistakes, the quality of others' information, the correlation of tastes, etc. It would be hard to distinguish these confounding elements in data from the field, which is why we turn to the lab. We conducted 2 types of sequential decision-making experiments: in treatment "sequence," agents can see the entire sequence of predecessors' decisions and in treatment "no-sequence" they only see the number of predecessors' choices for either option. Collecting data from both treatments allows us to connect our findings to prior literature, which mostly employs the sequence treatment, and enables us to evaluate the efficiency gains that may result from the additional information in treatment sequence.

[4] In these experiments, cascade breaks are informative and prevent the learning process from getting stuck. As a result, full information aggregation becomes possible in the limit as the number of agents grows large. Goeree et al. (2007) demonstrate how a logit quantal response model can account for much of the dynamics in the experiments.

[5] Previous social learning experiments have documented the tendency of subjects to overweigh their private information. This is not possible in our design, since subjects get either no information or information that is perfectly informative.

We find that informed subjects always follow their signal, i.e., they always pick the correct option. On average, uninformed subjects tend to follow unanimous predecessors close to 90% of the time both in treatments sequence and no-sequence. Furthermore, the frequency with which unanimous predecessors are followed significantly rises (to levels between 90% and 100%) as the number of predecessors grows. This high percentage of rational choices is maybe not surprising given that the decision problem faced by an uninformed agent is relatively easy when all predecessors agree. When there is a deviator in treatment sequence, subjects tend to follow the deviator only 72% of the time. The frequency with which a deviator is followed is significantly higher when the deviator's choice belongs to the majority (80%) than when it belongs to the minority of previous choices (58%). Finally, in treatment no-sequence, subjects tend to follow the minority only 28% of the time. This percentage significantly declines when the difference between the number of majority and minority choices grows.

While observed choices deviate from theoretical predictions, they are approximate best responses to the empirical distribution of play in the following sense. Given the choices of others, and given the signals used in the experiment, the cost of not following unanimous predecessors in the experiment is \$1.53 on average. Likewise, the cost of not following a deviator in treatment sequence is \$1.23 on average, and the cost of not following the minority is −\$0.36 on average. In other words, subjects who are (imperfect) profit maximizers would nearly always follow unanimous predecessors, would more likely than not follow a deviator in treatment sequence (although not as frequently as they would follow unanimous predecessors), and would follow the *majority* in treatment no-sequence. The canonical model that captures this type of imperfect maximization behavior is the (logit) QRE model. We show that, on an aggregate level, logit-QRE is able to reproduce the main features of our data quite well.

In the logit-QRE model, however, agents are assumed to be *ex ante* symmetric, which is clearly not true in our data. While some subjects make rational choice in all ten periods of the experiment, others do so in less than half the periods. This type of heterogeneity is captured by models that allow for different levels of strategic thinking, such as the level-k model (e.g., Crawford and Iriberri 2007a, 2007b) and cognitive hierarchy (Camerer, et al. 2004). When we apply level-k and cognitive hierarchy to the data, we find they produce a worse fit both on an individual and aggregate level. The reason for their poor performance is that these models subsume best response behavior (given beliefs), except for level-0 who randomizes when uninformed. The best-response assumption often conflicts with intuitive comparative statics observed in the data, e.g., subjects tend to follow unanimous predecessors more frequently when the group of predecessors is large. In addition, the best-response assumption implies that a subject who mostly but not always makes a rational choice, is classified as level-0 even though most of her choices suggest a higher level of thinking.

Goeree and Holt (2004) propose a "noisy introspection" model that blends the idea of different levels of strategic thinking with noisy responses (trembling). In particular, the noisy introspection model replaces the strict best responses of the level-k model with logit "better responses." Importantly, agents in the noisy introspection model are assumed to be aware that others tremble. For example, when computing the probability that the minority choice is correct in treatment no-sequence, agents take into account the possibility that the minority arose because of trembles. As a result, the model can predict why subsequent choices favor the majority (not the minority) even for agents with high levels of strategic thinking.

We find that the noisy introspection model provides a significant improvement in fit relative to logit-QRE and a dramatic improvement relative to level-k and cognitive hierarchy. To illustrate the importance of the "common-knowledge-of-trembling" assumption that underlies noisy introspection, we also estimate a noisy version of the level-k model in which agents tremble but assume others do not. We find that the noisy level-k model provides a better fit than level-k and cognitive hierarchy, but does not do nearly as well as noisy introspection. We also estimate 2 versions of the cognitive hierarchy model with trembles, one in which agents are aware that others tremble and one in which they are not. Both of these models fit the data equally well as the noisy introspection model.

The chapter is organized as follows. In the next section, we briefly discuss the theoretical predictions for the 2 treatments. The experimental design is presented in section 3.3. The results of the experiment can be found in section 3.4. In section 3.5, we apply alternative models of bounded rationality to explain individual and aggregate outcomes. Section 3.6 concludes.

3.2 Theoretical Predictions

Treatment sequence is a simple variant of the social learning model proposed by Bikhchandani et al. (1992). There are 2 options, A and B, that are equally likely to be correct and a finite set of n agents labeled $t = 1, 2, \ldots, n$. Each agent chooses either A or B after having observed a private signal s_t and the decisions of predecessors. Signals in the experiment are either fully informative or not informative at all: if ω denotes the correct option then $s_t = \omega$ with probability $q > 0$ and $s_t = \emptyset$ with probability $1 - q$.

Given that some agents may be fully informed, the perfect Bayesian equilibria of sequence treatment are easy to derive. First, an agent with $s_t = \omega$ chooses ω. Second, if all predecessors agree, then an uninformed agent follows the majority since either all predecessors were uninformed and the agent is indifferent or some predecessors were informed and the agent strictly prefers to follow the majority. Third, if predecessors were not unanimous, i.e. choices switched from one option to the other, then the first predecessor who "deviated" by not following her predecessors must have been informed. In this case, the agent should follow the deviator. Note that all 3 cases can be succinctly summarized as follows: under common knowledge of rationality, uninformed agents follow their immediate predecessor. Finally, if predecessors switched from one option to the other and then switched back, play is off the equilibrium path. In this case, the agent cannot infer anything from prior play and simply randomizes when $s_t = \emptyset$ and chooses ω when $s_t = \omega$.

In treatment no-sequence, informed agents with $s_t = \omega$ again choose ω. Also, uninformed agents simply follow the majority if all predecessors agree. The surprising result for treatment no-sequence is that uninformed agents may be better off following the *minority*, a result due to Callander and Hörner (2009). To glean some intuition, consider their illustrative example in which uninformed agent $t = 4$ has to choose after 2 predecessors have chosen option A and 1 predecessor has chosen option B. In this case, agent $t = 1$ must have been uninformed and must have chosen the wrong option since predecessors would otherwise have been unanimous. Furthermore, either (i) agent 2 was uninformed in which case she followed agent 1 and agent 3 was informed and choose differently from her predecessors, or (ii) agent 2 was informed and choose differently from agent 1 and agent 3 was uninformed and randomly followed agent 1, or (iii) both agents 2 and 3 were informed and chose differently from agent 1. Under scenario (i) the minority choice is correct, under (ii) both options are equally likely to be correct, and under (iii) the majority choice is correct. For $q < \frac{1}{2}$, the situation in which 2 agents are informed is less likely, and, hence, the minority is more likely to be correct. A simple calculation shows that for $q \in (0, 1)$,

$$P_{1,2}(\text{minority is correct}) \;=\; \frac{2-q}{3}. \tag{3.1}$$

Moreover, for small q, the probability that the minority is correct *grows* as the majority increases in size. For example, consider the case where 3 predecessors have chosen option A and 1 predecessor has chosen option B. The fourth agent must have faced (i) unanimous predecessors in which case the minority choice B is correct (if option A were correct the fourth agent should pick A whether or not informed), or (ii) a 2–1 majority for A in which case the minority is wrong (now if option B were correct the fourth agent should pick B whether or not informed, resulting in a tie). Note that situation (i) only requires the fourth agent to be informed while situation (ii) requires the fourth agent and at least 1 other agent to be informed. The chance that the minority is correct is now

$$P_{1,3}(\text{minority is correct}) \;=\; \frac{2(1-q)^2}{2(1-q)^2 + q(1+q)}, \tag{3.2}$$

and it is straightforward to verify that $P_{1,3} > P_{1,2}$ for small q. The minority wisdom becomes even stronger as the majority grows larger than 3: the outcome in which $m \geq 4$ predecessors chose A and only 1 predecessor chose B requires at least $m-1$ agents to be informed when the majority choice A is correct, but requires only the final agent to be informed when the minority choice B is correct. Furthermore, this logic extends to minorities of sizes other than 1.

Figures 3.1 and 3.2 and show the relevant probabilities for the experimental setup discussed below: Figure 3.1 corresponds to the case where the probability of being informed is $q = 0.2$ and the total number of agents is $n = 7$, and figure 3.2 corresponds to the case where $q = 0.1$ and $n = 13$. In figure 3.1, the number of predecessors (on the x-axis) is at most 6, and the minority size can be either 0, 1, or 2, as indicated by the labels next to the 3 lines that represent the probabilities the minority is correct for these cases. A minority of size 0 means that predecessors were unanimous in which case the minority choice is more likely to be incorrect, as indicated by the decreasing line that is everywhere below 0.5. The first 2 points of the line labeled "1" can be computed from (3.1) and (3.2) above for $q = 0.2$: note that $P_{1,m}$ for $m = 2, 3, 4, 5$ is everywhere above 0.5, i.e. the minority is more likely to be correct, and is increasing in m, i.e. the minority wisdom becomes stronger as the majority grows. The same is true for $P_{2,m}$ for $m = 3, 4$ as shown by the line labeled "2." Figure 3.2 establishes the same properties for our second treatment in which $q = 0.1$ and $n = 13$.

3.3 Experimental Design

The experiments were conducted in the Social Science Experimental Laboratory (SSEL) at Caltech using undergraduate and graduate students as subjects. Each subject was allowed to participate only once. For each session, we randomly determined the correct option in each of the ten periods. And for each period, we randomly determined for each subject whether she was informed or not. Since the specific sequence of informed and uninformed agents can affect efficiency and behavior, we used the same draws for both treatment sequence and treatment no-sequence. We ran ten sessions for both treatments.

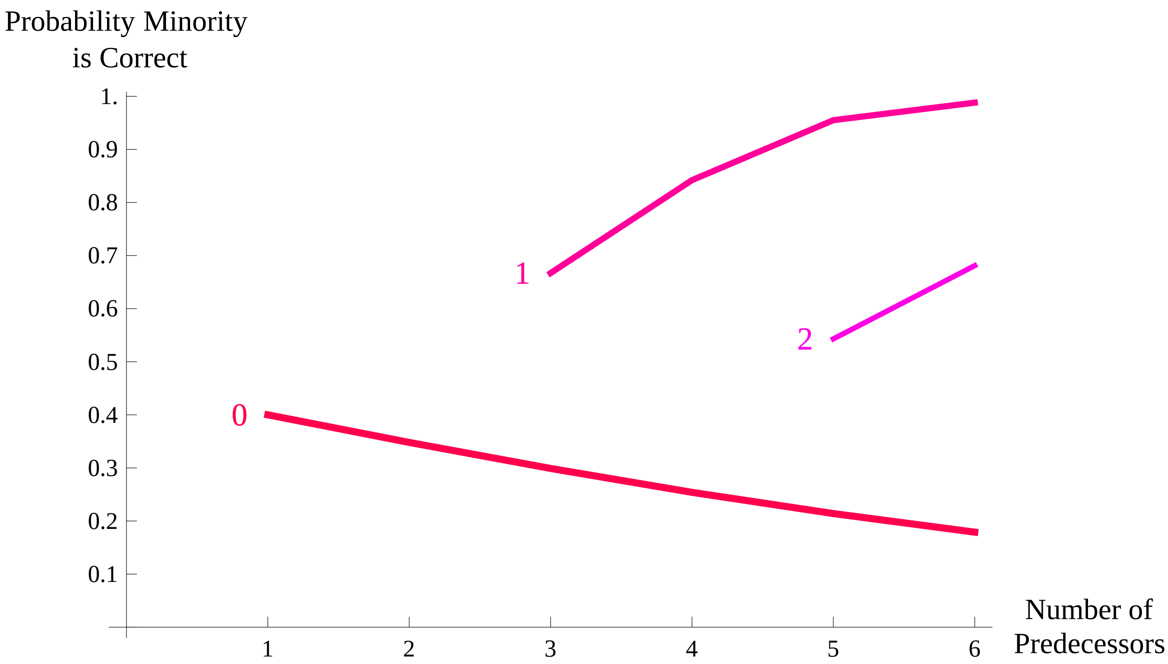


Figure 3.1. Probability Minority is Correct in Treatment No-Sequence when $q = 0.2$ as a Function of the Number of Predecessors (x-axis) and the Minority Size (Label Next to a Line).

The experiments were run by hand. After reading the instructions out loud, subjects had an opportunity to ask clarifying questions. We used an urn to select the subject who had to guess first. The experimenter then went to that subject's seat and indicated on the subject's record sheet whether that subject was informed, and, if so, what the correct option was. Then another draw without replacement determined which subject had to guess next, etc. Each time, the experimenter revealed the correct option only to informed subjects. In addition, all subjects, informed or uninformed, were told the sequence of predecessors' choices in treatment sequence and the number of choices for both options in treatment no-sequence.[6]

[6]To avoid a bias for one of the options, we denoted option A with a circle ($\circ$) and option B with a cross ($+$) since these symbols have no obvious ordering. Also, there was no practice period that

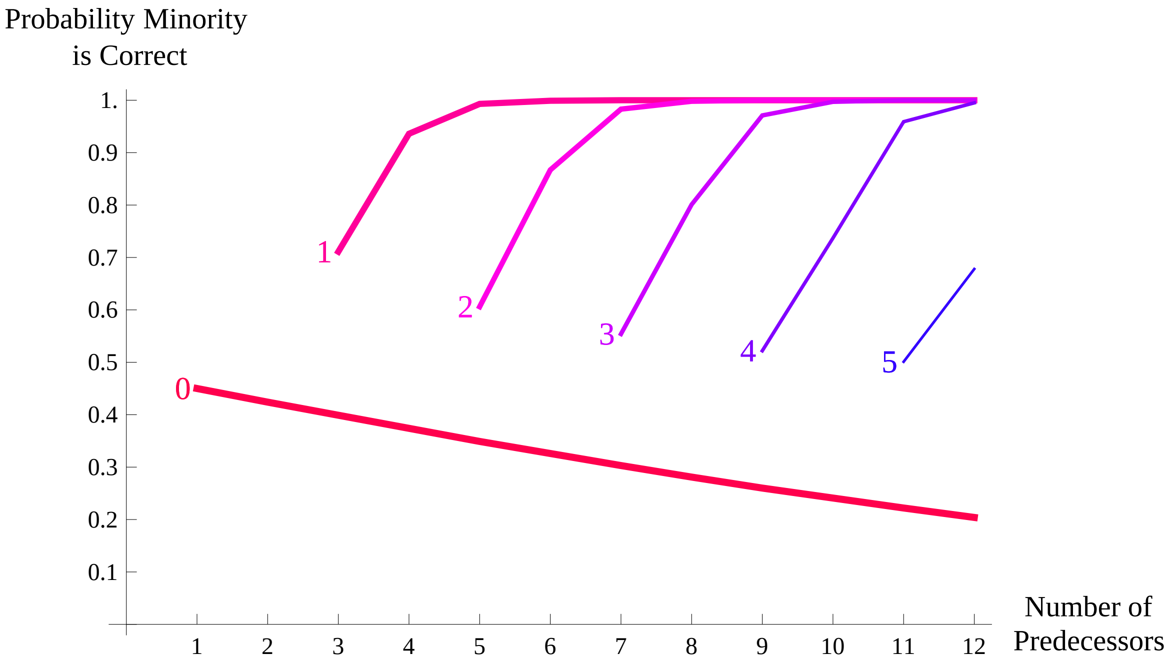


Figure 3.2. Probability Minority is Correct in Treatment No-Sequence when $q = 0.1$ as a Function of the Number of Predecessors (x-axis) and the Minority Size (Label Next to a Line).

The main design parameters are the probability that an agent is informed (q) and the number of subjects participating in a session (n). We chose q and n such that it was optimal to follow the minority in treatment no-sequence in all possible situations (assuming common knowledge of rationality). For $q = 0.1$ the maximum number is $n = 13$ and for $q = 0.2$ the maximum number is $n = 7$, see table 3.1.[7] At the end of each period, the correct option was revealed to everyone and subjects earned \$4 if

might have allowed subjects to coordinate their guesses at no cost. To test whether these measures were sufficient to avoid a bias, we considered the 143 cases were both options had the same number of choices in treatment no-sequence. In 82 of these cases, subjects chose A. Using a 2-tailed test and a 5% confidence level, we cannot reject the null hypothesis that the number of choices of A follows a binomial (143, 0.5) distribution.

[7] More generally, the larger q, the lower is the largest n for which following the minority is optimal (Callander and Hörner, 2009).

Table 3.1. Experimental Design Parameters and Subjects' Earnings.

Treatment	Prob Informed (q)	# Subjects (n)	# Sessions	Earnings
Sequence	0.1	12-13	5	$31.3
Sequence	0.2	7	5	$29.8
No-sequence	0.1	11-13	5	$27.1
No-sequence	0.2	7	5	$26.5

they had picked the correct option (and 0 otherwise). At the end of the experiment, subjects were paid their cumulative earnings plus a $5 show-up fee in cash. Average earnings for the different treatments are shown in the rightmost column of table 3.1. Sessions with 13 subjects usually took about 1 hour while sessions with only seven subjects typically lasted about 35 minutes.

3.4 Results

We first discuss the extent to which subject behavior conforms to theoretical predictions. We pool the data from the $q = 0.1$ and $q = 0.2$ sessions since there were no significant differences between them.

3.4.1 Subject Behavior

Not surprisingly, informed subjects in the experiment always follow their signals, and, hence, always select the correct option. In contrast, uninformed subjects do not always choose according to their optimal strategy under common knowledge of rationality. We differentiate among the following situations:

(S1) all predecessors chose the same option.

(S2) 1 predecessor deviated from his unanimous predecessors in treatment sequence.[8]

(S3) some option is chosen by a minority of predecessors in treatment no-sequence.

[8]We are not considering situations in which 2 or more agents deviated from their immediate predecessor. In these situations, the assumption of common knowledge of rationality is clearly violated.

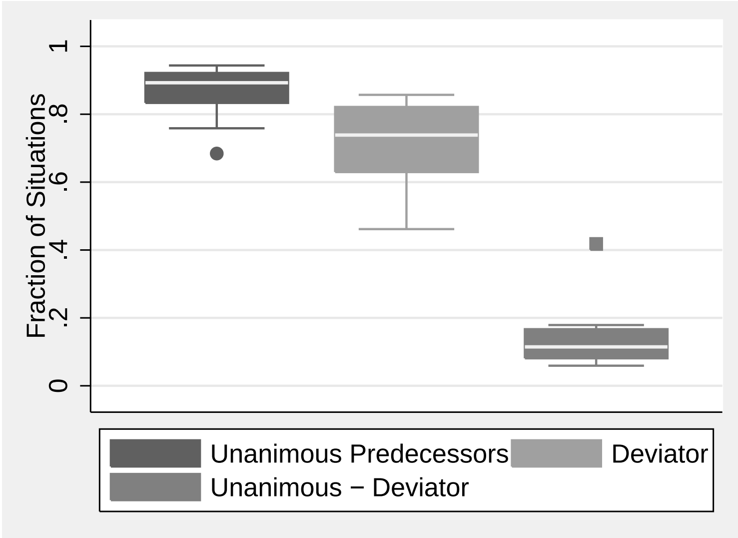


Figure 3.3. Subject Behavior in Treatment Sequence.

As explained above, rational agents would follow unanimous predecessors in both treatments, a deviator in treatment sequence, and the minority in treatment no-sequence. Figure 3.3 summarizes the behavior of uninformed subjects in treatment sequence. For each of the ten sessions we compute the fraction of situations in which subjects follow their optimal strategy under common knowledge of rationality. The box plot summarizes these ten independent data points.[9]

The leftmost box in Figure 3.3 summarizes behavior in treatment sequence when all predecessors chose the same option. The median frequency with which uninformed subjects follow their unanimous predecessors (across ten sessions) is 0.89. In other words, 89% of the time, subject behavior coincides with the optimal strategy (under common knowledge of rationality) for this specific situation. The middle box concerns situations in which 1 predecessor deviated from his unanimous predecessors. The median frequency with which uninformed subjects follow such a deviator is 74%. The rightmost box summarizes the difference between the frequency with which subjects follow unanimous predecessors and the frequency with which they follow a deviator.

[9]The boundaries of the box correspond to the first and third quartile, the line within the box marks the median. Observations that are more than 1.5 times the interquartile range higher (lower) than the third (first) quartile are considered to be outliers and shown separately. The whiskers show the position of the lowest (highest) observation that is not an outlier.

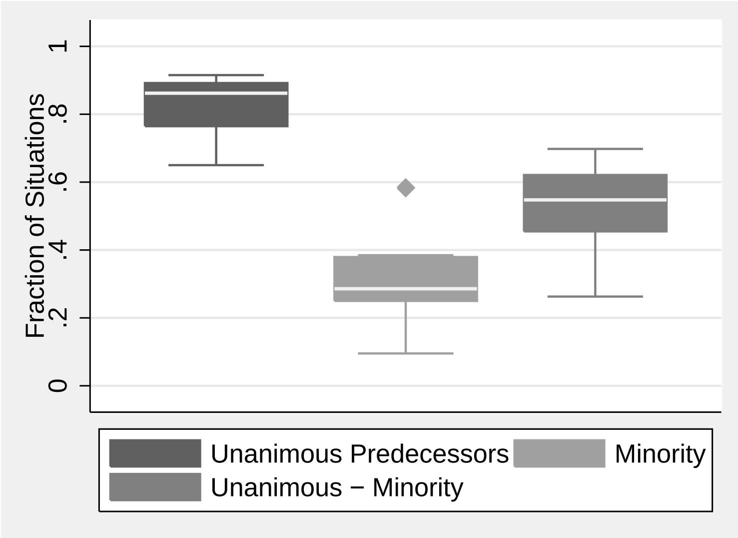


Figure 3.4. Subject Behavior in Treatment No-Sequence.

In each one of ten sessions, this difference is positive. Treating each session as an independent observation, this difference is therefore significant using any conventional test. Subjects in our experiments are more likely to follow unanimous predecessors than to follow a deviator in treatment sequence.

Figure 3.4 summarizes the behavior of uninformed subjects in treatment no-sequence. As in Figure 3.3, the leftmost box summarizes behavior when subjects face unanimous predecessors. Not surprisingly, the results are very similar to those reported for treatment sequence (since there is no reason why subjects should behave differently in the 2 treatments when all predecessors chose the same option). The median frequency with which subjects follow unanimous predecessors is now 86%. The middle box concerns situations in which there is a minority choice among predecessors. Obviously, behavior deviates very strongly from the optimal strategy under common knowledge of rationality: subjects only follow the minority about 29% of the time (median over all ten sessions). In fact, in all but one session, subjects follow the majority *more* frequently than the minority. Finally, for each of the ten sessions of treatment no-sequence, we compute the difference between the frequency with which subjects follow unanimous predecessors and the frequency with which they follow the minority. The results are captured by the rightmost box in Figure 3.4. Subjects

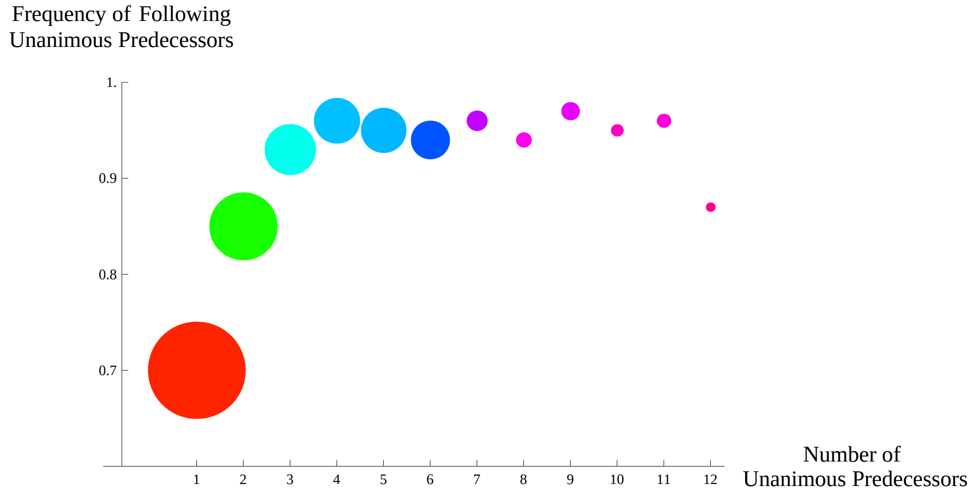


Figure 3.5. Frequency with Which Unanimous Predecessors Are Followed as a Function of the Number of Unanimous Predecessors.

in our experiments are clearly more likely to follow unanimous predecessors than to follow the minority.

Summing up our findings for the 3 situations:

(S1) While subjects do not always follow their unanimous predecessors, they do so more than 50% of the time in all 20 sessions.[10] Furthermore, the tendency of subjects to follow their unanimous predecessors becomes stronger as the number of unanimous predecessors grows. Figure 3.5 shows the frequency with which unanimous predecessors are followed when the number of predecessors varies from 1 to 12. The size of each circle reflects the number of occurrences. A simple Probit regression in which the dependent variable is the choice to follow unanimous predecessors, P_{follow}, and the independent variable is the number of predecessors, n, results in significant estimates: $P_{\text{follow}} = \Phi(\beta_0 + \beta_1 \cdot n)$ yields $\beta_0 = 0.7(0.1)$ and $\beta_1 = 0.12(0.04)$, where the numbers in parentheses denote the robust standard errors (clustered by subject).

(S2) In nine out of ten sessions, subjects follow a deviator in treatment sequence

[10] We can easily reject the null hypothesis that subjects follow their unanimous predecessors in 50% of all instances using a Wilcoxon signed-rank test ($p < 0.001$).

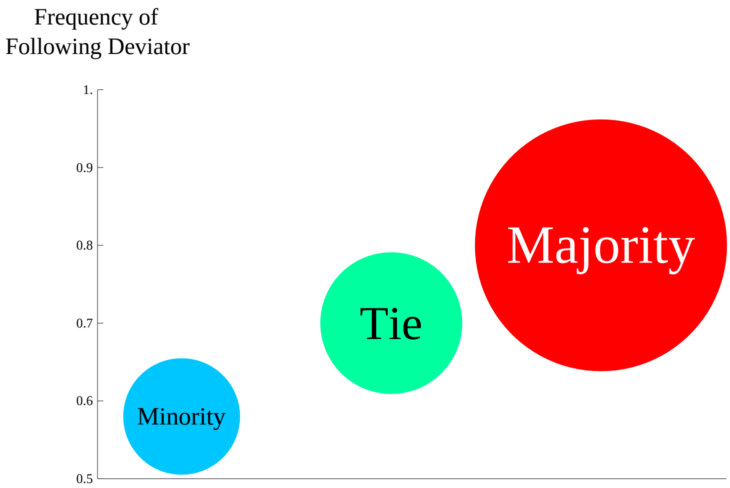


Figure 3.6. Frequency with Which a Deviator's Choice Is Followed when it Is in the Minority, Creates a Tie, or Is in the Majority.

more than 50% of the time.[11] Furthermore, the tendency to follow a deviator is stronger if the choice of the deviator coincides with the majority choice. Figure 3.6 shows the frequency with which a deviator is followed when the deviator's choice is in the minority (left circle), when there is a tie (middle circle), and when the deviator's choice is in the majority (right circle). The size of each circle reflects the number of occurrences. To verify whether the observed differences are statistically significant, we estimate a simple Probit model in which the chance of following a deviator, P_{follow}, is explained by whether there is minority, tie, or majority: $P_{\text{follow}} = \Phi(\beta_{\text{min}} + \beta_{\text{tie}} + \beta_{\text{maj}})$ yields $\beta_{\text{min}} = 0.20(0.23)$, $\beta_{\text{tie}} = 0.34(0.31)$, and $\beta_{\text{maj}} = 0.65(0.29)$, where the numbers in parentheses denote the robust standard errors (clustered by subject). The difference between β_{maj} and β_{min} is significant at the 5% level.

(S3) In all but one session of treatment no-sequence, subjects follow the majority more frequently than the minority.[12] Furthermore, subjects are increasingly *less* inclined to follow the minority the smaller the minority relative to the majority.

[11] The null that subjects follow a deviator half of the time in treatment sequence can be rejected using a Wilcoxon signed-rank test ($p = 0.004$).

[12] The null that subjects are equally likely to follow the minority or the majority in treatment no-sequence is rejected using a Wilcoxon signed-rank test ($p = 0.004$).

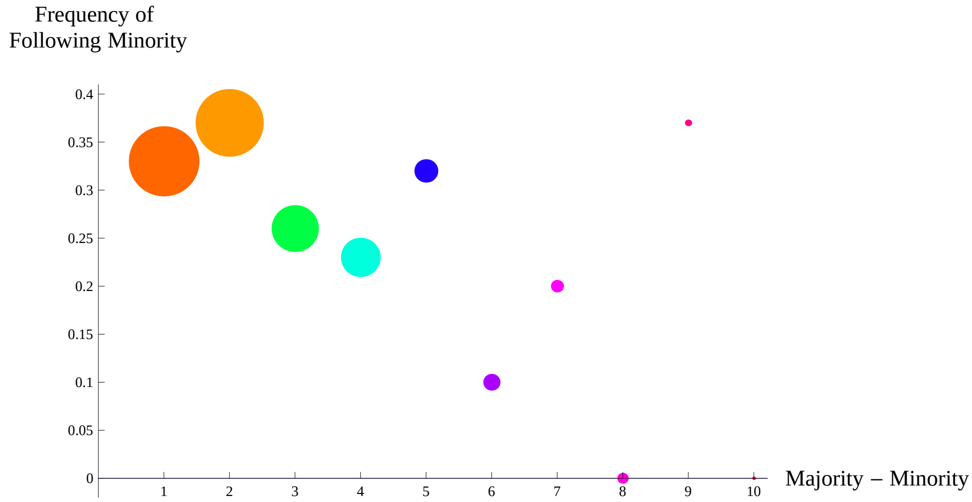


Figure 3.7. Frequency with Which the Minority Is Followed as a Function of the Difference between Majority and Minority Size.

> Figure 3.7 shows the frequency with which subjects follow the minority as a function of the difference between the size of the majority and the size of the minority. The size of each circle reflects the number of occurrences. A Probit analysis in which the chance of following the minority, P_{follow}, is explained by the difference between the size of the majority and the size of the minority, Δn, results in significant estimates: $P_{\text{follow}} = \Phi(\beta_0 + \beta_1 \cdot \Delta n)$ yields $\beta_0 = -0.26(0.14)$ and $\beta_1 = -0.098(0.039)$, where the numbers in parentheses denote the robust standard errors (clustered by subject).

The observed choice frequencies across the 3 situations can be ranked as follows:[13]

$$\text{P(follow unanimous predecessors)} > \text{P(follow deviator)} > \frac{1}{2} > \text{P(follow minority)}. \tag{3.3}$$

In other words, subjects are more likely to follow unanimous predecessors than a deviator , are more likely to follow unanimous predecessors than the minority, and are more likely to follow deviators than the minority. It is interesting to compare the

[13]Using a Wilcoxon matched-pairs signed-rank test, we can reject the null that subjects follow unanimous predecessors equally frequently as a deviator in treatment sequence ($p = 0.002$). Likewise, we can reject the null that subjects are equally likely to follow unanimous predecessors as the minority in treatment no-sequence ($p = 0.002$) and the null that subjects follow deviators equally frequently as the minority ($p < 0.001$ using a Wilcoxon rank-sum test).

ranking of choice frequencies with that of the associated average payoff differences, computed using the actual draws and choices in the experiment:

$$\Delta\pi(\text{unanimous predecessors}) > \Delta\pi(\text{deviator}) > 0 > \Delta\pi(\text{minority}),$$

where $\Delta\pi$ measures the average payoff difference between following and not following for each one of the 3 situations. To summarize, given the observed play of others, not following a deviator is a costly mistake ($1.23), but it is less of a mistake than not following unanimous predecessors ($1.53). Furthermore, not following the minority is *not* a mistake, i.e. in the no-sequence sessions, following the majority yields higher earnings ($0.36).

3.4.2 Efficiency

For the specific draws used in the experiment, the theoretically expected fraction of correct choices is 78% in treatment sequence and 72% in treatment no-sequence.[14] Actual efficiency was 69% in treatment sequence and 62% in treatment no-sequence. Finally, the predicted fraction of correct choices when uninformed agents choose randomly and informed agents choose correctly is 58%.

Treating each session as an independent observation ($n = 10$), we test the null hypothesis that the distribution of the fraction of correct choices is identical under both treatments. A Wilcoxon matched-pairs signed-rank test allows us to reject the null hypothesis ($p = 0.039$). Therefore, efficiency in treatment sequence is significantly higher than in treatment no-sequence. We also test whether observed efficiency is significantly higher than when uninformed subjects choose randomly and informed subjects choose correctly. Using the same test, we can reject the null for treatment sequence ($p = 0.006$) but not for treatment no-sequence ($p = 0.432$). This result suggests that information about predecessors' choices improves efficiency only when the entire sequence of decisions is known – subjects are not able to improve their decisions

[14] We computed the fraction of correct choices separately for each session and then took the average across sessions.

significantly when observing only the number of prior choices for each option.

3.5 Explaining Subject Behavior

In this section, we analyze the observed deviations through the lens of alternative models of bounded rationality, including logit-QRE as well as 2 models in which agents exhibit different levels of strategic thinking (level-k and cognitive hierarchy). We find that models that combine heterogeneity in strategic thinking with error-prone behavior (e.g., a noisy introspection model, Goeree and Holt, 2004) fit our data best.

3.5.1 Logit-QRE

Consider again the example discussed in Section 3.2, where 2 predecessors chose option A and 1 predecessor chose option B. Whether the minority or majority is more likely to be correct depends on whether the sequence {uninformed, uninformed, informed}, for which the minority is correct, is more likely to occur than the sequence {uninformed, informed, informed}, for which the majority is correct. For $q < 0.5$, subjects should follow the minority under common knowledge of rationality. If subjects tremble, however, then the minority is not necessarily correct even for sequence {uninformed, uninformed, informed} since the second uninformed subject may not have followed the first one (such a tremble is not unlikely since the associated payoff loss is small); following the majority may be better as a result.

The above intuition can be formalized using a logit-QRE model in which subjects' choice probabilities are positively but not perfectly related to expected payoffs, i.e. subjects are "better responders" not necessarily "best responders." For subject t, let H_t denote the profile of predecessors' choices and s_t the subject's signal. Then the probability that subject t chooses c_t is

$$P(c_t|H_t, s_t) = \frac{1}{1 + \exp(\lambda\pi(1 - 2P(\omega = c_t|H_t, s_t))}, \tag{3.4}$$

where λ is a "rationality" parameter that determines how sensitive choice probabilities

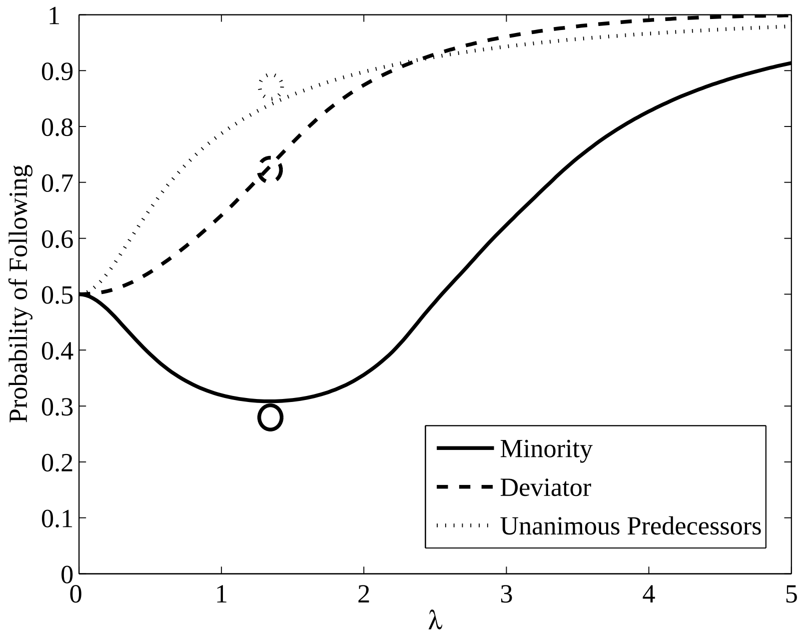


Figure 3.8. Uninformed Subjects' Behavior as a Function of λ in a Logit QRE Model for Situations that Occurred in the Experiment.

are with respect to expected payoffs and $\pi = \$4$ is the payoff of picking the correct option. Note that, since $\lambda\pi > 0$, agents more often choose the option that is more likely to be correct.

The curves in Figure 3.8 show the predicted probability of following unanimous predecessors, a deviator, or the minority respectively for different values of λ (to generate this figure we averaged predicted logit choice probabilities over all occurrences of the particular situation in the experiment). For low levels of λ, logit-QRE predicts that uninformed subjects are more likely to follow the majority than to follow the minority. The predicted probability of following a deviator is higher than 0.5 for all λ, and the same is true for the predicted probability of following unanimous predecessors.

Using maximum-likelihood techniques (treating each decision as an independent observation) we estimate $\lambda = 1.4$ (0.1) for treatment no-sequence, $\lambda = 1.3$ (0.1) for treatment sequence, and $\lambda = 1.3$ (0.1) for the pooled data, where the number in parentheses denotes the robust standard error (clustered by subject). The estimates are listed in the topmost panel of table 3.2 below (second column labeled "λ"), together with the associated likelihood (third column labeled "$\mathrm{LogL_{obs}}$"), the likelihood that results if all uninformed subjects chose randomly (fourth column labeled "$\mathrm{LogL_{random}}$"), and the best-possible likelihood (fifth column labeled "$\mathrm{LogL_{best}}$") that results by using the observed fraction for either option as the predicted fraction (separate for each different situation that occurred in the experiment).[15] The rightmost column in table 3.2 labeled "% Explained" provides a "goodness-of-fit" defined as follows:

$$\%\ \text{Explained} = \frac{\mathrm{LogL_{obs}} - \mathrm{LogL_{random}}}{\mathrm{LogL_{best}} - \mathrm{LogL_{random}}} \times 100\%.$$

The circles in Figure 3.8 show the average frequency with which subjects in the experiment follow unanimous predecessors, a deviator, or the minority respectively. The data averages are shown at the pooled estimate $\lambda = 1.3$ to facilitate comparison with logit predictions. Obviously, Logit-QRE can explain the qualitative patterns of our data as summarized in (3.3) extremely well.

3.5.2 Subject Heterogeneity

The homogeneous logit-QRE model assumes that agents are *ex ante* symmetric, i.e. they are equally likely to make mistakes. To test whether this is a reasonable assumption, we compute the fraction of rational choices for each subject by treatment.[16] If all subjects were equally rational, this fraction would not vary much across subjects.

Figure 3.9 illustrates the distribution of rational choices separately for treatments sequence (left) and no-sequence (right). Clearly, there is substantial subject het-

[15] We treat 2 situations as different if and only if at least one of the models we estimate makes a different prediction for the 2 situations. More specifically, (1) we distinguish between informed and uninformed agents, but all situations informed agents encounter are considered equivalent, (2) the situation where n (m) predecessors have chosen A (B) is considered equivalent to the situation where m (n) predecessors have chosen A (B), and (3) situations with an equal number of choices for both options are equivalent.

[16] We only count choices made by uninformed subjects. Moreover, only situations in which there is a rational and an irrational choice are considered. Hence, we drop choices made by first agents in the sequence. We also drop choices made by agents who face the same number of predecessors' choices for both options. In treatment sequence, we drop ("off-the-equilibrium-path") choices made after 2 or more predecessors deviated from their immediate predecessor's choice.

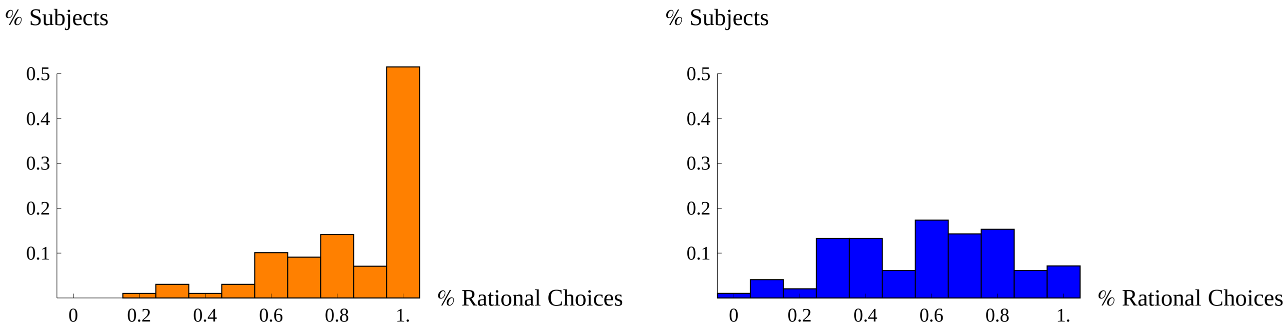


Figure 3.9. Fraction of Rational Choices in Sequence (left) and No-Sequence (right).

erogeneity. While the standard deviation of the fraction of rational choices is very similar in both treatments (0.20 in sequence and 0.21 in no-sequence), the shape of the distribution is quite different. In treatment sequence, a fairly high proportion of subjects make choices that are always consistent with their optimal strategy (under common knowledge of rationality), but this is not the case in treatment no-sequence.

3.5.3 Level-k and Cognitive Hierarchy

To account for the observed heterogeneity, we next consider a “level-k model” (e.g., Crawford and Iriberri, 2007a, 2007b) that allows for different levels of strategic thinking. We first describe how different types are defined in this model and then discuss the model’s predictions and the estimation of the type distribution.

In the level-k model, type k best responds believing others are of type $k-1$. We assume that type 0 chooses randomly when uninformed and follows her signal when informed.[17] In treatment sequence, an uninformed type 1 then follows the majority of predecessors, if all others are of type 0, the majority is more likely to be correct since type 0 picks the correct option when informed. Uninformed type 2 follows unanimous predecessors, but since type 1 deviates from the majority only

[17] Under the assumption that type 0 chooses randomly even when informed, type 1 would choose randomly when uninformed but would follow his signal when informed. Therefore, type 1 would correspond to type 0 under our assumptions. The only change to the model would be an introduction of an additional type that can only rarely be distinguished from type 1 (since informed agents are rare). Moreover, informed agents in our experiments always follow their signal.

when informed, uninformed type 2 follows such deviations. Note that it is possible that type 2 observes a sequence that is inconsistent with her beliefs[18] in which case type 2 is assumed to randomly pick an option when uninformed and to follow her signal when informed. Uninformed type 3 behaves the same as type 2 except that there are more situations in which type 3's beliefs are contradicted, and, hence, type 3 randomizes more often.[19] All higher types (4 and above) are identical to type 3.

In treatment no-sequence, uninformed type 1 follows the majority of predecessors. Uninformed type 2 follows unanimous predecessors, and since type 1 only deviates from the majority when informed, uninformed type 2 follows the minority when predecessors are not unanimous. Callander and Hörner (2009) show that it is optimal to follow the minority when all others do, so uninformed types 3 and higher are identical to type 2. In treatment no-sequence, there are no situations in which beliefs about lower-types' behavior are contradicted.

To facilitate comparison with the one-parameter logit-QRE model, we assume that types follow a Poisson distribution with parameter τ, which we truncate at the highest type that can be distinguished (2 in treatment no-sequence and 3 in treatment sequence).[20] Figure 3.10 summarizes the predictions of the level-k model for the 3 situations of interest for different levels of τ. For low τ, the level-k model is able to reproduce the main features of our data as summarized in (3.3) although the fit is not as good as that of logit-QRE (cf. Figure 3.8).

We estimate τ separately for treatment sequence and treatment no-sequence by maximizing

$$L(\tau) = \prod_{t=1}^{n} \sum_{k=0}^{m} \frac{\frac{e^{-\tau}\tau^k}{k!}}{\sum_{v=1}^{m} \frac{e^{-\tau}\tau^v}{v!}} \operatorname{Prob}(c_t \,|\, \text{type} = k), \tag{3.5}$$

where n is the number of subjects, m the number of different types (2 in treatment

[18]For instance, if the majority choice switches from A to B after a deviation, then a next deviation (back to A) would violate the assumption that all other agents are of type 1.

[19]Observing 2 or more deviations from the immediate predecessors' choice violates the assumption that all agents are of type 2, but if these deviations occur in favor of the majority's choice then they are compatible with the assumption that all other agents are of type 1. Therefore, the set of situations that violate the assumptions made by type 2 is a strict subset of the set of situations that violate the assumptions made by type 3.

[20]We also estimated the type distribution nonparametrically, see footnote 22.

no-sequence and 3 in treatment sequence), and c_t the ten choices made by subject t, i.e. we assume that a subject's type is the same for all ten periods.

The estimation results for the level-k model are shown in the second panel of table 3.2, and the third panel provides results for the closely related poisson cognitive hierarchy model (Camerer et al. 2004).[21] Vuong tests for overlapping models confirm that logit QRE fits all datasets (sequence, no-sequence and the pooled data) significantly better than either level-k or cognitive hierarchy while the difference between level-k and cognitive hierarchy is not significant. [22]

The worse fit could have been expected from Figure 3.10, which shows that observed choice frequencies for the 3 situations of interest (represented by the circles) cannot be matched. More importantly, because of the best-response assumption underlying level-k, it cannot reproduce the intuitive data patterns of Figures 3.5–3.7. For example, in level-k, types 1 and 2 follow unanimous predecessors and they do so irrespective of whether the number of predecessors is 1 or 12. Clearly, this prediction is refuted by the data, see Figure 3.5 (and a similar argument applies to Figures 3.6 and 3.7).

The data patterns of Figures 3.5–3.7 are consistent with models in which choice probabilities vary continuously with expected payoff differences, i.e. when best responses are replaced by (logit) better responses. Introducing trembles has the addi-

[21] Cognitive hierarchy differs from level-k only in that type k is aware of the entire distribution of types lower than k, while agents in the level-k model assume all other agents are of type $k-1$. Choice probabilities in cognitive hierarchy do not vary continuously with τ since a change in τ not only alters the type distribution but it can also shift the best response (since expected payoff calculations depend on the type-distribution). The likelihood function for the cognitive hierarchy model is not differentiable, and we introduce very small logit trembles (using $\lambda = 20$) to facilitate estimation. As for the level-k model, we assume type 0 picks randomly when uninformed and chooses the correct option when informed.

[22] Logit-QRE, level-k and cognitive hierarchy all predict completely random behavior for some parameter values. Therefore, these models are overlapping in the sense of Vuong (1989). To run these tests, we assume each subject produces 1 independent observation. This assumption is consistent with the requirement that all subjects are of the same type in each one of the ten periods as well as with computing standard errors clustered by subject. Note also that level-k still fits worse than logit-QRE when we estimate the type distribution nonparametrically. In treatment sequence, the estimated fractions of types $(0, 1, 2, 3)$ are $(0.68, 0.17, 0.09, 0.05)$ and the associated loglikelihood is -512. In treatment no-sequence, the estimated fractions of types $(0, 1, 2)$ are $(0.70, 0.26, 0.04)$ and the associated loglikelihood is -504. In other words, allowing for an arbitrary type distribution hardly improves the fit compared to the assumed Poisson distribution of types.

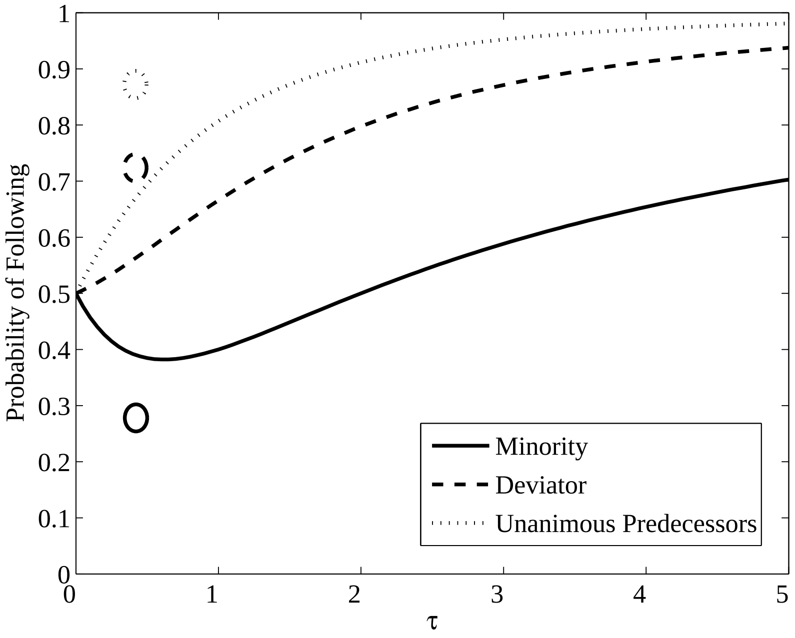


Figure 3.10. Uninformed Subjects' Behavior as a Function of τ in a Level-k Model for Situations that Occurred in the Experiment.

tional benefit that the type distribution can be determined more robustly. Compared to previous experiments, we find a high fraction of type 0: Using our estimate for τ, we obtain an expected fraction of type 0 of 70% in treatment no-sequence and 63% in treatment sequence. [23] The reason is that a subject is classified as type 0 even if her behavior is incompatible with type 1 or 2 in only 1 single period. The model presented in the next section avoids this problem.

3.5.4 Noisy Introspection

The noisy introspection model proposed by Goeree and Holt (2004) combines subject heterogeneity and error-prone behavior by assuming that (i) subjects differ in levels

[23] Instead of estimating the distribution of types, one can also simply classify the observed sample. A subject is then assigned the type that produces the highest likelihood of observing that subjects' decisions. The aim of such a classification is not to make claims about the distribution of types in the general population but simply describes our sample. Using the types as defined by the level-k model, we classify 43 out of 99 subjects as type 0, 17 as type 1, 12 as type 2 and 11 as type 3 while 16 subjects cannot be conclusively classified in treatment sequence. In treatment no-sequence, 65 out of 98 subjects are classified as type 0, 26 as type 1 and 7 as type 2.

of strategic thinking and (ii) subjects realize others are "better responders" and not necessarily "best responders." Let $\phi_\lambda(\cdot)$ denote the logit response function, see (3.4), then type k's choice probabilities can be recursively defined as $P_k = \phi_\lambda \circ P_{k-1}$, i.e. type k better responds believing that observed choices are generated by better responders of type $k-1$. Equivalently:[24]

$$P_k = \underbrace{\phi_\lambda \circ \cdots \circ \phi_\lambda}_{k \text{ times}} \circ \phi_0. \tag{3.6}$$

So type k better responds to type $k-1$, who better responds to type $k-2$, ..., and type 1 better responds to type 0 who chooses randomly (since ϕ_0 results in uniform choice probabilities). The noisy introspection model in (3.6) reduces to the level-k model of the previous section when $\lambda = \infty$, i.e. when logit better responses are replaced by standard best responses.

For finite levels of λ, predicted behavior is different for all types (unlike in the model without trembles). To keep the model parsimonious and comparable to level-k, we again assume that the types follow a Poisson distribution. The estimated parameter values for τ and λ are shown in fifth panel of table 3.2. They are highly significant, and are higher than for the corresponding estimates in models that contain either λ or τ. Hence, compared to the level-k model, the introduction of trembles leads to a right-shift of the type distribution because subjects whose choices are almost always compatible with a type higher than 0 are now classified as such. Likewise, introducing types into logit-QRE increases the estimated rationality parameter because some of the randomness in observed choices is accounted for by the presence of type 0. Note that noisy introspection provides a significant improvement in fit relative to logit-QRE and a dramatic improvement relative to level-k and cognitive hierarchy. [25]

---

[24]Goeree and Holt (2004) define choice probabilities for the noisy introspection model by considering an infinite sequence of logit responses: $P = \lim_{n\to\infty} \phi_{\lambda_n} \circ \cdots \circ \phi_1 \circ \phi_0$, where $\lambda_1 \leq \cdots \leq \lambda_n$. Note that this is equivalent to (3.6) when $\lambda_n = \cdots = \lambda_{n-k+1} = \lambda$ and $\lambda_{n-k} = \cdots = \lambda_1 = 0$.

[25]Since level-k and noisy introspection are nested models, we can run a likelihood ratio test to determine whether the difference in goodness of fit is significant. That is indeed the case for all datasets (sequence, no-sequence and the pooled data). To compare noisy introspection to logit-QRE and cognitive hierarchy, we run Vuong tests for overlapping models, treating each subject as an independent observation. Noisy introspection fits significantly better on all datasets. Just like for

Table 3.2. Overview of Different Models, Robust Standard Errors are in Parentheses.

	τ	λ	$LogL_{obs}$	$LogL_{random}$	$LogL_{best}$	% Explained
QRE						
Sequence		1.3 (0.1)	-417	-686	-364	83.5%
No-sequence		1.4 (0.1)	-438	-679	-407	88.6%
Pooled		1.3 (0.1)	-856	-1366	-771	85.7%
Level-k						
Sequence	0.5 (0.1)		-516	-686	-364	52.8%
No-sequence	0.4 (0.1)		-504	-679	-407	64.3%
Pooled	0.4 (0.1)		-1021	-1366	-771	58.0%
Cognitive Hierarchy						
Sequence	1.2 (0.1)		-489	-686	-364	61.2%
No-sequence	0.5 (0.1)		-504	-679	-407	64.3%
Pooled	1.0 (0.1)		-1003	-1366	-771	61.0%
Noisy Level-k						
Sequence	1.9 (0.3)	2.0 (0.2)	-402	-686	-364	88.2%
No-sequence	1.0 (0.1)	1.6 (0.2)	-457	-679	-407	81.6%
Pooled	1.3 (0.1)	1.7 (0.1)	-869	-1366	-771	83.5%
Noisy Introspection						
Sequence	2.3 (0.3)	1.8 (0.2)	-385	-686	-364	93.5%
No-sequence	2.1 (0.4)	1.7 (0.1)	-429	-679	-407	91.9%
Pooled	2.2 (0.3)	1.8 (0.1)	-815	-1366	-771	92.6%
Noisy Cognitive Hierarchy						
Sequence	2.2 (0.2)	2.0 (0.2)	-383	-686	-364	94.1%
No-sequence	1.6 (0.2)	1.9 (0.2)	-430	-679	-407	91.5%
Pooled	1.9 (0.2)	2.0 (0.1)	-815	-1366	-771	92.6%
Cognitive Hierarchy (Noisy Introspection Trembles)						
Sequence	2.5 (0.3)	2.3 (0.3)	-382	-686	-364	94.4%
No-sequence	2.5 (0.4)	1.9 (0.3)	-430	-679	-407	91.5%
Pooled	2.5 (0.2)	2.1 (0.2)	-813	-1366	-771	92.9%

We also estimate a cognitive hierarchy model in which agents "better respond" to their assumptions about the distribution of types. Like in the noisy introspection model, subjects are aware of the fact that others tremble. Not surprisingly, this combination of cognitive hierarchy and logit-QRE also yields a very good fit (see the last panel in table 3.2). A Vuong test for overlapping models does not allow us to reject the null that the cognitive hierarchy model with noisy introspection trembles fits equally well as the noisy introspection model on any of the datasets.

Previous papers that have allowed for trembles within the level-k framework typically have assumed that subjects are unaware that others tremble (i.e. only the econometrician assumes trembles). In other words, subjects compute their expected payoffs under the assumption that others are best responders (even though they are better responders themselves). To gauge the importance of the "common-knowledge-of-trembles" assumption, we also estimate a noisy level-k model that follows by replacing others' logit responses by best responses:

$$\tilde{P}_k = \phi_\lambda \circ \underbrace{\phi_\infty \circ \cdots \circ \phi_\infty}_{k-1 \text{ times}} \circ \phi_0 \tag{3.7}$$

Note that in this model, there can be situations in which beliefs are inconsistent with observed play. Also, there are now 4 (instead of 3) distinct types in treatment sequence[26] (the number of distinct types does not change for treatment no-sequence). The estimation results in panel 4 of table 3.2 demonstrate that this model fits much better than level-k and cognitive hierarchy. Likelihood ratio tests (for level-k) respec-

the standard level-k model, we can also assign a type to each subject in the framework of a noisy introspection model. Instead of estimating the type distribution, we simply classify each subject as the type that yields the highest likelihood and only estimate the rationality parameter λ. When doing so, 11 out of 99 subjects are classified as type 0, 41 as type 1, 37 as type 2, 8 as type 3 and 1 as type 4 for treatment sequence. In treatment no-sequence, 18 out 98 subjects are classified as type 0, 42 as type 1, 17 as type 2, 7 as type 3, 9 as type 4 and 5 as type 5.

[26]Like type 2, type 3 follows the minority. While type 2 assumes that everybody else follows the majority, type 3 assumes all other agents follow the minority. Therefore, type 2 and type 3 arrive at different expected probabilities that option A is correct and as a consequence also exhibit different trembles. Like type 3, type 4 assumes that everybody follows the minority. Since agents in the simple noisy level-k model are unaware of the fact that other agents tremble, type 4 is identical to type 3.

tively Vuong tests (for cognitive hierarchy) yield significant differences on all datasets. However, Vuong tests for overlapping models allow us to safely reject the null of equal fit in favor of superior fit by noisy introspection as well as by the cognitive hierarchy model with noisy introspection trembles. [27]

Similarly, we also estimate a noisy version of the cognitive hierarchy model and summarize the results in panel 6 of table 3.2. While logistic trembles under the "common-knowledge-of-trembles" assumption generate a substantially better fit than mere logistic responses in the framework of a level-k model, both types of trembles yield a comparable and excellent fit when combined with cognitive hierarchy (in fact, a Vuong test for overlapping models does not allow us to reject the null of equal fit by the cognitive hierarchy models with different types of trembles). Since agents in the cognitive hierarchy model are aware of the entire distribution of lower types, they understand that others can make mistakes (they can be type 0), even in a model without trembles. As a result, the assumption that agents are aware that others tremble seems to be less crucial in a cognitive hierarchy framework compared to the level-k model.

3.6 Conclusion

The "wisdom of the crowds" typically refers to the observation that a group may produce better decisions than any of its members could have. The common explanation is that groups aggregate diverse opinions and preferences, yielding more accurate information or more widely acceptable policies and rules (e.g., Surowiecki, 2004). The wisdom of the crowds is a central outcome of most social-learning models in which imperfectly informed agents infer valuable information from predecessors' choices. In the canonical social learning model developed by Bikhchandani et al. (1992), for instance, herding occurs frequently and almost immediately after a few decisions have

[27]Classifying subjects in the framework of the noisy level-k model for treatment sequence results in 19 out of 99 subjects being assigned type 0, 38 type 1, 12 type 2 and 16 type 3 while the type of 14 subjects remains unidentified. For treatment no-sequence, 24 out of 98 subjects are classified as type 0, 59 as type 1, 4 as type 2 and 11 as type 3.

been observed.

In a clever and novel contribution, Callander and Hörner (2009) discuss settings where the *minority* is predicted to be correct. Necessary conditions for this prediction to hold are that agents are differentially informed and only observe the number of times each option is chosen (and not the exact sequence of prior choices as in Bikhchandani et al., 1992). In this paper, we test the Callander–Hörner model in a controlled laboratory setting. We employ a simplified version of their model where each agent is either perfectly informed (with small probability) or not informed at all (with complementary high probability). We report data from 2 treatments: in treatment "sequence," subjects could observe the entire sequence of predecessors' choices, while in treatment "no-sequence" they could see only how many times either option had been chosen. For our setup, the predictions are that subjects follow their immediate predecessors in treatment sequence and follow the minority in treatment no-sequence.

In a nutshell, the data may be characterized as follows: subjects tend to overwhelmingly follow unanimous predecessors in both treatments (87%), they more frequently than not follow a deviator in treatment sequence (72%), and they *do not* follow the minority in treatment sequence (28%). The observed deviations from theoretical predictions are approximate best responses to the empirical distribution of play. Given the choices of others, and given the draws used in the experiment, not following unanimous predecessors is very costly ($1.53 on average), not following a deviator in treatment sequence is less costly ($1.23 on average), and not following the minority *pays* ($0.36 on average).

We analyze the deviations observed in our data using alternative models of bounded rationality. In the logit Quantal Response Equilibrium (QRE), for instance, agents are predicted to tremble, which can overturn the logic for why the minority is correct. Intuitively, if the probability of being informed is very low then a "deviant minority" is more likely the result of a tremble that should rationally be ignored. When applying the logit-QRE to our data we find that it is able to reproduce the main (aggregate) features. Zooming in on individual level data, however, reveals a substantial amount of heterogeneity that cannot be explained by the symmetric logit-QRE.

Heterogeneity in levels of strategic thinking is more naturally explained by the cognitive hierarchy model (Camerer et al. 2004) or the closely related level-k model (e.g., Crawford and Iriberri 2007b). The level-k model, for example, assumes that agents of type k best respond to their beliefs that others are of type $k-1$. We apply both models to our data and find they are also able to reproduce the main aggregate features, although not as well as logit-QRE. Furthermore, their fit of the individual data is substantially worse. The main reason for these shortcomings is the underlying best-response assumption, which conflicts with several intuitive comparative statics observed in our data (e.g., subjects tend to follow unanimous predecessors more frequently when the group of predecessors is large). In addition, the best-response assumption skews the estimated type distribution towards lower types.

The noisy introspection model proposed by Goeree and Holt (2004) combines heterogeneity in strategic thinking with noisy behavior by replacing level-k's best response with a logit "better response." Importantly, agents in the noisy introspection model know that others tremble. For example, when computing the probability that the minority choice is correct in treatment no-sequence, agents take into account the possibility that the minority arose because of trembles. As a result, the model can predict why subsequent choices favor the majority even for agents with high levels of strategic thinking. We illustrate the importance of the "common-knowledge-of-trembling" assumption by also estimating a noisy version of the level-k model in which agents tremble but assume others do not (as is typically done in the literature) and show that it fits significantly worse.

A combination of the cognitive hierarchy model with either simple logistic trembles or introspective trembles also yields a very good fit. Like the noisy introspection model, these version of cognitive hierarchy allow for heterogeneity in subject behavior as well as trembles. Realizing that others can be error prone, sophisticated agents in these models benefit from following the wisdom of the crowds.

Chapter 4

An Experimental Test of Flexible Spectrum Auction Formats

This chapter reproduces a paper written jointly with Jacob K. Goeree, Charles A. Holt, and John O. Ledyard.

4.1 Introduction

Simultaneous auctions for multiple items are often used when the values of the items are interrelated. An example of such a situation is the sale of spectrum rights by the Federal Communications Commission (FCC). If a telecommunications company is already operating in a certain area, the cost of operating in adjacent areas tends to be lower. In addition, consumers may value larger networks that reduce the cost and inconvenience of "roaming." As a consequence, the value of a collection of spectrum licenses for adjacent areas can be higher than the sum of the values for separate licenses.[1] Value complementarities arise naturally in many other contexts, e.g., aircraft takeoff and landing slots, pollution emissions allowances for consecutive years, and coordinated advertising time slots. This paper reports a series of laboratory experiments to evaluate alternative methods of running multi-unit auctions, in both high- and low-complementarities environments.

[1]There can also be important synergies in the spectrum frequency dimension, where adjacent bands may improve capacity and reduce interference. For instance, in the FCC auction for air-to-ground communications frequencies in May 2006, a package of 3 bandwidth units sold for about 4.5 times as much as a single unit, and similar synergies were implied by unsuccessful bids.

Various auction formats have been suggested for selling multiple items with interrelated values. The most widely discussed format is the simultaneous multiple round (SMR) auction, first used by the FCC in 1994. In the SMR auction, bidders are only allowed to bid on single licenses in a series of "rounds," and the auction stops when no new bids are submitted on any license. To win a valuable package of licenses in this type of auction, bidders with value complementarities may have to bid more for some licenses than they are worth individually, which may result in losses when only a subset is won. Avoidance of this "exposure problem" may lead to conservative bidding, lower revenue, and inefficient allocations.[2]

The obvious solution to the exposure problem is to allow bidding for packages of items. In such combinatorial auctions, bidders can make sure they either win the entire package or nothing at all. As a result, bids can reflect value complementarities, which should raise efficiency and seller revenue. Combinatorial bidding, however, may introduce new problems. Consider a situation in which a large bidder submits a package bid for several licenses. If other bidders are interested in buying different subsets of licenses contained in the package, they might find it hard to coordinate their actions, even if the sum of their values is higher than the value of the package to the large bidder (the threshold problem). Thus, there is no clear presumption that package bidding will improve auction performance. The FCC has increasingly relied on laboratory experiments to evaluate the performance of alternative spectrum auctions (see also Goeree and Holt 2008). The next section summarizes the main features of the auction formats to be considered.

4.2 Alternative Auction Formats

The various combinatorial auctions to be considered are best understood in terms of how they differ from the incumbent standard, the FCC's simultaneous multi-round

[2]In the recent AWS auction (FCC auction 66), for example, the total cost of acquiring 20 MHz of nationwide coverage was \$2.268 billion for all 734 individual licenses in the "A-block" while the total cost was \$4.174 billion for the 12 larger regions in the "F-block." Presumably there was a larger exposure problem in the A-block because it consisted of a larger number of small licenses.

auction procedure. Therefore, we will begin by explaining how the SMR auction was implemented in the experiments. Each auction consists of multiple rounds in which bidders have a fixed amount of time to submit their bids. Once the round ends, the highest bid on each license is announced as a provisional winner. Once no more bids are submitted, the auction stops and the provisionally winning bids become the final winning bids.

There are 2 constraints on bidding. The first constraint is the FCC "activity limit" that determines the maximum number of different licenses for which a bidder can submit bids. Each bidder is assigned a prespecified activity limit at the beginning of the auction. A bidder's activity limit falls if the number of submitted bids (plus the number of provisionally winning bids in the previous round) is less than the bidder's activity limit in the previous round. Activity is transferable, so a bidder with a limit of 3 could bid on licenses A, B, and C in 1 round and on licenses E, F, and G in the next round, for example. The second restriction is that each bid must exceed the previous high bid for that license by a specified bid increment. This requirement is a minimum, and new bids can exceed the "provisionally winning" bid by up to eight bid increments. The only exception to the increment rule is that the provisionally winning bidder is not required to raise that bid. Bidders can observe others' previous bids and can see which of those were provisionally winning.[3] The effect of activity limits and bid increments is to force bids upward, although there are limited opportunities for withdrawing bids.[4] The auction stops after a round in which no new bids are submitted and no withdrawals occur.

This multi-round procedure can be adapted to allow for bids on both individual licenses and packages, and this approach has been shown to improve auction

[3]In most FCC auctions to date, bidders' identities are revealed during the auction. More recently, the FCC has contemplated revealing bid amounts but not bidder identities (anonymous or "blind" bidding).

[4]As a partial remedy to the exposure problem, the FCC allows bidders to withdraw their provisionally winning bids in at most 2 rounds, at a penalty that equals the difference between their withdrawn bids and the subsequent sale price if that is lower. Porter (1999) reports laboratory data showing that the introduction of this withdrawal rule increases the efficiency of the final allocation as well as the seller's revenue.

performance in some cases.[5] With package bidding, the relevant price of a license is not necessarily the highest bid on that license; indeed there may not even be a (non-package) bid on a particular license. One approach is to calculate the revenue-maximizing allocation of licenses after each round, and to use "shadow" prices that represent marginal valuations in terms of maximized revenue. Then the price of a package is the sum of the prices for individual items, and new bids in the subsequent round then have to improve on these prices by some minimum increment that depends on the size of the package. As with SMR, bidders are given the option of selecting 1 of a series of pre-specified higher increments. This approach, known as RAD (Resource Allocation Design) pricing, is due to Kwasnica et al. (2005). One advantage of the RAD approach is that prices may convey information about how high a bidder must go to "get into the action" on a particular license or package.[6] The revenue maximization at the close of each round uses all bids for all completed rounds. This maximization routine results in provisionally winning bids (on licenses or packages) and associated RAD prices. As in the SMR auction, a specified bid increment is added to the RAD price to determine the minimum acceptable bid for the license in the next round.[7] The minimum acceptable bid for a package is simply the sum of minimum acceptable bids for the licenses it contains. Bidders were allowed to submit multiple bids on licenses and/or packages. The treatment of activity limits is analogous to SMR, with activity being calculated as the number of different licenses being bid for or being provisionally won in the previous round (separately or as part of a package). The auction stops when no new bids are submitted, and the "provisional winning bids" for that round become the final winning bids (withdrawals are not needed with this format).

The FCC developed a variant of RAD pricing, called SMRPB. Of the 4 formats

[5]Rassenti et al. (1982) first used experiments to compare the performance of sealed-bid auctions with and without package bidding. Ledyard et al. (1997) provide data comparing several iterative processes. The combinatorial auction produces higher efficiencies in both designs.

[6]Ideally, the license prices should represent the revenue value of relaxing the constraint that there is only 1 of each license. The discreteness in license definitions may, however, result in nonexistence of dual prices, and Kwasnica et al. (2005) propose a method of computing approximate prices.

[7]In order to prevent cycles, the bid increment is raised after a round in which revenue does not increase.

considered, SMRPB is the only auction procedure that employs an "XOR" bidding rule, which means that each bidder can have at most 1 winning bid. For example, the XOR rule means that a bidder who is interested in both licenses A and/or B must bid on A, B, and the package AB, since a bid on AB alone would preclude winning either license separately while bids on A and B only would preclude winning the package. Since XOR bidding typically calls for making bids on lots of combinations, the activity rule used with the FCC's SMRPB auction is based on the size of the largest package bid, so a bidder with activity 3 could bid on both ABD and ABC, for example, but not on ABCD. Another difference with RAD concerns the pricing rule: in the SMRPB version, prices adjust slower in response to excess demand because they are "anchored" with respect to prices in the previous round.[8]

An alternative approach to the pricing problem is to have prices rise automatically and incrementally in response to excess demand, via a "clock" mechanism (Porter et al., 2003). In each round of the combinatorial clock (CC) auction, the price of a combination is the sum of the prices for each component, and bidders can indicate demands for 1 or more individual items or combinations of items. If more than 1 bidder is bidding for an item in the current round, either separately or as part of a package, the clock price for that item rises by the bid increment. Otherwise, the price remains the same. There are no provisional winners, but other aspects of this auction are analogous, e.g., activity is defined in terms of the number of different licenses for which a bidder indicates a demand.[9] The auction typically stops when there is no longer any excess demand for any item.[10] One possible advantage of an

[8]See Appendix D in the Goeree and Holt (2005) experiment design report for more details.

[9]Note that Porter et al. (2003) did not use activity limits in their combinatorial clock auctions.

[10]When there is no more excess demand for any of the licenses but some are in excess supply, the revenue maximizing allocation is calculated using all bids in the current and previous rounds. If this process results in a failure to sell to the remaining bidder for an item, the clock is restarted to let that bidder have another chance to obtain the item. This restart procedure can be illustrated with a simple 3-license example, which is taken from the instructions to subjects. Suppose bidder 1 only wants license A and is willing to bid up to 40 for A, bidder 2 only wants license B and is willing to bid up to 40 for B, and bidder 3 only wants license C and is willing to bid up to 80 for C. Finally, bidder 4 only wants package ABC and is willing to bid up to 150 for ABC. Initially there is excess demand for all licenses, which causes prices to rise. Bidders 1 and 2 drop out when prices rise to 45, 45, 45, but since there is still competition for license C its price continues to rise. Bidder 4 is willing to keep bidding on ABC as long as the price of C does not exceed 60. So when the price of C rises to 65, bidder 4 drops out. At prices of 45, 45, 65 no one is bidding for licenses A and B. At

incremental clock auction is that it prevents aggressive "jump bids," which have been observed by McCabe et al. (1988) in the laboratory and by McAfee and McMillan (1996) in an FCC auction.[11] The clock-driven price increments may also alleviate the threshold problem of coordinating small bidders' responses to large package bids. In addition it is possible to add a final round of sealed-bids to the clock phase. This final or shootout phase could be structured as a first-price (pay-as-bid) auction or a second-price auction (proxy bidding, see Ausubel et al. 2006).

Results of laboratory experiments suggest that these and other forms of package bidding may enhance performance measures, especially in environments with high complementarities.[12] In the Porter et al. (2003) experiment, for example, the combinatorial clock auction attained 100% efficiency in 23 sessions and 99% efficiency in 2 other sessions. Previous experiments have mainly focused on specific auction formats. This chapter provides a systematic and parallel consideration of SMR and its most widely discussed alternatives, including the one developed by the FCC.[13]

this point, bidder 3 is the only one bidding on C but the computer finds it better to assign ABC to bidder 4 (for a total of 45+45+60 = 150) than to assign A to bidder 1, B to bidder 2, and C to bidder 3 (for a total of 40+40+65 = 145). To allow bidder 3 (who has a value of 80 for C) to get back into the action on license C, the computer will raise the price of C further to 70, 75, until (i) either bidder 3 drops out or (ii) the revenue from assigning A to bidder 1, B to bidder 2, and C to bidder 3 exceeds that of assigning ABC to bidder 4. In this manner the price of a license can rise even though only 1 bidder is still bidding for it. Also, bidders may be assigned a license or package even though they were no longer bidding in the final round (as for bidders 1 and 2 in the example).

[11] In the recent AWS auction, for example, 1 of the bidders made the maximum allowed jump bid for the hotly contested Northeast and West regional licenses, effectively doubling the prices to about $1.5 billion. The main competitors for these licenses ceased bidding immediately afterwards.

[12] Banks et al. (1989) proposed a different type of combinatorial auction, called Adaptive User Selection Mechanism (AUSM). In this auction, bidders can submit bids for individual licenses and packages in continuous time. A new bid becomes provisionally winning if revenue can be increased by an allocation that includes the new bid. Kwasnica et al. (2005) compare RAD and AUSM to SMR in a laboratory setting. Efficiencies observed with RAD and AUSM are similar and higher than those for SMR, but revenue is higher in SMR since many bidders lose money due to the exposure problem. (If we assume that bidders default on bids on which they make losses and thus set the prices of such bids to zero, revenues are in fact higher under AUSM and RAD than under SMR.) Charles River and Associates also developed a combinatorial auction, called Combinatorial Multi-Round Auction (CMA). In this auction, only bids that are sufficiently high allow bidders to maintain their activity. A bid is sufficiently high when it is at least 5% higher than the currently highest combination of bids that spans the same licenses. Banks et al. (2003) ran an experiment to compare the CMA and SMR auction formats. They find that the CMA leads to more efficient allocations but less revenue since many bidders incur losses in their SMR auction experiments due to the exposure problem. Porter et al. (2003) also compare CMA to SMR auctions and also find that CMA tends to lead to more efficient allocations.

[13] A common feature of the combinatorial formats discussed in this paper is that they permit

4.3 Experimental Design

Our design involves groups of eight bidders and 12 licenses, a size that was selected to provide enough added complexity, while still permitting us to obtain sufficient independent observations for a broad range of auction formats and value structures. Bidders' values for the licenses were randomly determined for each auction, which resulted in a rich variety of market structures. There are 2 types of bidders in this design: small regional bidders (labeled 1 through 6) and large "national" bidders (labeled 7 and 8). A graphical representation of bidders' interests is shown in Figure 4.1. Each diamond represents a different region, and the licenses along the center line (A, D, E, H, I, and L) are the ones of interest to the 2 national bidders. In the diamond shaped region on the far left, for example, the regional bidders, 1, 2, 5 and 6, are interested in licenses B and C, and in addition, each is interested in 1 of the licenses (A or D) that are targets for the 2 national bidders. Similarly, in the middle region, small bidders 1, 2, 3, and 4, are interested in licenses F and G, and each one is interested in 1 of the licenses (E and H) that are also of interest to the national bidders. The far-right diamond shaped region has a similar structure. Notice that each regional bidder has interests in 2 adjacent regions, e.g., the left and center diamonds for bidders 1 and 2. Subjects' ID numbers stayed the same throughout the experiment, and, hence, so did their roles as regional or national bidders.

Regional bidders can acquire at most 3 licenses, and complementarities occur only when licenses in the same region are acquired. For example, if bidder 1 wins the combination ABE, then the value synergies would only apply to A and B, which are in the same region in Figure 4.1. Since value synergies do not apply across regions, a group of licenses in 1 region is a substitute for a group of licenses from another region, which creates an interesting "fitting problem." For example, under the SMR

"flexible" package bidding, i.e., bidders can construct arbitrary "customized" packages. An alternative approach is to restrict bidding to pre-specified packages as was done in the FCC air-to-ground auction in 2006. Rothkopf et al. (1998) have suggested hierarchically structured sets of pre-defined packages to reduce the complexity of the (revenue-based) assignment problem. Goeree and Holt (2008) propose a simple pricing mechanism for hierarchically structured packages and test the resulting auction in the lab.

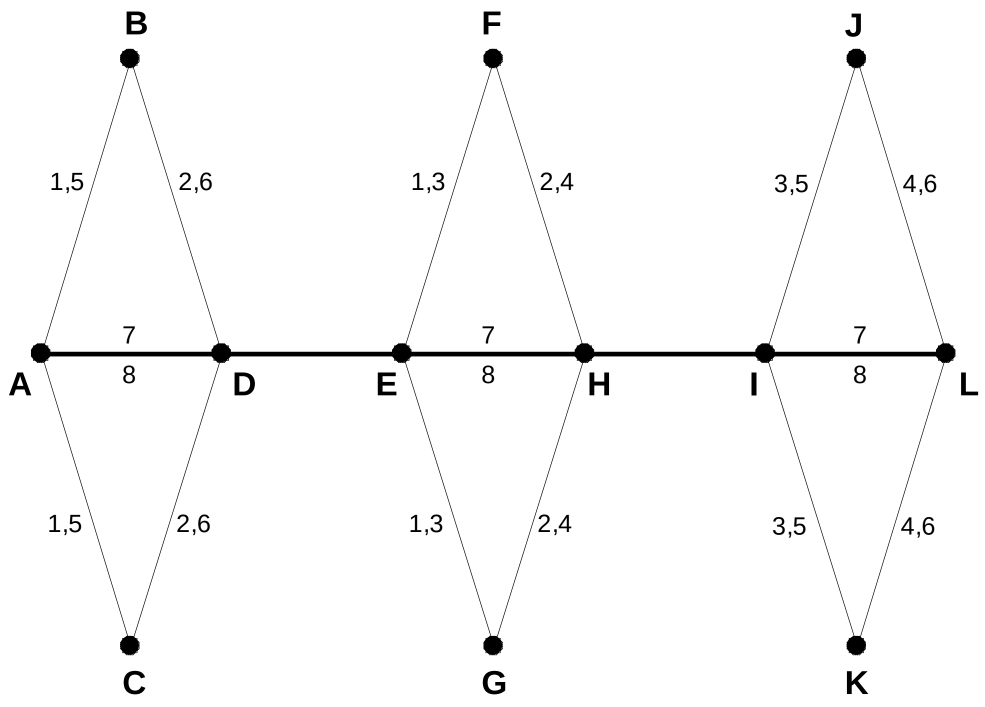


Figure 4.1. Eight-Bidder Design with 3 Regions. Regional bidders (1-6) are interested in 1 side of 1 of 2 diamond-shaped regions. National bidders (7-8) are interested in the middle line connecting all 3 diamond-shaped regions.

procedure, bidder 1 with an activity of 3 could either bid on licenses A, B, and C in the left region or E, F, and G in the middle region to capture the regional synergies. Likewise, under the RAD and CC procedures, bidder 1 could either bid to obtain synergies for the ABC package or the EFG package. The "XOR" rule used in SMRPB facilitates the regional bidders' "choice of region" problem because it allows them to bid on packages from *both* regions knowing that at most 1 bid can be winning. An additional advantage of the "XOR" rule is that bidders always know the maximum financial liability they face, i.e., the highest dollar amount of any of their bids.

National bidders can acquire up to 6 licenses and they have value complementarities for all 6 licenses in some treatments and for only 4 licenses in other treatments. The larger number of licenses subject to complementarities creates a larger exposure problem for the national bidders. The total number of possible allocations with this setup is 13,080,488.

Auction formats. The 4 auction formats are described in detail in Appendices A-C. They include 3 combinatorial formats (SMRPB, RAD, and CC) and 1 non-combinatorial format (SMR). The main modification of the basic SMR procedures described above is that bid withdrawals were permitted in at most 2 rounds of an auction. For example, a bidder who withdraws any number of bids in rounds 8 and 10 would not be able to make any withdrawals in subsequent rounds. If a withdrawn bid caused the final sale price to go down, the bidder had to pay the difference. If a license with a withdrawn bid went unsold, however, then the bidder was only responsible for 25% of the withdrawn bid, which represents a penalty intended to mimic the effect of having to pay the difference between a withdrawn bid and a lower sale price in a subsequent auction. A key feature of the withdrawal provisions is that the seller (FCC) becomes the provisionally winning bidder at the second-highest bid (minus a bid increment), so that the person who originally made the second-highest bid would be able to reenter at that level if the bidder has activity and interest to do so. This provision can benefit a bidder whose interests have changed, perhaps to a different region.

For each auction format, the experiments cover 4 different treatments: high/low overlap in national bidders' interests (HO vs. LO) and high/low complementarities (HC vs. LC). For example, treatment HOHC has high overlap and high complementarities. We next describe the treatment variations in more detail.

Complementarities. Payoffs in the experiment were expressed in terms of points, where each point was worth \$0.40 to subjects. (The bid increment was 5 points in all auctions.) The baseline draw distributions are uniform on the range [5, 45] for each license of interest to national bidders, and on the range [5, 75] for each license of interest to regional bidders. Synergies between licenses are modeled in a linear manner: when a bidder acquires K licenses the value of each goes up by a factor $1 + \alpha(K - 1)$. In the high-complementarities treatment, the synergy factor (α) for national bidders was 0.2. Thus each license acquired by a national bidder goes up in value by 20% (with 2 licenses), by 40% (with 3 licenses), by 60% (with 4 licenses), by 80% (with 5 licenses) and by 100% (with all 6 licenses). With low complementarities, these numbers are 1%, 2%, 3%, 4% and 5%, corresponding to $\alpha = 0.01$. With high complementarities (HC), each license acquired in the same region by a regional bidder goes up in value by 12.5% (with 2 licenses in the same region), and by 25% (with 3 licenses in the same region), so $\alpha = 0.125$. With low complementarities, these numbers are 1% and 2% for regional bidders. These minimal complementarities in the LC treatment allowed us to maintain parallelism in instructions and procedures. Participants were informed about the synergies that applied to regional and national bidders and about the distributions of possible values (but not about others' value draws).

Overlap. With high overlap (HO), each national bidder, 7 and 8, has value draws from the same distribution for all 6 licenses on the base of Figure 4.1, and the complementarities apply equally to all 6 licenses. In this sense, each national bidder is equally strong across the line. With low overlap (LO), national bidder 7 only receives complementarities for the 4 licenses on the left side of the base (A, D, E, and H). Conversely, national bidder 8 receives complementarities for the 4 licenses on the right side (E, H, I, and L). Thus with high complementarities and low overlap, each national bidder has a natural focus of interest that only partially overlaps with the other national bidder's area. One issue of interest is whether this type of partial separation may yield tacit collusion and less aggressive bidding in the center.

Treatment structure. The 2-by-2 treatment design yields 4 treatments for each of the 4 auction formats, for a total of 16 treatments. We used the same value draws across auction formats so that differences cannot be attributed to specific sequences of value draws. Each session consisted of 1 or 2 practice auctions and a series of 6 auctions for cash payments. The treatment and auction type was unchanged for all auctions in a session, but the randomly generated value draws changed from one auction to the next. In addition, we used new sequences of random draws for each of 3 "waves" of 16 sessions that spanned all treatments. To summarize, there were 18 (3 waves times 6 auctions) independent sets of value draws that were used in all 4 auction formats.

Subjects and sessions. Before conducting the sessions that form waves 1-3, we trained over 128 Caltech subjects in 16 sessions of eight people. These inexperienced sessions ("wave 0") involved both SMR and combinatorial auctions and were conducted to familiarize subjects with the auction software and bidding environment.[14] For these inexperienced sessions, we promised to pay each person a $60 bonus (in addition to other earnings) if they returned 3 more times.[15] This decision to use experienced bidders was based on the complexity of the auction formats and on earlier pilot experiments. For the subsequent data analysis, only the data from waves 1-3 but not from wave 0 is used. In waves 1-3, earnings averaged $50 per person per session, including $10 show-up fees and bonuses, for sessions that lasted from 1.5 to 2 hours, depending on the number of auctions.[16] In total, there were 16 training sessions and 48 sessions (3 ×16) with experienced subjects, each involving a group of eight subjects.

4.4 Results

One way to measure market efficiency is to divide the sum of all bidder values for licenses they won, the actual surplus (S_{actual}), by the maximum possible surplus ($S_{optimal}$). It is well known that this simple efficiency measure may be difficult to interpret. For example, adding a constant to all value amounts will tend to raise this efficiency ratio, since efficiency losses are affected by differences in valuations, not absolute levels. A more natural measure of efficiency is calculated on the basis of the difference between the actual surplus and the surplus resulting from a random allocation (S_{random}), this being normalized by the maximum such difference.

[14]The experiments were run using *jAuctions*, which has been developed at Caltech by Jacob Goeree. The *jAuctions* software consists of a flexible suite of Java-based auction programs designed to handle a wide range of auction formats and bidding environments, including combinatorial auctions with bid-driven or clock-driven prices, private and common valuations, etc. Instructions, which are available on request, were structured around relevant screen shots of the *jAuctions* program.

[15]As a consequence, most subjects participated in more than 1 auction format.

[16]In some cases, subjects ended the session with negative earnings, and these subjects were only paid the show-up fee. An alternative would have been to rotate bidder roles during the session, which would have avoided negative final earnings. This is the procedure followed in Goeree and Holt (2008).

$$efficiency = \frac{S_{actual} - S_{random}}{S_{optimal} - S_{random}} * 100\%$$

The value of a random allocation can be computed by taking the average of the surplus over all possible allocations, of which there are 13,080,488 in total for the design in Figure 4.1.[17] This definition of efficiency measures how much the auction raises surplus relative to a random allocation mechanism. In the analysis that follows, we will use these normalized efficiency measures.

Similarly, revenues will be measured as the difference between actual auction revenue and the revenue from a random allocation in which bidders pay their full values for all licenses and packages they receive ($R_{random} = S_{random}$). This difference is then divided by the difference between the maximum possible revenue ($R_{optimal} = S_{optimal}$) and the revenue from a random allocation. Note that the optimal revenue benchmark is the revenue obtained if bids equal full value on all licenses and packages leaving zero profits for the bidders, i.e., full rent extraction.

$$revenue = \frac{R_{actual} - R_{random}}{R_{optimal} - R_{random}} * 100\%$$

Since $R_{random} = S_{random}$ and $R_{optimal} = S_{optimal}$, the denominators of the normalized efficiency and revenue measures are equal, and the normalized sum of bidders' profits is simply equal to the difference between efficiency and revenue:

$$profit = \frac{S_{actual} - R_{actual}}{S_{optimal} - R_{random}} * 100\% = \frac{\sum_i \pi^i_{actual}}{S_{optimal} - S_{random}} * 100\%.$$

All efficiency, revenue, and profit measures reported below are normalized in this manner for the specific value sequences used in each auction for each of the 3 waves of sessions with experienced bidders.[18]

---

[17] Without the restriction that regional bidders can acquire at most 3 licenses, the total number of allocations would be 16,777,216.

[18] There were occasional glitches in the data recording, i.e., when a bidder's computer would temporarily be offline. A detailed analysis of all the bid books reveals that less than 1% of all bids were lost. Since a bidder's activity was determined by the bids she submitted (not according to the bids recorded by the server) this had no adverse effects for the bidder's activity. Unless the round in which this occurred was the final round, the effect of lost bids was negligible.

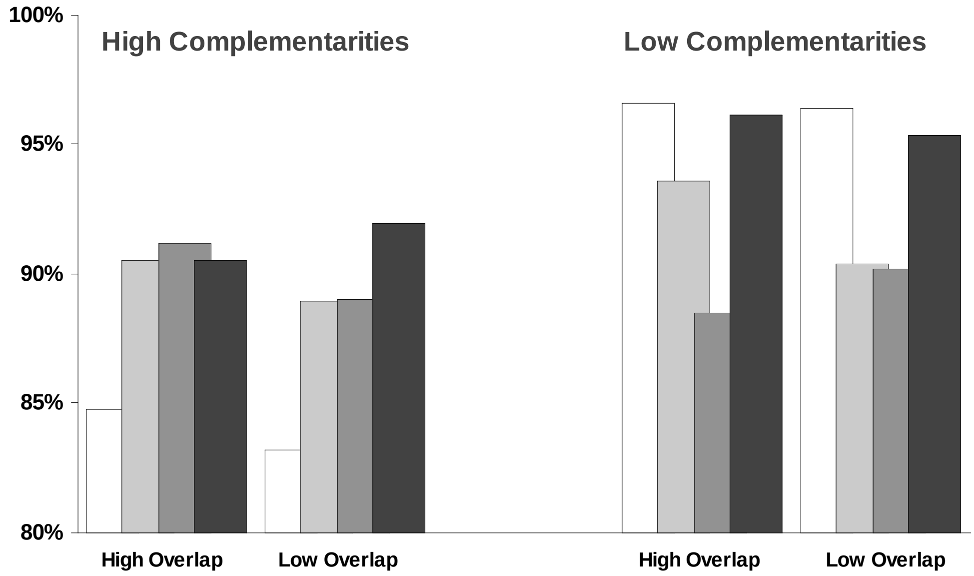


Figure 4.2. Efficiency by Auction Format. The bars from light to dark (left to right) correspond to SMR, CC, SMRPB and RAD respectively.

Efficiency. Package bidding is designed to help bidders avoid the "exposure problem" of bidding high for licenses with high complementarities. As expected, switching from SMR to a combinatorial format raises efficiency in the high complementarities treatments. The differences between SMR and the combinatorial formats occur for both of the high complementarities treatments, as can be seen from the left side of Figure 4.2. In this and subsequent figures the colorcoding is as follows: from light to dark the bars correspond to SMR, CC, SMRPB, and RAD respectively.

In contrast, the switch to combinatorial auctions reduces efficiency when complementarities are minimal (our "low complementarities" treatment). The efficiency levels are now 97% for SMR and 89%, 92%, and 96% for SMRPB, CC, and RAD respectively. Again, this difference shows up in both LC treatments shown on the right side of Figure 4.2. Result 1 summarizes our findings, where we use the following notation: $\sim$ implies a pairwise difference is not significant, $\succ^{*}$ indicates significance at the 10% level, $\succ^{**}$ indicates significance at the 5% level, and $\succ^{***}$ indicates significance at the 1% level.

Result 1 *With high complementarities, efficiency levels are highest for the 3 combi-*

natorial formats and are ranked

$$RAD \sim SMRPB \sim CC \succ^{**} SMR.$$

With low complementarities, efficiency levels are ranked

$$RAD \sim SMR \succ^{**} CC \sim SMRPB.$$

Pooling the low and high complementarities treatments, efficiency levels are ranked

$$RAD \succ^{*} CC \sim SMRPB \sim SMR.$$

Support. Session averages are grouped by wave and auction format in Appendix C.1. For example, consider the efficiencies for the HC treatments (pooling high and low overlap) shown in the eight columns on the left side of Appendix refapD (top 3 rows). It is important to compare the auction formats for the same wave, since the valuation draws change from one wave to another. All 6 of the paired comparisons for the HC treatments show higher efficiencies for any of the package bidding auctions compared to SMR. This effect is significant using a Wilcoxon matched-pairs signed-rank test (p = 0.03). These results are generally reversed with low complementarities, where all paired comparisons between SMR and CC and SMRPB go in the opposite direction (higher efficiency for SMR): this effect is significant (p = 0.03). The only combinatorial auction that is not statistically different from SMR in terms of efficiency is RAD (p = 0.41). When pooling the data from the low and high complementarities treatments, RAD is more efficient than SMRPB (p = 0.02), CC (p = 0.09), and SMR (p = 0.09). There are no significant differences between SMR, SMRPB, and CC.

In addition to being statistically significant, the differences in observed efficiencies are also economically significant. With high complementarities (combining the low and high overlap treatments and data from all 3 waves), the average efficiency in SMR is 84% while it is 90%, 90%, and 91% in SMRPB, CC, and RAD respectively.

One reason why SMR leads to low efficiencies with high complementarities is the incidence of unsold licenses, which happens with rates of 4% and 7% in the high-and low-overlap treatments respectively, see Appendix C.1. Unsold licenses typically result from withdrawals late in the auction when a bidder realizes that it will not be possible to obtain the value synergies associated with multiple licenses. After a withdrawal, recall that the seller becomes the provisional winner at the second highest bid, and the person who made that bid previously may have lost activity or interest in that license, which causes it to go unsold.[19] Withdrawals are not permitted in the combinatorial auctions, where the exposure problem is addressed directly by allowing package bids, so these auctions do not result in unsold licenses. The difference between SMR and any of the combinatorial formats in terms of license sales rates is significant with a Wilcoxon matched-pairs signed-ranks test (p = 0.05).

Revenues. Figure 4.3 shows the revenues by auction format and treatment averaged across sessions (session averages for each parameter/experience wave can be found in rows 4-6 of the table in Appendix C.1). What is obvious from Figure 4.3 is that the combinatorial clock auction extracts more rents for the seller in all treatments, even when it is less efficient than other formats.

Result 2 *Revenues are highest for the combinatorial clock auction and are ranked* $CC \succ^{***} RAD \sim SMRPB \sim SMR$.

Support. There are 3 rows in the Revenue section of Appendix C.1, 1 for each wave of parameter values. In each row, there are 4 paired comparisons between CC and a particular alternative format, so overall there are 12 paired comparisons. The CC provides higher revenue in all 12 pairwise comparisons with each of the alternatives, except for RAD where CC yields higher revenues in 11 of 12 cases. These comparisons are significant using a Wilcoxon matched-pairs signed-rank test (p = 0.001). Basically,

[19]Since the seller's value for a license is assumed to be 0, an unsold license was given a value of 0 in the efficiency calculations. This calculation provides a lower bound for the efficiency since unsold licenses are typically sold in later auctions (there is, however, an efficiency loss associated with delays in spectrum use). Alternatively, a rough estimate of the upper bound for the efficiency would be to scale the actual efficiency by 1 + x where x is the proportion of unsold licenses. By scaling up the efficiencies for the SMR auction in the high-complementarities treatments using an "x factor" of 0.05 (proportion of unsold licenses averaged across treatments) raises average efficiency from 84% to 88%. This scaled up efficiency is only slightly lower than those for the combinatorial formats, indicating that a large part of the efficiency loss in the SMR auction is due to unsold licenses. Note that these calculations ignore "selection effects," i.e., low-value licenses are more likely to go unsold.

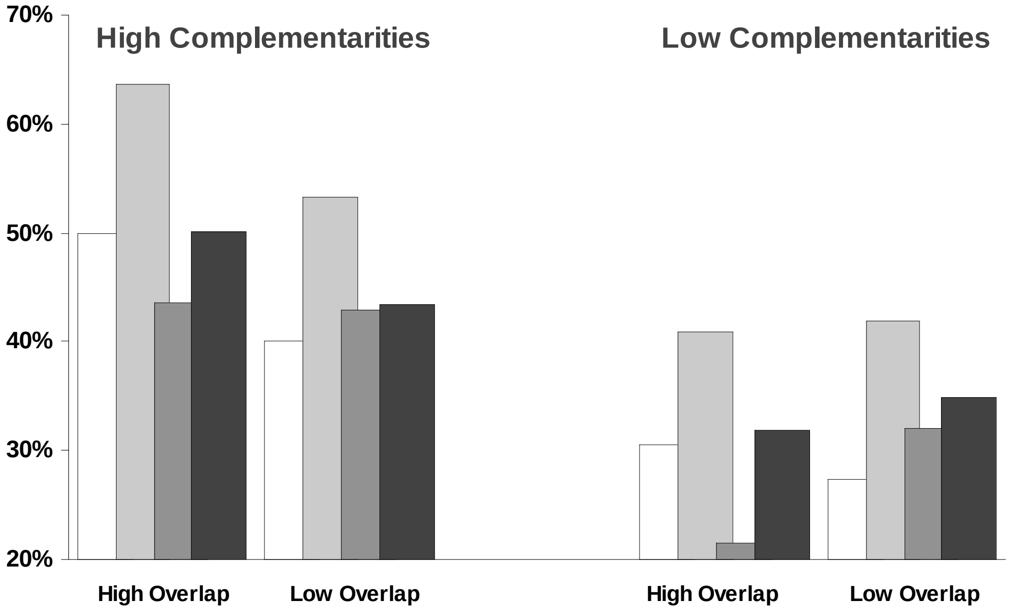


Figure 4.3. Revenue by Auction Format. The bars from light to dark (left to right) correspond to SMR, CC, SMRPB and RAD respectively.

CC is higher than the others with both low and high complementarities (p = 0.03), except for the comparison with RAD in the low-complementarities (LC) treatment (p = 0.06). Revenue under RAD is borderline significantly higher than SMRPB when we pool all data (p = 0.109). Revenue under RAD is not significantly different from SMR, and SMR and SMRPB raise the same revenues.

These revenue differences are also significant economically: averaging over all treatments and all waves, the revenue from the combinatorial clock format is 50% as compared to 37%, 40%, and 35% for the SMR, RAD, and SMRPB auctions respectively.

With high overlap, national bidders own more licenses and, hence, can create bigger packages with higher associated values especially when complementarities are high. The result is that revenues are higher for the High Overlap and High Complementarities bars on the left side of Figure 4.3. Moreover, national bidders earn more in the high overlap and high complementarities treatments, while the regional bidders do worse (see the table in Appendix C.1). These revenue and profits results fit with our prior expectation that there could be more tacit collusion in the low overlap treatments where there is less head-to-head competition between the national bidders.

Profits. Figure 4.4 shows bidders' profits by auction format and by treatment. The ability to bid for combinations allows bidders to bid high on packages and avoid the exposure problem, an effect that is mainly relevant with high complementarities. But if all bidders bid higher, the effect on bidder profits is unclear.

Result 3 *Bidders' profits are lowest in the combinatorial clock auction and are ranked* $RAD \sim SMRPB \sim SMR \succ^{***} CC$.

Support. Normalized profits are calculated as the differences between entries in the efficiency and revenue rows of Appendix C.1. With 3 waves and 4 treatments, there are 12 paired profit comparisons between CC and a particular alternative format, and the CC provides lower profits in all 12 pairwise comparisons with SMR, and for 11 of the 12 comparisons with RAD and SMRPB. These comparisons are significant using a Wilcoxon matched-pairs signed-rank test (p = 0.001). Averaged over treatments, profits for CC are 40% while the profits for the other formats are all in a narrow range from 53% to 55%.

The exposure problem can be alleviated to some extent by the (limited) bid withdrawal provisions built into the SMR bidding rules under consideration. In this manner a bidder may compete aggressively for a package and then decide to withdraw, paying a penalty equal to the difference between the withdrawn bid and the final sale price if it is higher. Withdrawals are more frequent (and the associated penalties higher) with high complementarities, as would be expected. While the possibility of bid withdrawals helps bidders deal with the exposure problem to some extent, some losses did occur when bidders decided not withdraw or had to pay a penalty after a withdrawal.[20]

Summary. Pooling data across treatments and sessions, the revenue and efficiency results by auction format are given in Table 4.1. In terms of seller revenues and

[20]National bidders' penalties averaged 2% in the HC treatments and were negligible in the LC treatments. Regional bidders' penalties averaged 1% in the HC treatments and were negligible in the LC treatments.

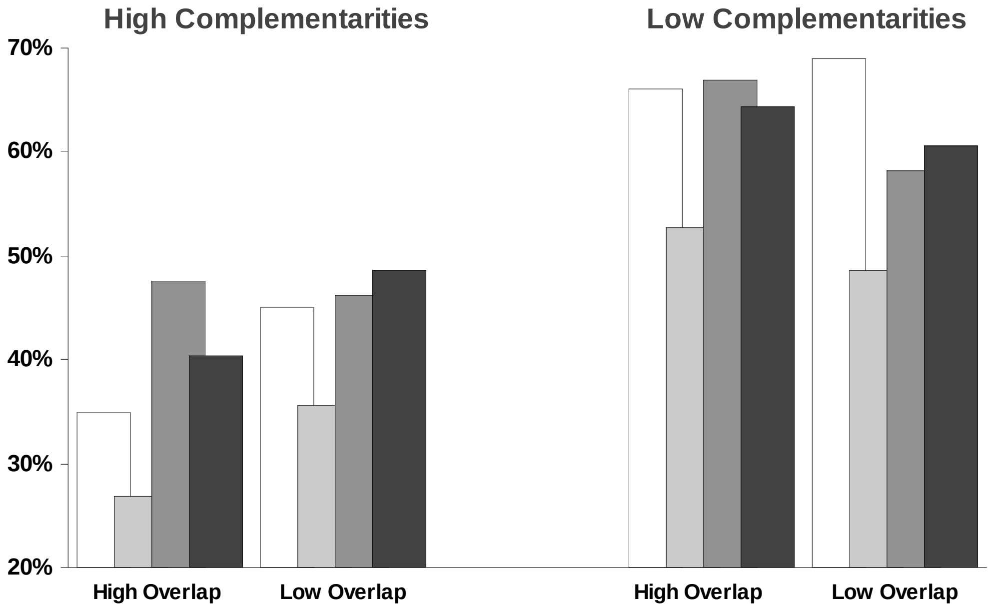


Figure 4.4. Bidder's Profits by Auction Format. The bars from light to dark (left to right) correspond to SMR, CC, SMRPB and RAD respectively.

bidder profits,[21] the combinatorial clock auction is best for the seller and worst for the bidders, but these results are not caused by bidder losses, which are not present in the CC sessions (see the losses rows for Nationals and Regionals in Appendix C.1). In a comparison with the other formats, the SMRPB auction with XOR bidding is the worst from the seller's point of view (lowest revenue and efficiency), and it is the best from the bidders' point of view (sum of profits for regionals and nationals).

The bottom row of Table 4.1 provides a perspective on the levels of the realized bidders' profits. The percentages in this row are calculated as ratios of actual bidders' profits (national profits plus regional profits) to profits that would result under collusion, i.e., when all bidders drop out at zero price levels resulting in a random allocation with the corresponding surplus, Srandom, being divided among the bidders. Note that realized profits are far from collusive levels, especially for the combinatorial clock format.

One reason why the SMRPB auction performs the worst in terms of efficiency

[21]In Table 4.1, the "Profit Regionals" row lists the total profit (as a percentage) for the group of 6 regional bidders, while the "Profit Nationals" row lists the combined profits (as a percentage) for the 2 national bidders.

Table 4.1. Summary Statistics by Auction Format.

	SMR	CC	RAD	SMRPB
Efficiency	90.2%	90.8%	93.4%	89.7%
Revenue	37.1%	50.2%	40.2%	35.1%
Profit Regionals	52.0%	38.5%	50.0%	51.3%
Profit Nationals	1.5%	2.3%	3.5%	3.5%
Profits/Collusive Profits	34.2%	26.0%	34.1%	34.9%

is that, in the presence of minimal complementarities, the requirement that bidders can only have 1 bid accepted (the XOR rule) may reduce efficiency, since bidders have to bid on many combinations of licenses to find all possible efficiency gains. (Even though our design, in which regional bidders face a choice of region problem, favors the XOR rule.) RAD, in contrast, more or less reduces to SMR with low complementarities while it enables bidders to extract the extra efficiency gains when complementarities are high. Another consideration may be that the inertia in the SMRPB price adjustment algorithm could exaggerate the threshold problem, since attempts to unseat large package bids may have delayed effects due to inertia. The next section explores the treatment differences in greater detail.

4.5 Individual Bidding Behavior

This section provides an analysis of bidding patterns to explain the main qualitative features of the aggregate data. In particular, with high complementarities, efficiency is significantly lower in SMR than in the 3 combinatorial auction formats. This suggests that bidders are competing conservatively for larger packages when package bids are not allowed, which could lead to an inefficient allocation.

In order to quantify the effect of "exposure risk" on bidder behavior in SMR, we consider a conditional logit model in which bidders choose among all combinations of licenses that are still feasible given their current activity limits. The conditional logit model includes 4 variables to explain the choice of a specific bidding basket, see Table 4.2. Since some sources of noise are individualspecific, we estimate robust standard errors allowing for correlation among observations generated from the same subject.

The "profit" variable is the difference between the value of the basket (the value of the combination of licenses that the bidder is either bidding on or provisionally winning) and the minimum required bid. The value calculations include possible synergies. The coefficient of this "Profit" variable shown in the top row of Table 4.2 is highly significant; as expected, bidders are more likely to bid on a basket when it yields a higher profit.[22]

The second row of the table shows the effect of the binary variable "PW," which assumes a value of 1 if the bidder is already provisionally winning at least 1 of the licenses in the basket. The highly negative coefficient indicates that subjects are not likely to raise their bids on licenses they are already provisionally winning, which is again intuitive. The third variable, "Inertia" is a dummy variable that is 1 if the set of licenses that the bidder was provisionally winning or bidding for in the last round is the same as the set of licenses the bidder is bidding for or provisionally winning in the current round.

Finally, "Exposure" is measured as the largest possible loss that a bidder might sustain when bidding on a certain combination of licenses.[23] We include an interaction term "Exposure * HC" to allow for the possibility that exposure has less of an effect with high complementarities. Note that exposure is significant and negative in both treatments,[24] suggesting that bidders are less likely to bid on baskets that entail the risk of winning licenses at prices above private values. The sign of the coefficient is robust across a variety of different specifications.

To illustrate the importance of exposure, let us consider a simple example for the case of high complementarities. Suppose the national bidder is interested in winning

[22]These estimations are based on all bids, including those that could result in negative profits. Such bids are more prevalent in the high-complementarities treatments (9.2%) than in the low-complementarities treatments (6.3%).

[23]Profit always refers to the difference between value and the minimum required bid for all licenses that the bidder is bidding on or provisionally winning. As noted earlier, we only consider baskets that yield a positive profit. However, when the bidder ends up winning only some of the licenses that he is currently provisionally winning or bidding for, he might sustain a loss.

[24]Running separate regressions for the low and high complementarities treatments yields exposure coefficients of -0.05 (0.007) with high complementarities and -0.11 (0.0038) with low complementarities.

Table 4.2. Bidding Behavior in SMR.

Conditional (fixed-effects) logistic regression

N	184,884	Log pseudolikelihood	-6853
Wald chi2(4)	1642	Pseudo R2	0.69
Prob > chi2	0		

Bid	**Coef.**	**Robust Std. Err.**	**z**	**P>|z|**
Profit	0.09	0.008	10.9	0
PW	-5.22	0.412	-12.6	0
Inertia	1.44	0.081	17.7	0
Exposure	-0.13	0.035	-3.7	0
Exposure*HC	0.08	0.036	2.2	0.03

either the national package ADEHIL or nothing at all (and is not the provisional winner on any license). License values are 25 on average so the national package is worth 300 on average. Consider a situation where license prices are 42 each so the minimum required bid would be 47 for each license, totaling 282 for the package. In this case, profit would be 18 but exposure would be 36, i.e., when the national bidder ends up winning only 3 out of the 6 possible licenses. Hence, the national bidder prefers stop bidding for the national package when its price is 252 (6 times the license price of 42) even though the value of the package is 300.

The second qualitative feature of the data is that efficiency is higher in RAD than in SMRPB for the minimal complementarities treatment. While activity in RAD is maintained by bidding or provisionally winning a sufficiently large number of different licenses, bidders in SMRPB have to bid on sufficiently large packages to maintain activity. As a result, bidders in RAD can simply bid on single licenses when there are no complementarities. In SMRPB, however, they typically bid on some profitable large package in order to maintain activity and are not also bidding on the subsets of that large package. Therefore, the number of possible allocations in SMRPB tends to be far lower than in RAD. If a bidder has high values for licenses A, B and C, for example, that bidder will typically bid on all 3 licenses separately in RAD. In SMRPB, the same bidder typically bids on package ABC only.

Table 4.3. Average Bid Characteristics when Complementarities Are Low (standard deviation).

Auction	Bidding Activity	Number of bids	Size of bids
SMRPB	2.91 (0.09)	1.62 (0.08)	2.44 (0.13)
SMR	2.60 (0.07)	2.60 (0.07)	1.00 (0.00)
RAD	2.69 (0.09)	2.11 (0.28)	1.60 (0.24)
CC	2.66 (0.07)	3.15 (0.52)	1.69 (0.13)

To evaluate why efficiency is lower in SMRPB when complementarities are low, we compare bidding behavior in terms of numbers and sizes of bids. The bidding activity column in Table 4.3 indicates that with low complementarities, subjects are bidding for or provisionally winning about 3 different licenses on average in all 4 formats. However, in SMRPB, bidders do so by bidding on fewer packages of a larger average size than in the other formats. The average number of bids submitted is lower in SMRPB than in any other auction format in every single one of the 6 sessions with low complementarities. Using a Wilcoxon matched-pairs signed-rank test, this difference is therefore significant ($p = 0.03$, $n = 6$). Similarly, the average size of the bids under SMRPB is higher in all 6 sessions. This bid size effect is significant in comparisons between SMRPB and any of the other auction formats ($p = 0.03$, $n = 6$). The consequence of having fewer bids of larger size is to create a fitting problem under SMRPB. This problem is not present for the other formats where activity can be maintained by submitting many smaller bids.

The seller's revenue is higher in the combinatorial clock auction than in any other auction format in all treatments of our experiment. In SMR, efficiency and thus also revenue is negatively affected by the exposure problem when complementarities are high. Moreover, the option to withdraw bids leads to a higher fraction of unsold licenses in SMR, which further reduces the seller's revenue.

In RAD and SMRPB, the threshold problem can potentially reduce the seller's revenue, since large bidders may be able to win packages at low prices when small bidders are unable to coordinate their actions. In order to test for the effects of the threshold problem in SMRPB and RAD, we look at whether small bidders bid up to

Table 4.4. Bidding up to Value.

Auction	Mean	Standard Deviation
CC	99.2%	9.9%
SMRPB	87.1%	5.3%
RAD	85.8%	3.0%

their values in periods in which they end up winning nothing. Since the threshold problem only pertains to small bids, we only look at bids on individual licenses and packages containing 2 licenses.

Recall that the combinatorial clock auction solves the threshold problem by forcing bidders to increase bids together on licenses for which there is excess demand. On average, small losing bids are closer to bidders' values in the combinatorial clock auction than in either RAD or SMRPB ($p < 0.001$ for both comparisons using a Wilcoxon matched-pairs signed-rank test with 12 observations). The differences between RAD and SMRPB are not significant (p = 0.23, n = 12), see also Table 4.4.

Table 4.5. Size of Jumpbid (Bid - Minimum Required Bid).

Auction	Treatment	Mean	Standard Deviation
CC	All	0	0
RAD	HC	3.1	2.2
RAD	LC	1.3	0.8
SMRPB	HC	4.1	2.1
SMRPB	LC	2.6	1.2

If the threshold problem is indeed the reason why small bidders fail to bid up to their values, one would expect large bidders in SMRPB and RAD to submit aggressive "jump bids," i.e., to bid more than the minimum required bid in the early rounds of the auction. Taking the average across all bids submitted during the first 5 rounds, the difference between the bid price and the minimum required bid is higher in SMRPB than in RAD both with high and low complementarities, see Table 4.5. However, these differences are not significant (pooling data from low and high complementarities yields p = 0.15 using a Wilcoxon matched-pairs signed-rank test with 12 observations). Since the combinatorial clock auction does not allow jump bids, both these differences are higher for RAD and SMRPB than for CC (see Table 4.5).

4.6 Conclusion

The simultaneous multi-round auction is considered to be a remarkably successful application of game theory, with careful attention to the details of implementation by policy makers. This auction format is currently used around the world, and government officials in other agencies now routinely consult the FCC on auction design matters. Concerns about the effects of value complementarities have convinced many people that new procedures need to be developed and tested. In particular, the FCC developed a package bidding variant of the SMR auction, known as SMRPB. This chapter compares the performance of these 2 alternatives and that of 2 other package-bidding formats: the combinatorial clock (Porter et al. 2003) and the RAD auction (Kwasnica et al. 2005).

The experiments were conducted with a common *jAuctions* bidder interface and parallel sets of value draws, for an array of structural and auction format treatments. The combinatorial auction procedures used (RAD, SMRPB, and CC) all result in higher efficiency than the currently used SMR procedure when value complementarities are present. It is important to emphasize that value complementarities are not just a theoretical possibility; a package of 3 bandwidth segments sold for about 5 times as much as a single segment in a recent FCC auction that offered a very limited menu of pre-specified package bidding options. Complementarities are almost surely significant for other potential applications of package bidding such as emissions permits for successive years.

However, of the 3 combinatorial auction types, SMRPB performed worst in terms of revenue and efficiency. One distinguishing feature of SMRPB is the XOR rule, which allows each bidder to have at most a single winning bid. A bidder who is interested in obtaining 1 or more licenses in a certain region thus has to bid on all possible combinations of those licenses. In the experiment, however, bidders submit only a few bids per round, in which case the additional constraint of at-most-one-winning bid per bidder becomes detrimental for efficiency and revenue. The poor performance of SMRPB reported here was a main factor in the FCC's decision not to implement it.

The FCC subsequently decided to implement package bidding for 1 of the 5 blocks in the recently conducted 700 MHz auction. Unlike the fully flexible package bidding formats considered in this paper, the FCC opted to use a simple format with a single 50-state package and 2 additional packages (Atlantic and Pacific). This setup is a simple version of the Hierarchical Package Bidding mechanism proposed by Goeree and Holt (2008). Under this mechanism, predefined packages are structured in a hierarchical manner and after each round of bidding, prices for all licenses and packages are determined such that they signal the bid amounts required to unseat the current provisional winners.

Without extensive knowledge of bidders' valuations, there will be some efficiency loss associated with using predefined packages. However, the simplicity of the hierarchical package structure (e.g., individual licenses, non-overlapping regional packages, and a single nationwide package) avoids fitting problems that can arise with fully flexible package bidding. For example, if a nationwide package bid is winning then the non-overlapping nature of the regional packages together with the simple pricing feedback allows regional bidders to avoid the threshold problem. This approach was tested using laboratory experiments based on 2-layer and 3-layer hierarchies and the results were cited by the FCC as a factor in their decision to use Hierarchical Package Bidding in the recent 700 MHz auction.

Chapter 5

A Correction and Reexamination of "Stationary Concepts for Experimental 2×2 Games"

This chapter is based on a paper written jointly with Colin F. Camerer, and Jacob K. Goeree.

5.1 Introduction

A recent paper by Reinhard Selten and Thorsten Chmura (2008) (henceforth SC) reports laboratory results for 12 different 2×2 games with a unique mixed-strategy equilibrium. These binary-choice games are relatively simple and provide a natural testbed for alternative models that aim to predict long-run, or stationary, outcomes of play. SC consider 5 such models: Nash equilibrium, quantal response equilibrium, action-sampling equilibrium, payoff-sampling equilibrium, and impulse-balance equilibrium.

Nash equilibrium assumes that players have correct beliefs about others' play and that players best respond to those beliefs. Quantal response equilibrium (QRE) replaces the requirement of best responses with "better responses," i.e., players are *more likely* to choose the option with the higher expected payoff but they do not necessarily choose the best option all the time. QRE does assume that players' beliefs are correct on average, i.e., beliefs are not systematically biased. Action-sampling equilibrium describes the long-run outcome when players best respond to a finite sample of their opponent's previous actions.[1] Payoff-sampling equilibrium describes the long-run outcome when players form 2 finite samples of their past payoffs, 1 for each option, and select the one with the highest total payoff. Finally, impulse-balance equilibrium is based on the idea that players take into account foregone payoffs. If the option not chosen would have yielded higher payoffs, then there is an "impulse" to change (and, importantly, "losses" of foregone payoff are weighted twice as heavily as gains). impulse-balance equilibrium corresponds to the long-run outcome where, for both players, expected impulses are equal across the 2 options.

SC conclude that Nash and QRE fit worse than the other 3 concepts. They write,

> It is remarkable that the newer concepts of impulse-balance equilibrium, payoff-sampling equilibrium, and action-sampling equilibrium clearly outperform the more established concepts of quantal response equilibrium [QRE] and Nash equilibrium. All the relevant comparisons are highly significant. This is perhaps the most important result of the statistical tests. (p. 962)

The first purpose of this comment is to correct the reported model fits for 2 of the 5 concepts, QRE and action sampling, which are incorrect for all 12 games in SC. We report the correct results for these 2 models (and some other small corrections). The corrected fits for QRE are close to the other 3 non-Nash concepts, which overturns the most novel (and to some, surprising) part of their conclusion, viz., QRE fits as well as the other concepts, not worse (as SC concluded).[2]

[1] Action-sampling equilibrium is closely related to the "stochastic learning equilibrium" concept introduced by Goeree and Holt (2002) where players make a noisy best response to a weighted average of their opponents' past decisions. Rather than putting a weight of 1 on a fixed number of past observations and a weight of 0 on observations that are in the more distant past, as in action-sampling equilibrium, the stochastic learning equilibrium assumes that weights decline continuously for more distant observations (e.g., geometrically). Stochastic learning equilibrium has been shown to yield an improved fit over QRE in some contexts (see, e.g., Capra et al. 2002).

[2] It is true that with the corrected analysis, Nash predictions *do* fit worse than the other 4 concepts. However, the ability of other models to explain deviations from Nash play has been shown in many previous experiments (see Camerer 2003 for a book-length summary). This part of their conclusion is solid but is only original in its emphasis on the sampling and impulse balance models.

Fit measures and statistical tests show that the 4 non-Nash models are about equally accurate. SC note this fact (but for 3 models, not all 4) and suggest a research direction as follows:

> It is not easy to understand why the predictions of the 3 newer concepts are not very far apart, in spite of the fact that they are based on very different principles. This is perhaps peculiar to our sample. *It would be desirable to devise experiments that permit a better discrimination among the 3 concepts.* (p. 965, emphasis ours).

The second purpose of our comment is to extend their analysis, by showing how 2 different games reported several years ago *do* "permit a better discrimination" among some of the concepts. The first game was explicitly designed to show that *no* quantal response equilibrium (logit or otherwise) could explain observed behavior (see Game 4 and Proposition 1 in Goeree et al., 2003). Applying impulse-balance equilibrium to this game works like "magic:" it explains observed behavior almost perfectly. So this game is capable of differentiating between 2 of the concepts, impulse-balance equilibrium and risk-neutral QRE, which fit equally well in SC's data.

The results also highlight one of the crucial assumptions underlying impulse-balance equilibrium: impulses are defined relative to a security level (the max-min payoff) and it is assumed that losses with respect to this security level are weighed *twice as much* as gains. While impulse-balance equilibrium is ostensibly a parameter-free concept (since the loss aversion coefficient is fixed to 2), this additional assumption about players' different reactions to foregone losses and gains is not innocuous. For the game designed by Goeree et al. (2003), it is the assumption of loss aversion that makes impulse-balance equilibrium predict well.[3] The favorable comparison of impulse-balance equilibrium (with loss-aversion built in) to risk-neutral QRE reinforces the main point of the Goeree et al. (2003) paper, i.e., that some degree of

[3] In a 1-parameter extension of impulse-balance equilibrium where the weight for gains is fixed to be 1 but the weight for losses is a free parameter, γ, we obtain an estimate of $\gamma = 2.07$. In other words, the degree of loss aversion ($\gamma = 2$) that is hardwired into the impulse-balance equilibrium concept is nearly optimal for the data set considered.

risk or loss aversion is needed to explain behavior in their game. Indeed, as we show below, if the other concepts are augmented with loss aversion, they predict behavior quite well (and even better than impulse-balance equilibrium).

The second class of games that discriminate among concepts are asymmetric 2×2 matching pennies games (e.g., Ochs, 1995). We report new analyses using the data of McKelvey et al. (2000). In these games, loss aversion plays no role since security levels are 0 and payoffs are non-negative. We find that impulse-balance equilibrium fits a little better than QRE but much worse than action- or payoff-sampling. These 2 reanalyses of older data take up the search for games that discriminate better among stationary concepts that SC called for, and show that the loss-aversion built into impulse-balance equilibrium accounts for some of that concept's success on risky games such as game 4 from Goeree et al. (2003).

5.2 Reexamining the Selten and Chmura (2008) Results

Table 5.1 shows data averages and model predictions for each of the 12 experimental games that SC ran. This table, and all subsequent tables and figures report corrections of their results in a visual form identical to their originals. The bold numbers indicate discrepancies between our results and those of SC. In particular, we find (i) a different impulse balance prediction for Game 1, (ii) a different data average for Game 3, (iii) a different optimal sample size ($n = 12$) and, hence, different predictions for action-sampling equilibrium (see Figure 5.1 for the mean-squared distances by sample size), and (iv) vastly different predictions for the QRE model: the precision parameter we estimate using the mean-squared distance objective function is $\lambda = 1.05$, much lower than the estimate reported by SC ($\lambda = 8.84$).[4]

At this lower value of λ, the QRE predictions (see Table 5.1) are much different from Nash predictions and much closer to the data. The improved fit is illustrated

[4]Using maximum-likelihood techniques yields an estimate $\lambda = 0.99$.

Table 5.1. 5 Stationary Concepts Together with the Observed Relative Frequencies for Each of the Experimental Games. Note: λ is the logit precision parameter, n is the optimal sampling size for action or payoff sampling.

		Nash	QRE **(λ = 1.05)**	Action-sampling **(n = 12)**	Payoff-sampling (n = 6)	Impulse Balance	Observed average of 12 observations
Game 1	U	0.091	**0.042**	**0.090**	0.071	**0.068**	0.079
	L	0.909	**0.637**	**0.705**	0.643	0.580	0.690
Game 2	U	0.182	**0.154**	**0.193**	0.185	0.172	0.217
	L	0.727	**0.579**	**0.584**	0.569	0.491	0.527
Game 3	U	0.273	**0.168**	**0.208**	0.152	0.161	**0.163**
	L	0.909	**0.770**	**0.774**	0.771	0.765	0.793
Game 4	U	0.364	**0.275**	**0.302**	0.285	0.259	0.286
	L	0.818	**0.734**	**0.719**	0.726	0.710	0.736
Game 5	U	0.364	**0.307**	**0.329**	0.307	0.297	0.327
	L	0.727	**0.657**	**0.643**	0.654	0.628	0.664
Game 6	U	0.455	**0.417**	**0.426**	0.427	0.400	0.445
	L	0.636	**0.607**	**0.596**	0.597	0.600	0.596
		Nash	QRE **(λ = 1.05)**	Action-sampling **(n = 12)**	Payoff-sampling (n = 6)	Impulse Balance	Observed average of 6 observations
Game 7	U	0.091	**0.042**	**0.090**	0.060	0.104	0.141
	L	0.909	**0.637**	**0.705**	0.691	0.634	0.564
Game 8	U	0.182	**0.154**	**0.193**	0.222	0.258	0.250
	L	0.727	**0.579**	**0.584**	0.602	0.561	0.587
Game 9	U	0.273	**0.168**	**0.208**	0.154	0.188	0.254
	L	0.909	**0.770**	**0.774**	0.767	0.764	0.827
Game 10	U	0.364	**0.275**	**0.302**	0.308	0.304	0.366
	L	0.818	**0.734**	**0.719**	0.730	0.724	0.700
Game 11	U	0.364	**0.307**	**0.329**	0.338	0.354	0.331
	L	0.727	**0.657**	**0.643**	0.650	0.646	0.652
Game 12	U	0.455	**0.417**	**0.426**	0.404	0.466	0.439
	L	0.636	**0.607**	**0.596**	0.599	0.604	0.604

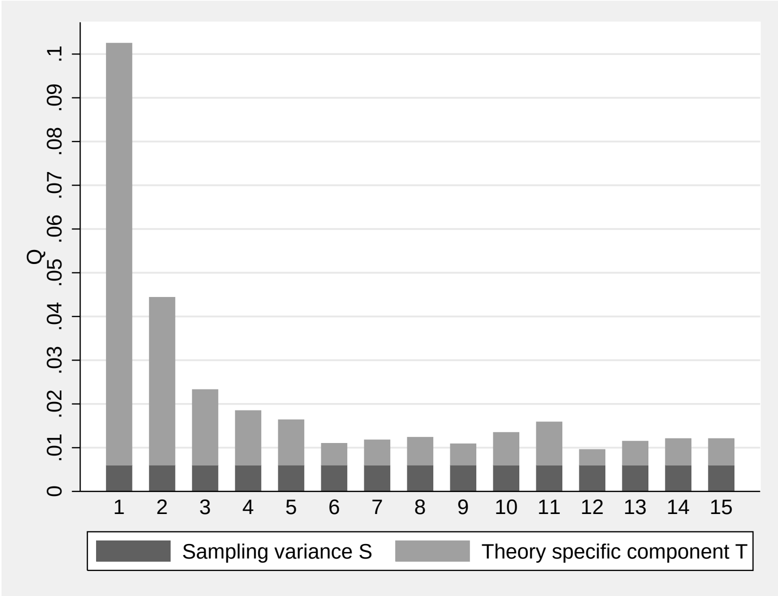


Figure 5.1. Overall Mean Squared Distance for the Action-Sampling Equilibria with Different Sample Sizes (cf. SC Figure 9).

by Figure 5.2, which shows data averages and model predictions and parallels Figure 8 in SC. Using an "ocular metric" suggests that the predictions of the alternative models are remarkably close to each other and to the data averages. To quantify this we also computed the sample variance and theory-specific variance as in SC, which are shown in Figure 5.3 (cf. Figure 12 in SC).

Following SC, it is useful to evaluate the stationary concepts using data from the first 100 periods and final 100 periods, see their Figure 13. The correction can be found in Figure 5.4, which displays the theory-specific variances for the different concepts (excluding Nash) by blocks of 100 periods and for all 200 periods. Note that impulse-balance equilibrium is the best model in the first 100 periods and the worst in the second 100 periods. Furthermore, compared to its overall performance, impulse-balance equilibrium predicts 30% worse in the first 100 periods *and* 14% worse in the final 100 periods. In contrast, the other concepts do better in the final 100 periods compared to the first 100 periods (or all 200 periods), and outperform impulse-balance equilibrium in the final 100 periods.

To test for significant differences, SC report ten pairwise comparisons of the 5

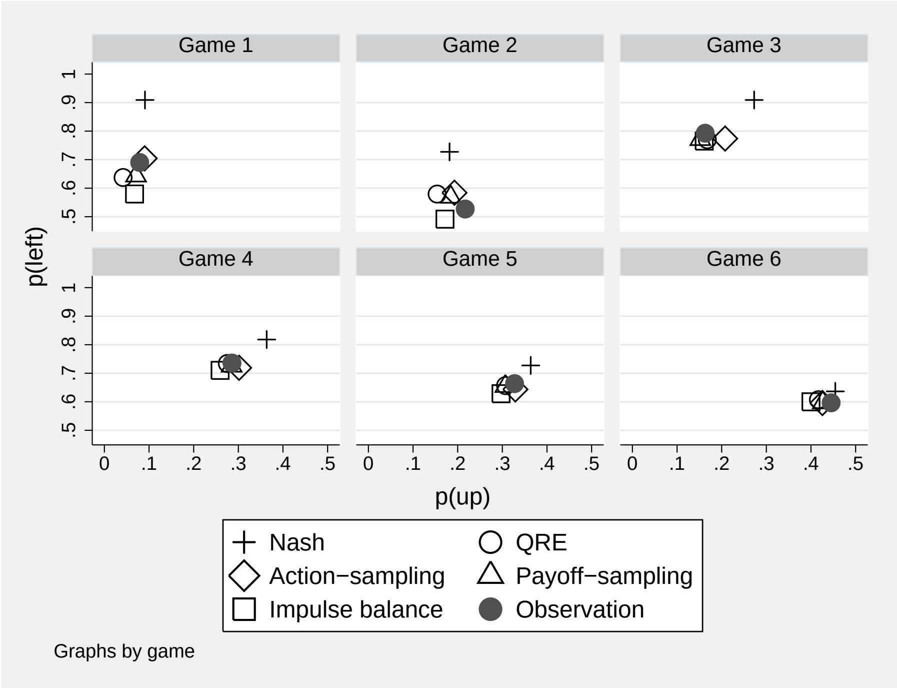


Figure 5.2. Visualization of the Theoretical Equilibria and the Observed Average in the Constant Sum Games (cf. SC Figure 8).

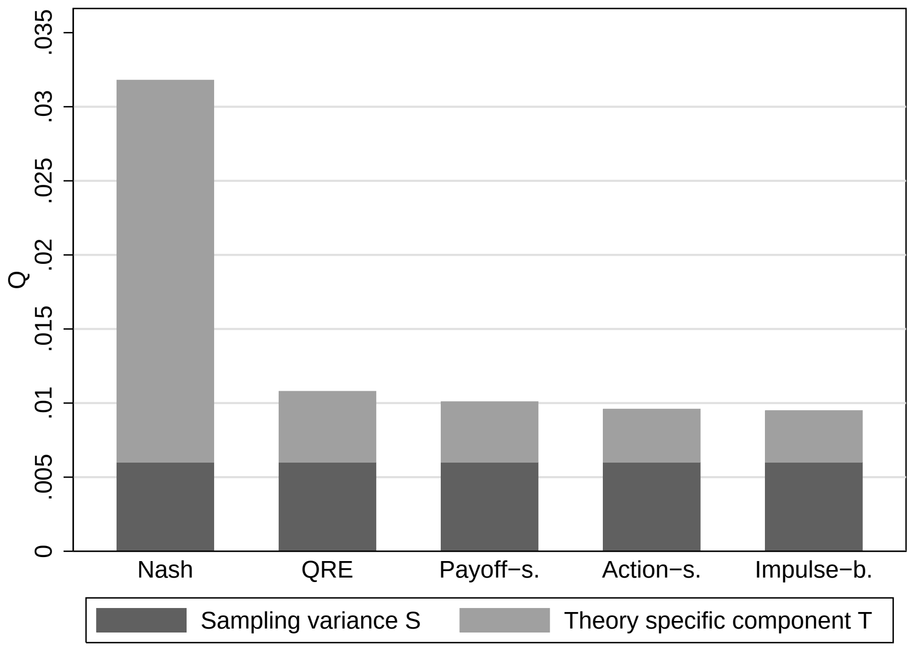


Figure 5.3. Overall Mean Squared Distances of the 5 Stationary Concepts Compared to the Observed Average (cf. SC Figure 12).

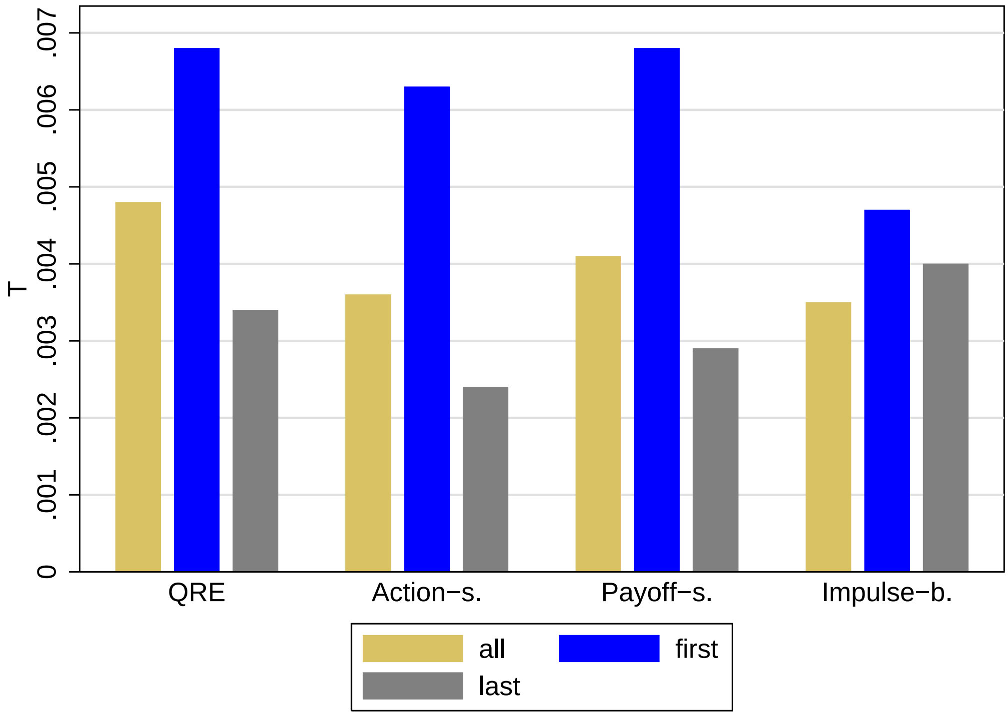


Figure 5.4. Theory Specific Squared Distances of the 5 Stationary Concepts Compared to the Observed Average by Blocks of 100 Periods (cf. SC Figure 13).

different models based on the Wilcoxon matched-pairs signed-rank test. Each model generates a squared deviation (between observed and predicted frequencies) for each game, and the Wilcoxon test is applied to the differences in these squared deviations, treating each session as an independent observation. As noted by a referee, the Wilcoxon test assumes that the distributions of these differences have the same shape, which is not necessarily true in the data. The Kolmogorov–Smirnov test relaxes this assumption, and when we apply it to the ten different pairs of models, we cannot reject the null hypothesis of identical distributions except when comparing action sampling and impulse-balance equilibrium (5% level) and when comparing any of the non-Nash models with Nash. These results are corroborated using a robust rank-order test (Fligner and Policello 1981). This test also relaxes the assumption of identical shape and only requires both distributions to be symmetric. We find that none of

Table 5.2. P-Values in Favor of Row Concepts, 2-Tailed Matched-Pairs Wilcoxon Signed-Rank Tests, n = 108 (Rounded to the Next Higher Level Among 0.1 Percent, 0.2 Percent, 0.5 Percent, 1 Percent, 2 Percent and 10 Percent).

	Action-sampling equilibrium	Payoff-sampling equilibrium	Quantal response equilibrium	Impulse balance equilibrium	Nash equilibrium
Action-sampling equilibrium		**n.s.** n.s.	**2 percent** *10 percent* 10 percent	**n.s.** *5 percent*	**0.1 percent** *0.1 percent* 0.1 percent
Payoff-sampling equilibrium	*n.s.*		**5 percent** *0.1 percent*	**n.s.** *0.1 percent*	**0.1 percent** *0.1 percent* 0.1 percent
Quantal response equilibrium		n.s.		**n.s.** *0.1 percent*	**0.1 percent** *0.1 percent* 1 percent
Impulse balance equilibrium	n.s.	5 percent	0.5 percent		**0.1 percent** *0.1 percent* 0.1 percent

Above: All 108 Experiments; *Middle: 72 Constant-Sum Game Experiments*; Below: 36 Non-Constant Sum Game Experiments.

the pairwise comparisons are significant except when comparing Nash to any of the non-Nash models. For completeness, Table 5.2 reproduces SC's pairwise comparisons using the Wilcoxon matched-pairs signed-rank test (see their Table 3): the entries display rounded p-values for 2-tailed tests. Combined, the various statistical tests confirm the "no difference" result suggested by Figure 5.3, there is no clear ranking among 4 of the stationary concepts, except that that all 4 models always do better than Nash.

To summarize, *the SC design does not differentiate well among the stationary concepts they consider.* As noted in the Introduction, an extension to games which do differentiate well across concepts is therefore of interest in addressing the central point of the SC paper.

5.3 Differentiating Stationary Concepts in Other Data Sets

5.3.1 A Matching Pennies Game with Safe and Risky Choices

Goeree et al. (2003) designed the game in the left panel of Figure 5.5 to illustrate the limitations of QRE in terms of explaining behavior when other factors, such as risk aversion, are likely to play a role. In particular, both players have a "safe" choice that rewards either 160 or 200 and a "risky" choice that rewards either 10 or 370. Goeree et al. (2003) prove that in *any* quantal response equilibrium (logit or otherwise), the column player's probability of playing Right is greater than 0.5. Risk aversion, however, will naturally steer players towards the safer option of playing Left.

In the experiment, the aggregate choice frequencies were 65% for Left and 47% for Up, which contradicts QRE predictions. To compute the impulse-balance equilibrium of the game, note that the max-min payoff is 160 for both players. Subtracting 160 from all payoffs and multiplying by 2 if the resulting number is negative yields the transformed game in the right panel of Figure 5.5. The condition that, for both players, the expected impulses even out yields: $300\,p_D\,q_R = 170\,p_U\,q_L$ and $300\,p_U\,q_R = 170\,p_D\,q_L$, which implies that $\frac{p_U=1}{2}$ and $q_L = \frac{30}{47} \approx 0.64$. impulse-balance equilibrium fits almost perfectly!

	Left	Right
Up	200, 160	160, 10
Down	370, 200	10, 370

	Left	Right
Up	40, 0	0, -300
Down	210, 40	-300, 210

Figure 5.5. A Matching Pennies Game with "Safe" (200/160) and "Risky" (370/10) Choices (Left) and the Transformed Game (Right).

However, it is important to point out the importance of the loss aversion that is built into the impulse-balance equilibrium concept. If losses and gains were weighed equally, the relevant conditions would be: $150\,p_D\,q_R = 170\,p_U\,q_L$ and $150\,p_U\,q_R = 170\,p_D\,q_L$, which implies that $\frac{p_U=1}{2}$ and $q_L = \frac{15}{32} \approx 0.47$. In other words, without loss aversion, the impulse-balance equilibrium predictions are on the wrong side of

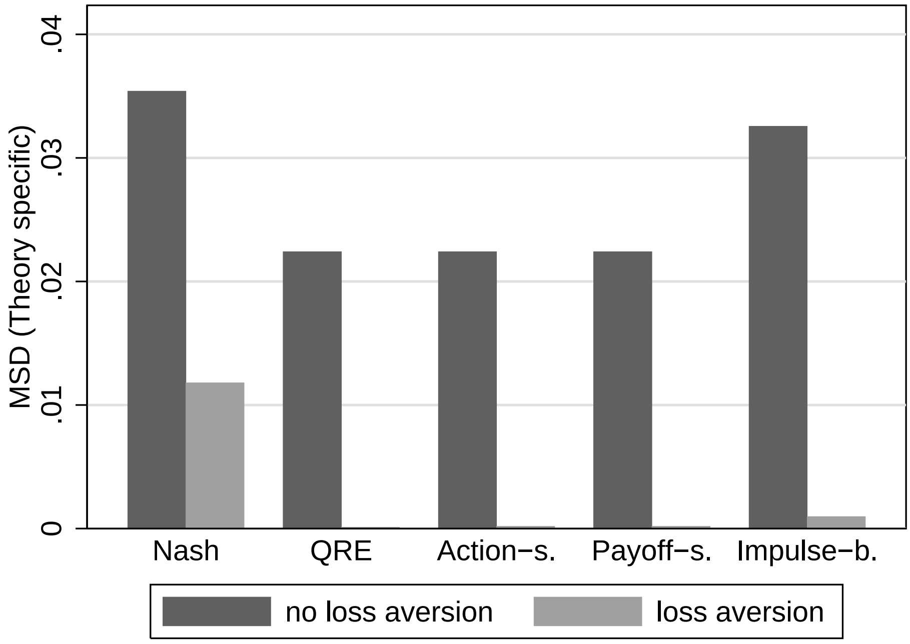


Figure 5.6. Theory-Specific Mean-Squared Distances for Game 4 from Goeree et al. (2003) for Models with and without Loss Aversion.

0.5 just like the risk-neutral QRE predictions reported by Goeree et al. (2003). The rightmost bars in Figure 5.6 show the theory-specific mean-squared deviations for impulse balance, with and without loss aversion. The other pairs of bars display the analogous results for Nash and non-Nash models. The latter do slightly better than impulse-balance equilibrium once they are also augmented with loss aversion (weighing losses twice as much as gains). Clearly, it is the loss aversion assumption that drives the goodness of fit for this game.

5.3.2 Asymmetric Matching Pennies Games

A second test of the stationary concepts is provided by the experiments of McKelvey et al. (2000). They used games with an "asymmetric matching pennies" structure (Ochs, 1995), shown in Figure 5.7. The Row player earns a positive amount if the players match on "Heads" or "Tails" (and then the Column player earns nothing), or the Column player earns a positive amount if the players mismatch (and then the

Row player earns nothing). McKelvey et al. (2000) consider 4 related games: in game A, $X = 9$; in game D, $X = 4$; game B payoffs are the same as game A's except Column payoffs are multiplied by 4; game C payoffs are the same as game A's except all payoffs are multiplied by 4.

	Heads	Tails
Heads	X, 0	0, 1
Tails	0, 1	1, 0

Figure 5.7. An Asymmetric Matching Pennies Game.

To compute the impulse balance equilibria for these games note that the max-min payoff is 0 for both players (as is the second-lowest payoff) so the transformed games are equivalent to the original games. A simple calculation shows that for the game in Figure 5.7, the impulse-balance equilibrium predictions for the Row and Column players are[5]

$$p_H = \frac{\sqrt{X}}{1+\sqrt{X}}, \quad q_H = \frac{1}{1+\sqrt{X}}. \tag{5.1}$$

Since multiplicative factors drop out of the impulse-balance equilibrium calculations, the predictions for games A, B, and C are identical: $p_H = 0.75$ for Row and $q_H = 0.25$ for Column, while for game D we have $p_H = 0.67$ for Row and $q_H = 0.33$ for Column.

The aggregate choice frequencies observed in the experiments are shown in Table 5.3 together with the predictions of the 5 stationary concepts as well as parameter estimates. We apply the same tests that we applied to the SC data to establish whether there are any significant differences in terms of goodness of fit. For that purpose, we compute the mean squared deviation separately for each one of 16 ex-

[5]Interestingly, the predictions in (5.1) are identical to those obtained from a "Luce"-type quantal response equilibrium where choice probabilities are proportional to expected payoffs, e.g.,

$$p_H = \frac{\pi_H^R}{\pi_H^R + \pi_T^R}, \quad q_H = \frac{\pi_H^C}{\pi_H^C + \pi_T^C}.$$

The expected payoffs on the right side depend on the choice probabilities: $\pi_H^R = q_H X$, $\pi_T^R = 1 - q_H$ and $\pi_H^C = 1 - p_H$, $\pi_T^C = p_H$. It is straightforward to show the fixed-point probabilities are those in (5.1).

Table 5.3. 5 Stationary Concepts Together with the Observed Relative Frequencies for Each of the Experimental Games in McKelvey et al. (2000). Note: λ is the logit precision parameter, n is the optimal sampling size for action or payoff sampling.

		Nash	QRE (λ = 5.35)	Action-sampling (n=3)	Payoff-sampling (n=3)	Impulse Balance	Observed average	n
Game A	U	0.500	0.691	0.643	0.625	0.750	0.643	6
	L	0.100	0.115	0.291	0.276	0.250	0.241	
Game B	U	0.500	0.550	0.643	0.625	0.750	0.630	4
	L	0.100	0.104	0.291	0.276	0.250	0.244	
Game C	U	0.500	0.551	0.643	0.625	0.750	0.594	4
	L	0.100	0.101	0.291	0.276	0.250	0.257	
Game D	U	0.500	0.619	0.643	0.625	0.667	0.550	2
	L	0.200	0.218	0.291	0.276	0.333	0.328	
MSD		0.039	0.028	0.011	0.010	0.023		

perimental games.[6]

Table 5.4 contains the p-values of these pairwise comparisons. For each pairwise comparison, we run a Kolmogorov-Smirnov test (KS), a robust rank order test (FP) and a Wilcoxon matched-pairs signed-rank test (W). Both payoff sampling and action sampling perform substantially better than any of the other equilibrium concepts. Moreover, impulse balance and QRE yield a significantly better fit than the Nash equilibrium.

5.4 Conclusion

Correcting for errors in estimating 2 of the 5 stationary concepts SC compare, QRE and action-sampling equilibrium, we find that their design does not differentiate among the non-Nash stationary concepts that were considered. They also suggest it is desirable to create games which discriminate among these non-Nash theories, a direction we pursue. We first compare the fit of the 5 stationary concepts on a 2×2 game with a risky and a safe choice (game 4 from Goeree et al., 2003) and find that

[6]McKelvey et al. 2000 run 2 different games within each session. We treat each one of these 2 games as an independent observation even though the same subjects participate in both games.

Table 5.4. P-Values in Favor of Row Concepts for Pairwise Tests on the McKelvey et al. 2000 Data. Column "KS" contains p-values of a Kolmogorov-Smirnov test, "FP" p-values of a robust rank-order test (Fligner and Policello 1982) and column "W" p-values of a Wilcoxon matched-pairs signed-rank test, $n = 16$.

	Action-sampling equilibrium			Impulse balance equilibrium			Quantal response equilibrium			Nash equilibrium		
	KS	FP	W	KS	FP	W	KS	FP	W	KS	FP	W
Payoff-sampling equilibrium	0.91	0.64	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Action-sampling equilibrium				0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Impulse balance equilibrium							0.91	0.58	0.23	0.02	0.00	0.02
Quantal response equilibrium										0.07	0.03	0.00

the impulse-balance equilibrium fits particularly well. It turns out that this superior fit is due to the hardwired loss aversion that characterizes impulse balance. When we incorporate loss aversion in the other non-Nash concepts, their fit is even slightly better than the fit of the impulse-balance equilibrium. Asymmetric matching pennies games in which loss aversion plays no role (games A-D from McKelvey et al. 2000) also allow us to discriminate among the 5 equilibrium concepts. We find that action sampling and payoff sampling fit significantly better than any of the other models while all models fit better than the Nash equilibrium.

While we do find differences in goodness of fit between the models we compare, they also share similar features. For some games, impulse-balance equilibrium *coincides* with a specific QRE model (see footnote 5). Furthermore, by varying the sample size in the action-sampling equilibrium from 1 to infinity, the implied behavior ranges from purely random to Nash behavior in the 2×2 games studied here. This is akin to varying the precision parameter in a QRE model. Both models trace out a 1-dimensional curve in the 2-dimensional unit square corresponding to players' choice probabilities, and payoff-sampling yields yet another such curve. Evaluating theories via a simple horse race is simply asking which of these curves comes closest to the observed data points. However, the value of the models we compare cannot be measured by their goodness of fit alone. Other considerations such as their analytical tractability or their theoretical appeal are also important. Moreover, further research is needed in order to establish how well these models can explain behavior in other types of games.

Appendix A

Appendix to Chapter 2

A.1 Treatment Sequence No-Cascade Periods

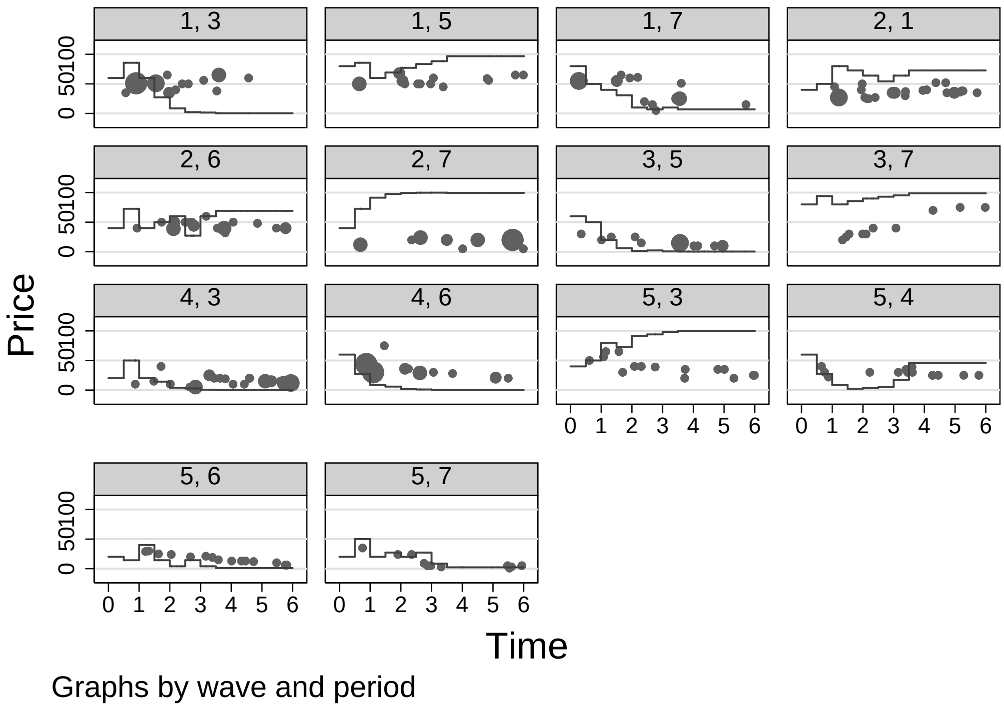


Figure A.1. Prices in Treatment Sequence no-Cascade Periods. The size of the dots is proportional to the number of units exchanged. The line corresponds to the expected value of the asset given all private signals.

A.2 Treatment Baseline No-Cascade Periods

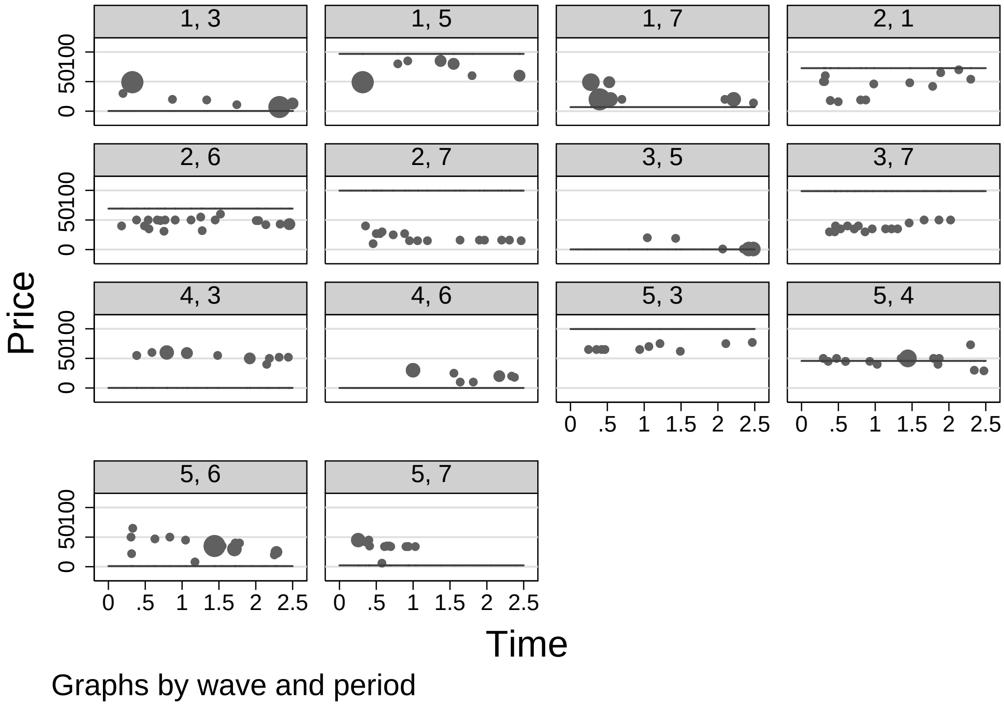


Figure A.2. Prices in Treatment Baseline no-Cascade Periods. The size of the dots is proportional to the number of units exchanged. The line corresponds to the expected value of the asset given all private signals.

A.3 Instructions for Treatment Baseline

The instructions for all treatments were presented as a powerpoint-presentation. The first 10 slides explain how the market works and each subsection corresponds to a slide. The last 10 slides describe the user interface and for that part of the instructions, each figure (figures A.4 – figure A.13) corresponds to a slide.

A.3.1 Welcome to SSEL

- Welcome and thank you for participating in today's experiment.

- Today's experiment will involve a series of markets. At the end of the experiment you will be paid in cash for your participation. Each of you may earn different amounts. The amount you earn depends on your decisions, chance, and on the decisions of others.
- You will be using the computer for the entire experiment, and all interaction between you and others will be through computer terminals. Please DO NOT socialize or talk during the experiment.
- If you have any questions, raise your hand and your question will be answered so everyone can hear.
- Please do not use any programs unrelated to the experiment.

A.3.2 Markets, Commodities, and Traders

- The experiment consists of a series of 7 market periods. Each period lasts 2.5 minutes during which you can submit as many bids and asks as you want.
- In each period, you will be in a group of 8 traders.
- In each period, one commodity labeled A will be traded.

A.3.3 Common Value

- The value of the commodity A is the same for all traders.
- This common value is either 0 or 100 points, each number being equally likely. Imagine we flip a fair coin in each market period to determine the common value of commodity A.
- Traders will NOT know the value of A at the time of trading. Instead, each trader receives a common-value signal (CVS) about the commodity's common value at the beginning of each period.

A.3.4 Common Value Signal I

- The common-value signal (CVS) is either 0 or 100. It can be either strong or weak.
  - If you receive a strong CVS, your signal corresponds to the actual common value with probability 0.8.
  - If you receive a weak CVS, your signal corresponds to the true common value with probability 0.6.
  - Each CVS is equally likely to be strong or weak. You will be told whether your signal is strong or weak.
- Each trader receives exactly one common-value signal at the beginning of each market period. You can only see your own CVS, not the signals of other traders.
- Common-value signals are independent across traders and periods.

A.3.5 Common Value Signal II

This slide graphically explains how the CVS is determined and consists of figure A.3.

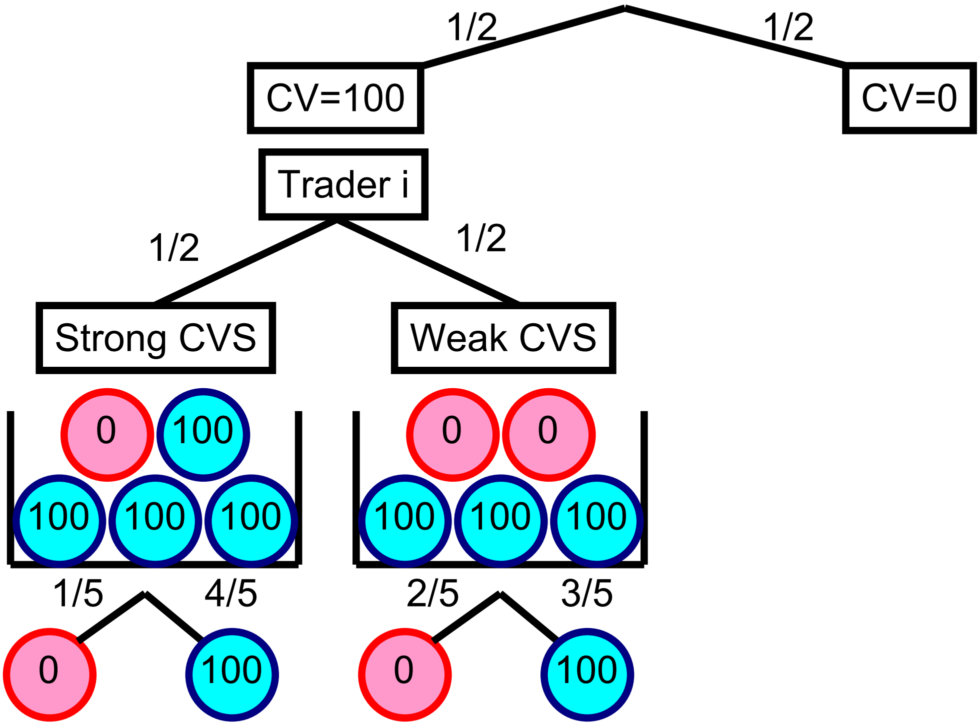


Figure A.3. Drawing Signals.

A.3.6 Making Profits from Trades

- Traders make money by buying or selling commodities.
  - Your profit from a purchase is simply: common value - price you pay.
  - Your profit from a sale is simply: price you get paid - common value.
- In other words, you make money by BUYING BELOW THE COMMON VALUE and/or by
- SELLING ABOVE THE COMMON VALUE.

A.3.7 Maximum Number of Units

- You will start with 5 units of A and you have values for at most 10 units of A.

- In other words, if you hold more than 10 units of A, the additional units will be worthless to you even though you paid for them—so buying additional units results in a loss!

- You will not be able to sell more units than you currently hold. In other words, short sales are not allowed.

A.3.8 Market Prices

Whenever someone submits a new order, the computer checks if it can transact orders. Here's an example:

- Trader 1 bids 10 for A

- Trader 2 bids 65 for A (*)

- Trader 3 asks 35 for A (*)

- Trader 2 asks 55 for A

Here the orders with a (*) transact. The transaction price will be 50, that is halfway between the best ask price (35) and the best bid price (65).

A.3.9 Summary

- The experiment consists of a series of 7 market periods preceded by 1 practice period that does not affect earnings.

- Each period lasts 2.5 minutes, and you can submit as many bids/asks as you want during that time. After each submission the computer checks which orders can transact and determines corresponding prices.

- A trader's profit is simply:

  - common value - price paid when a unit of A is bought

  - price received - common value when a unit of A is sold.

- In other words, you make money by buying below the common value or by selling above the common value.
- Traders have values for at most 10 units of commodity A.

A.3.10 Concluding Remarks

- The exchange rate used in the experiment is 100 points for $1.
- You also receive a $5 participation fee.
- You will be paid at the end of the experiment the total amount you have earned in all of the periods. You need not tell any other participant how much you earned.

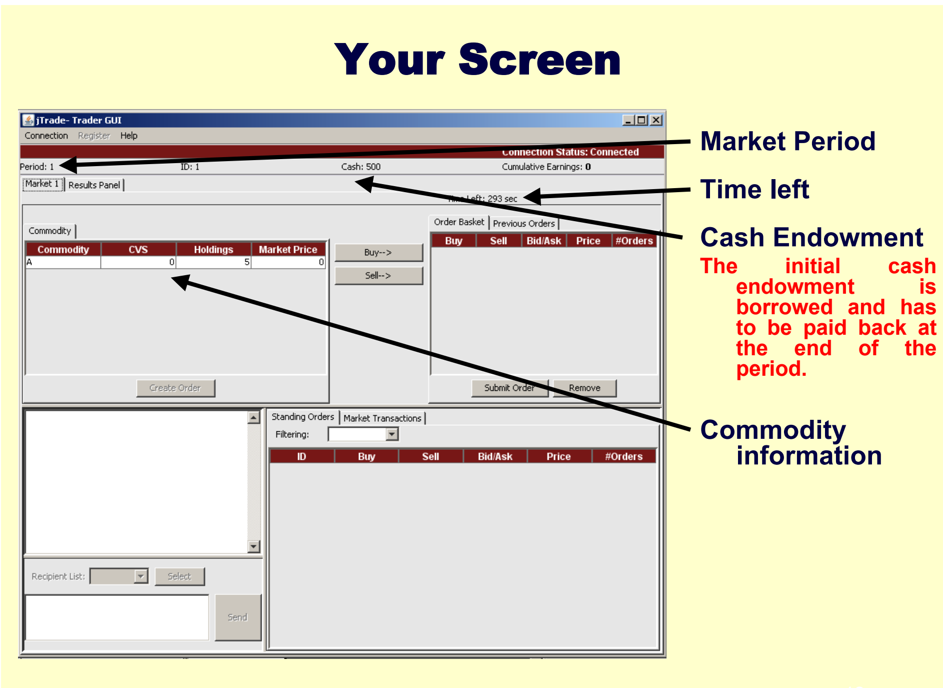


Figure A.4. Your Screen.

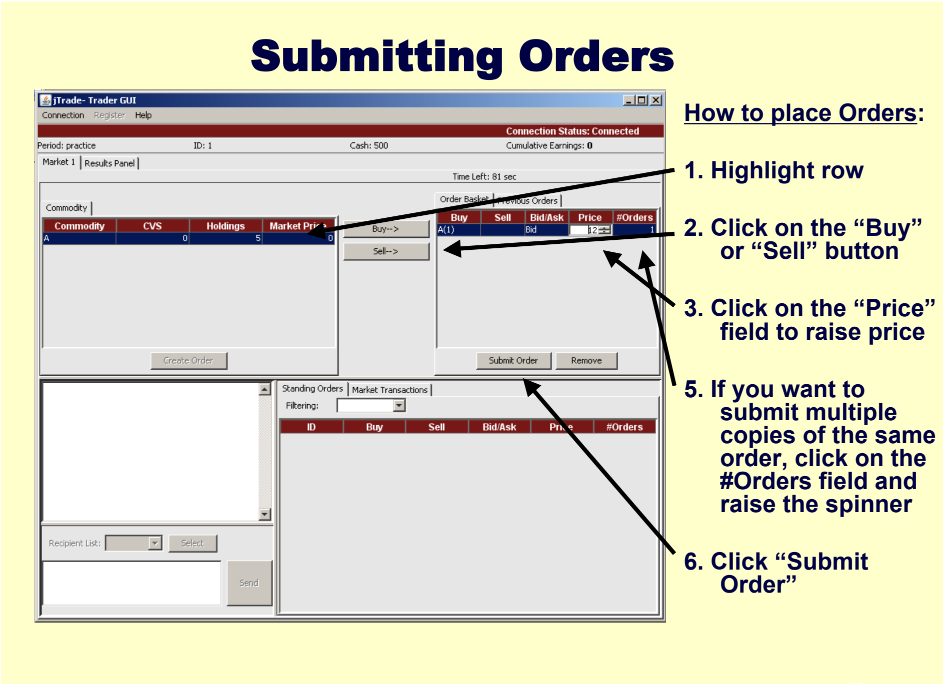


Figure A.5. Submitting Orders.

A.4 Instructions for Treatment Sequence

The instructions for treatment sequence are identical to the instructions for treatment baseline except for the description of the common value signal. The slide "common value signal I" is slightly different and there is one extra slide (common value signal III).

A.4.1 Common Value Signal I

- The common-value signal (CVS) is either 0 or 100. It can be either strong or weak.
  - If you receive a strong CVS, your signal corresponds to the actual common value with probability 0.8.
  - If you receive a weak CVS, your signal corresponds to the true common value with probability 0.6.

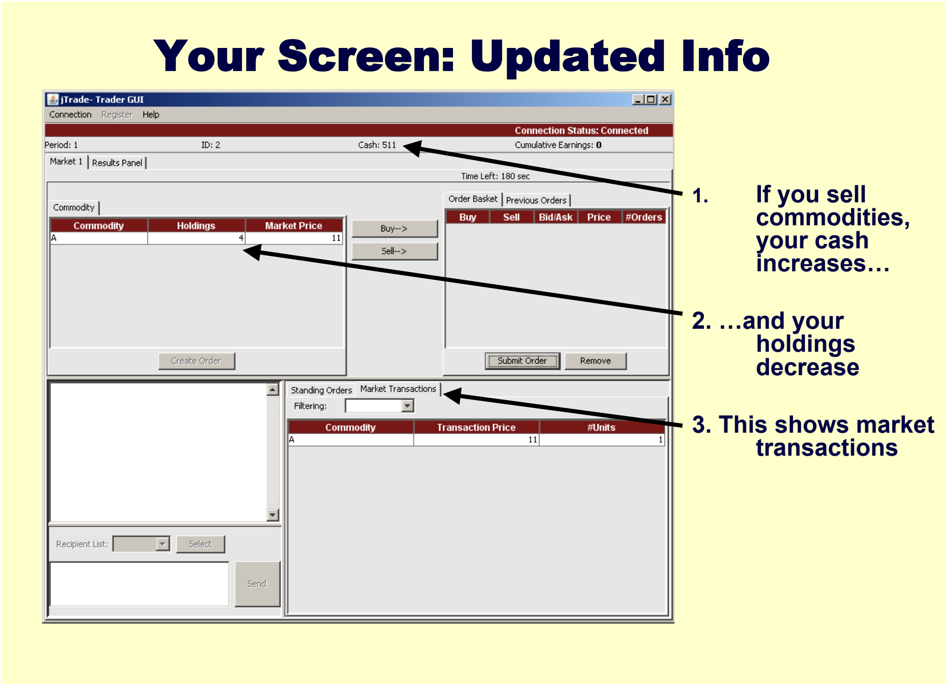


Figure A.6. Your Screen: Updated Info.

- Each CVS is equally likely to be strong or weak. You will be told whether your signal is strong or weak.

- Each trader receives exactly one common-value signal. Common-value signals are independent across traders and periods.

A.4.2 Common Value Signal III

- In each period, each trader receives exactly one common-value signal. However, different traders receive their signal at different times. Trader 1 receives his signal at the beginning of the market period. Trader 2 receives his signal 30 seconds after the beginning of the market period. Every 30 seconds, another trader receives his common-value signal. After the last trader (trader 8) received his signal, you have 2.5 minutes to trade before the market closes. Therefore, one period lasts 6 minutes.

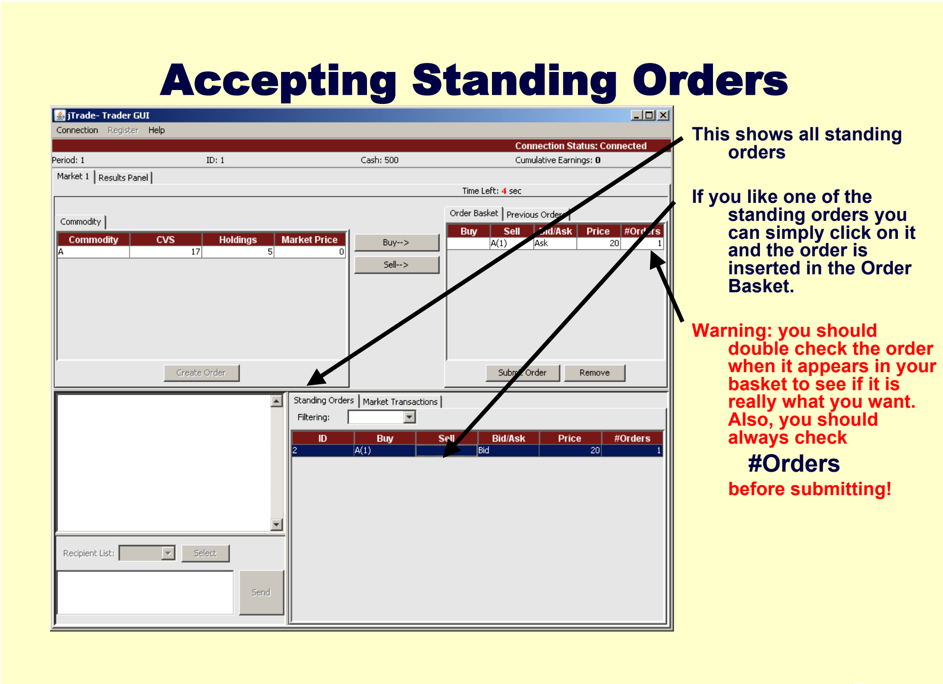


Figure A.7. Accepting Standing Orders.

- In each market period, your position is randomly determined. Therefore, the point of time at which you receive your common-value signal is likely to change from one period to the next.
- You will only see your own common value signal but not other traders' signals.
- You will be able to see the ID number of the trader for each bid/ask submitted. Therefore, you will always know whether the person who submitted a certain order has already received his common-value signal.

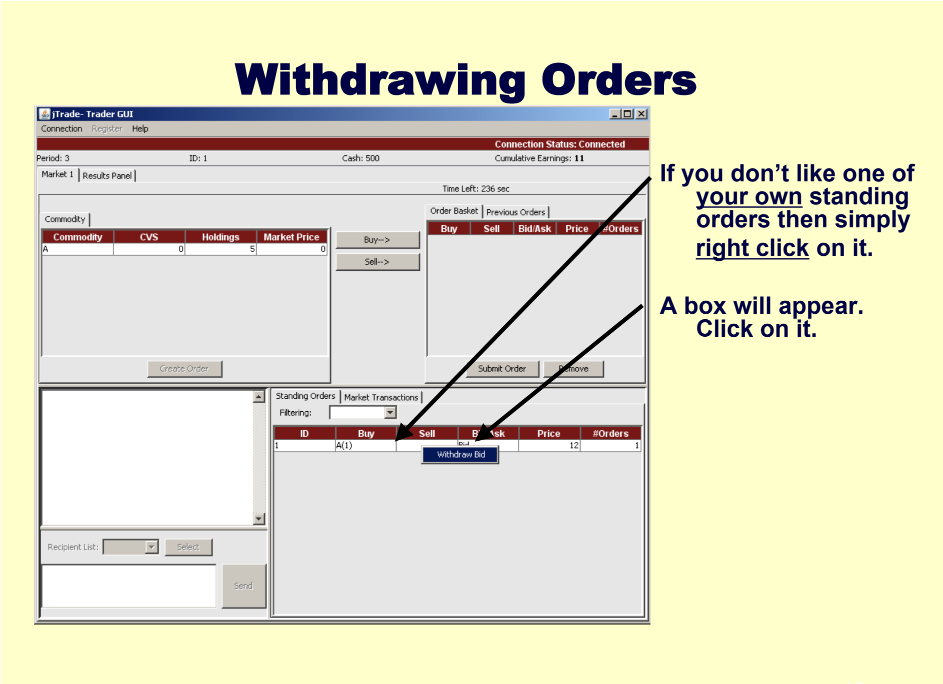


Figure A.8. Withdrawing Orders I.

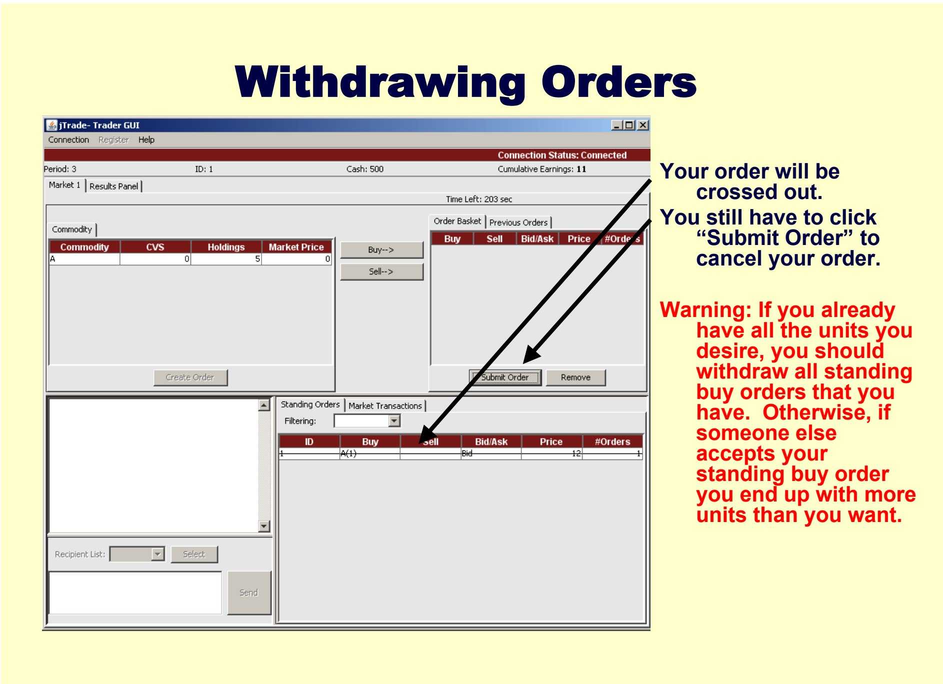


Figure A.9. Withdrawing Orders II.

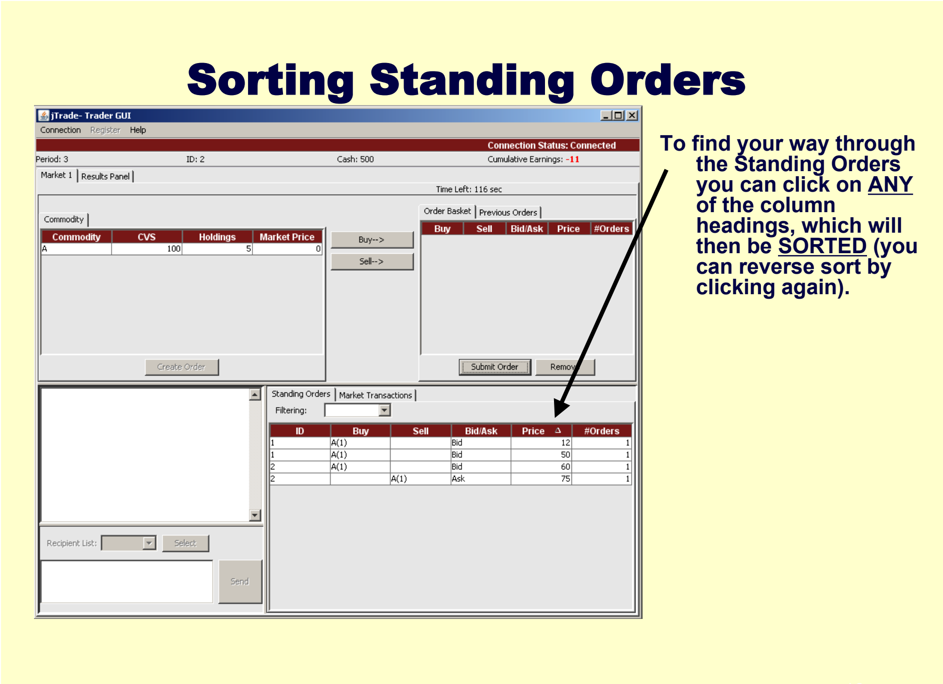


Figure A.10. Sorting Standing Orders.

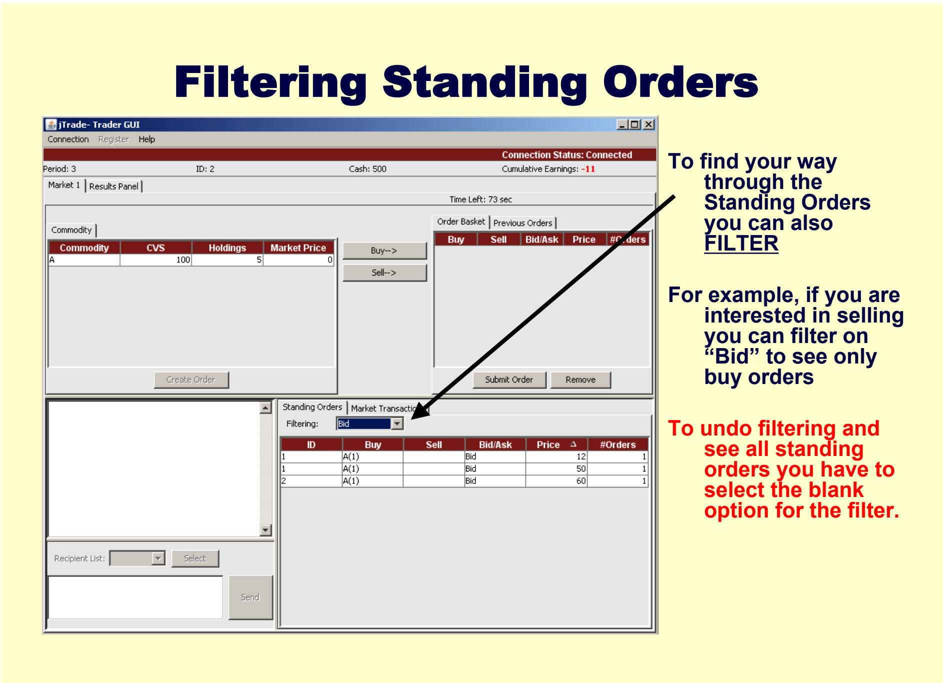


Figure A.11. Filtering Standing Orders.

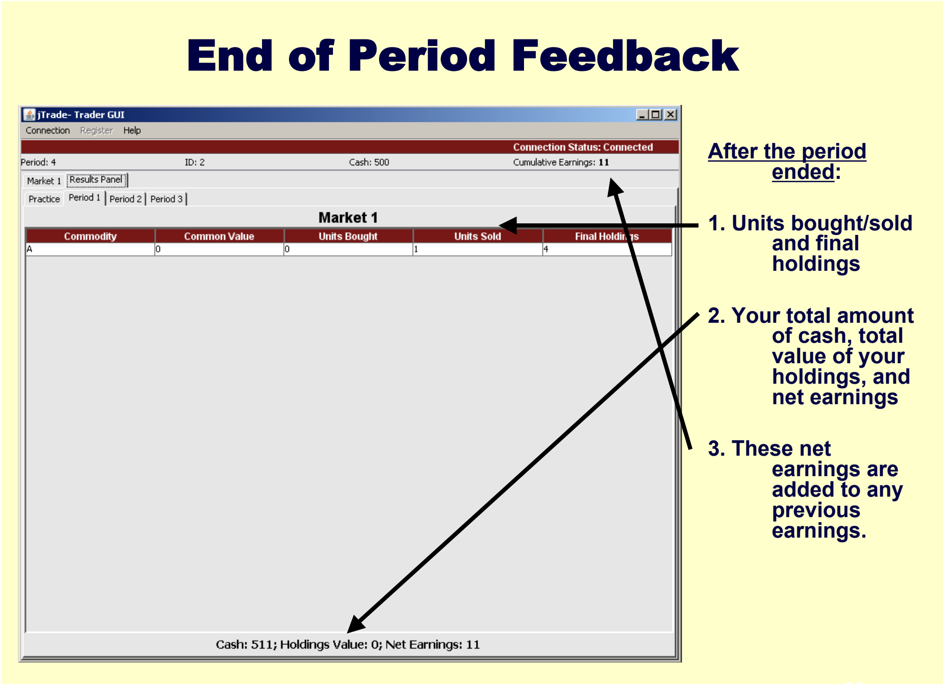


Figure A.12. End of Period Feedback.

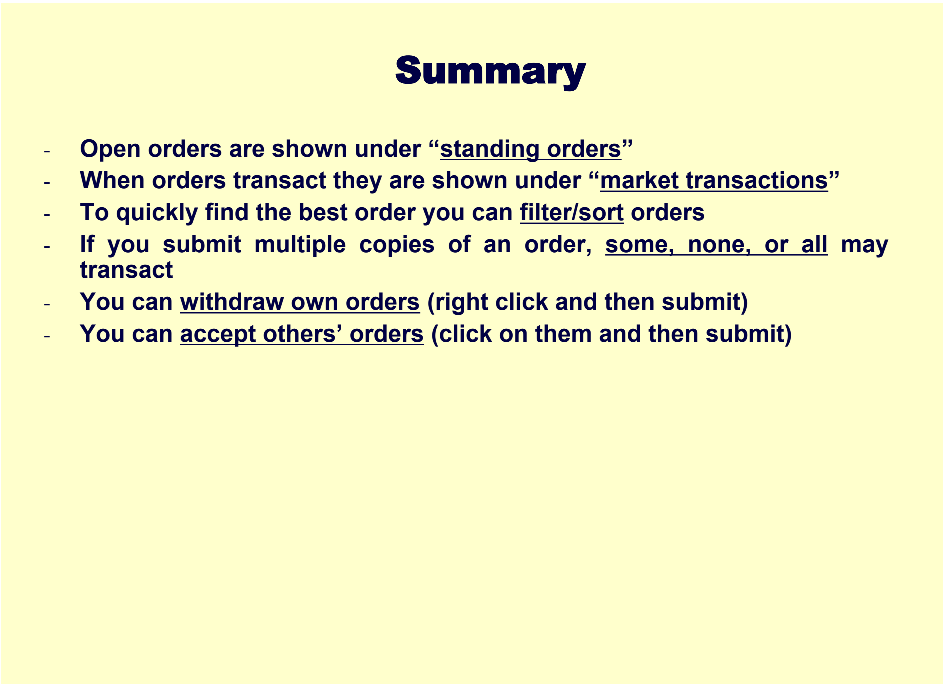


Figure A.13. Summary.

Appendix B

Appendix to Chapter 3

B.1 Instructions for Treatment Sequence

All experiments reported in chapter 3 were conducted without computers. Instead, subjects wrote down their decisions on a piece of paper. The instructions were presented in the form of a powerpoint-presentation and subjects were able to ask questions both during and after the presentation. The following is the text used in the presentation. There was a total number of 9 slides (one per subsection). The instructions for sessions with different values of q (the probability that a subject was informed) were identical.

B.1.1 Welcome to SSEL

- Welcome and thank you for participating in today's experiment.
- Place all of your personal belongings away, so we can have your complete attention.
- In today's experiment, you will not use a computer. Instead, you will write down all of your decisions on paper forms.
- You will make a total number of 10 decisions. For each correct decision, you will earn \$4. You will be paid in cash at the end of the experiment.

- Please do not talk during the experiment. If you have a question, raise your hand.

B.1.2 The Correct Option

In each one of 10 identical periods, you will choose among two options, cross (+) and circle ($\circ$). One of these two options is the "correct" option. If you choose the correct option, you receive $4, otherwise, you receive nothing in that period.

For each period, we flipped a fair coin to determine whether $\circ$ or + is the correct option. Therefore, in each period, both + and $\circ$ have an equal chance of being the correct option.

B.1.3 Sequential Guessing

You will guess what the correct option is one after another. The order in which you make your guesses is determined randomly using the following procedure: Each one of you has been assigned an ID number (at the top of your record sheet). An urn contains each one of these numbers exactly once. To decide who goes first, we draw a number from the urn. To decide who goes second, we again draw a number from the urn without replacing the number of the person who guessed first. This process is repeated until there are no numbers left.

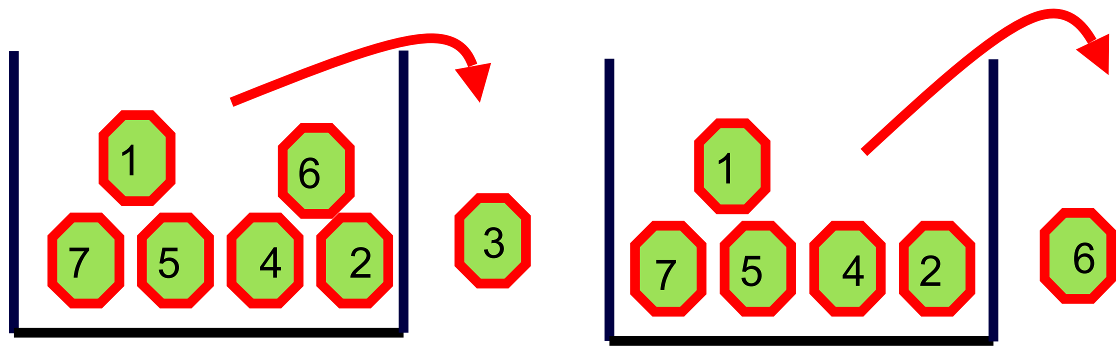


Figure B.1. Urn Representation.

B.1.4 Your Information I

When it is your turn to guess, we will randomly determine whether you are told what the correct option is or not. The probability that you are told what the correct option is is 0.2. In other words, you can imagine that we rolled a 5-sided die separately for each one of you in each period to determine whether you are told what the correct option is.

B.1.5 Your Information II

No matter whether you are told what the correct option is or not, you will always be told what your predecessors' decisions were.

For example, suppose the first decision maker chooses $\circ$, the second chooses $+$ and the third chooses $\circ$. If you are the forth decision maker, you will be shown the following table:

Period 1	1	2	3	4	5	6	7
Guess	○	+	○				

Figure B.2. Information Given to Subjects in Treatment Sequence.

B.1.6 Summary

- $+$ and $\circ$ have an equal chance of being the correct option.
- You will guess what the correct option is in a randomly determined sequence.
- With probability 0.2, you are told what the correct option is.
- You can see what your predecessors guessed.
- There are 10 periods. In each period, you receive $4 for correctly guessing what the correct option is.

- There will be no practice period. Please raise your hand if you have a question.

B.1.7 Keeping Track I

You all received an earnings sheet. Your ID number is indicated at the top. In each period,

- you record your position
- we record whether you are told what the correct option is
- and if so what it is
- you record your guess
- you record what the correct option was
- you record your earnings for the period

After the last period, please add up your earnings and also add $5 show-up fee.

B.1.8 Keeping Track II

This slide explains how subjects were supposed to keep records and consisted of figure B.3.

B.1.9 Concluding Remarks

For each correct decision, you will receive $4. In addition to that, you will receive a $5 show-up fee.

You will be paid at the end of the experiment the total amount you have earned. You need not tell any other participant how much you earned.

To ensure your privacy and that of others in the experiment, please pull out the dividers as far as they will go.

ID: 7					
Period	Position	Correct Option	Guess	Correct Option	Earnings
1		_			
2		-			
3		-			
4		-			
5		-			
6		-			
7		-			
8		-			
9		+			
10		-			
Show-up Fee					$5.00
Total					

We record
- if you don't know what the correct option is
+ if the correct option is +
○ if the correct option is ○

Figure B.3. Record Sheet.

B.2 Instructions for Treatment No-Sequence

The instructions for treatment no-sequence were identical to the instructions for treatment sequence with the exception of slides 5 (Your Information II) and 6 (Summary). Therefore, we only reproduce these two slides.

B.2.1 Your information II

No matter whether you are told what the correct option is or not, you will always be told how many of your predecessors chose + and how many of your predecessors chose ∘.

For example, suppose the first participant chooses ∘, the second chooses + and the third chooses $\circ$. If you are the fourth participant to guess the correct option, you will be told that 2 of your predecessors chose $\circ$ and 1 of your predecessors chose $+$.

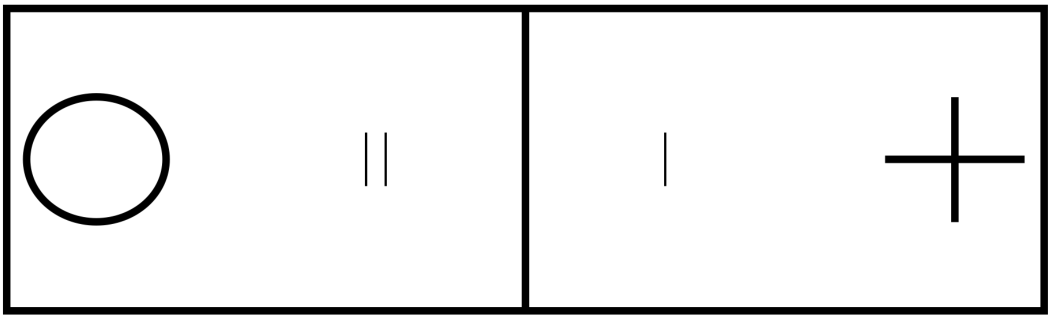


Figure B.4. Information Given to Subjects in Treatment No-Sequence.

B.2.2 Summary

- $+$ and $\circ$ have an equal chance of being the correct option.
- You will guess what the correct option is in a randomly determined sequence.
- With probability 0.2, you are told what the correct option is.
- You can see how many of your predecessors chose $\circ$ and how many of your predecessors chose $+$.
- There are 10 periods. In each period, you receive $4 for correctly guessing what the correct option is.
- There will be no practice period. Please raise your hand if you have a question.

Appendix C

Appendix to Chapter 4

C.1 Rules for the Simultaneous Multi-Round (SMR) Auctions

Rounds and bid structure. All licenses are put up for bid simultaneously, and participants may only submit bids on individual licenses. The auction consists of successive rounds in which participants may place bids. Following each round, the high bid for each license is posted. These high bids then become the standing bids for the subsequent round.

Acceptable bids. In the first round, an acceptable bid must be equal to or exceed the initial price of 0 by 5 points (each point equaled 40 cents in the experiment). Subsequently, in order to be acceptable, a bid must exceed the provisionally winning bid for the license by at least 5 points. Bidders are given the choice of making one of eight incrementally higher bids (in 5-point increments).

Bid withdrawal. Each bidder has at most 2 rounds in which they are permitted to withdraw any of their provisionally winning bids. After the withdrawal, the seller becomes the provisionally winning bidder for the withdrawn license and the minimum acceptable bid in the following round equals the second highest bid received on the license, which may be less than or equal to (in the case of tied bids) the amount of the withdrawn bid. A withdrawing bidder pays a penalty equal to the maximum of zero or the difference between the price at which the bidder withdrew its bid and the final sale price in the current auction. If the license goes unsold, the bidder would normally be responsible for paying the difference between the withdrawn bid and the sale price in a subsequent auction, plus a small percentage penalty of 3%. In the experiment, there is no subsequent auction, so these penalties for the case of an unsold license were implemented by requiring that the bidder pay a penalty of 25% of the withdrawn bid.

Bidding eligibility and activity. Each license in the experiment is assigned one bidding unit. The total number of bidding units available to the bidder establishes the bidder's maximum eligibility to bid. National bidders begin each auction with 6 activity units and regional bidders begin with 3. In each round, a bidder's activity is calculated as the number of licenses for which that bidder is a provisional winner, plus the number of licenses for which acceptable bids are submitted. If a bidder's activity falls below the bidder's current activity limit, that limit is reduced to equal the bidder's actual activity. There were no activity rule wavers in the experiment, so a reduction in activity would put an upper limit on the bidder's activity for all subsequent rounds of that auction.

End of round feedback. At the end of each round, bidders receive information on all provisionally winning bids, withdrawn bids, and the corresponding bidder ID numbers. Bidders also see the sum of their own values for the licenses that they are provisionally winning and prices that would be paid for the licenses if the auction had ended.

Closing rule. The auction closes after any round in which no new bids were placed and no bids were withdrawn. In this case provisionally winning bids become winning bids that are used to calculate auction earnings. The experiment did not allow for defaults on payments, so gains were added to cumulative earnings and losses were subtracted.

C.2 Rules for Resource Allocation Design (RAD) Pricing (in Bold) and the Simultaneous Multi-Round Auction with Package Bidding (SM-RPB) (in Italics)

Rounds and bid structure. This is a simultaneous, multi-round auction in which participants may submit one or more bids on individual licenses or on combinations of licenses (packages). Provisionally winning bids are calculated by maximizing seller revenue for the round, using all current and past bids. *(SMRPB: Bids have an exclusive XOR structure in the sense that each bidder can have at most one provisionally winning bid.)*

Acceptable bids. In the first round, an acceptable bid must be equal to or exceed the minimum opening bid of 0 by 5 points for each license, or by 5 points times the number of licenses in a package. After each subsequent round, prices are calculated for each license on the basis of bids received in the previous round. The pricing rule, as specified in Appendix D of the Experiment Design Report, calculates prices that reflect (as closely as possible) the marginal sales revenue of each license based on bids received. Prices for packages are the given by the sum of the prices for each license in the package. In order to be acceptable, a bid must exceed the price of a license or package at least 5 points times the number of licenses covered by the bid. Bidders are given the choice of making one of eight incrementally higher bids (in 5-point increments).

Bidding eligibility and activity. Each license in the experiment is assigned one bidding unit. The total number of bidding units available to the bidder establishes the bidder's maximum eligibility to bid (3 for regional bidders and 6 for national bidders). **(RAD: In each round, a bidder's activity is calculated as the number of different licenses for which that bidder is a provisional winner or for which that bidder places a bid, either singly or as part of a package.)** *(SMRPB: In each round, a bidder's activity is calculated as the maximum of (1) the* *size of the largest package the bidder is provisionally winning and (2) the size of the largest package the bidder is bidding for.)* If a bidder's activity falls below the bidder's current activity limit, that limit is reduced to equal the bidder's actual activity. There were no activity rule wavers in the experiment, so a reduction in activity would put an upper limit on the bidder's activity for all subsequent rounds of that auction.

End of round feedback. At the end of each round, bidders receive information on all provisionally winning bids (for licenses and packages) and the corresponding bidder ID numbers. Bidders also see the prices for all licenses, the sum of their own values for the licenses and packages that they are provisionally winning, and the sum of prices that would be paid for those licenses and packages if the auction had ended.

Closing rule. The auction closes after any round in which no new bids were placed. In this case provisionally winning bids become winning bids that are used to calculate auction earnings. The experiment did not allow for defaults on payments, so gains were added to cumulative earnings and losses were subtracted.

C.3 Rules for the Combinatorial Clock (CC)

Rounds and bid structure. This is a simultaneous, multi-round auction in which participants may submit one or more bids on individual licenses or on combinations of licenses (packages). Submitted bids stay active until they are removed (just as provisionally winning bids in the other formats are automatically renewed).

Acceptable bids. In the first round, an acceptable bid must be equal to or exceed the minimum opening bid of 0 by 5 points for each license, or by 5 points times the number of licenses in a package. After each subsequent round, prices are calculated for each license on the basis of bids received in the previous round. If a license is in the contained in a bid made by more than one bidder (individually or as part of a package), then the price for that license will rise by the bid increment (5), otherwise the price does not change. For example, if one bidder is bidding on A and AB, and if the only other bidder is bidding on B, then the price of B will increase and the price of A will stay unchanged. Prices for packages are the given by the sum of the prices for each license in the package, so the price of a package can increase by at most the product of the bid increment and the number of licenses in the package.

Bidding eligibility and activity. Each license in the experiment is assigned one bidding unit. The total number of bidding units available to the bidder establishes the bidder's maximum eligibility to bid (3 for regional bidders and 6 for national bidders). In each round, a bidder's activity is calculated as the number of different licenses for which that bidder places a bid, either singly or as part of a package. For example, a regional bidder with an initial activity limit of 3 would be able to bid on packages BC and on ABC, but not on ABC and E. If a bidder's activity falls below the bidder's current activity limit, that limit is reduced to equal the bidder's actual activity. There were no activity rule wavers in the experiment, so a reduction in activity would put an upper limit on the bidder's activity for all subsequent rounds of that auction.

End of round feedback. At the end of each round, bidders receive information on all prices and submitted bids (for licenses and packages), with the corresponding bidder ID numbers.

Closing rule. The auction generally closes after any round in which there is no excess demand, i.e., no license is in the bidding basket of more than one person. However, if there is excess supply at this point (one or more unclaimed licenses), then a revenue maximization routine is run using all submitted bids in all rounds in order to arrange the sale of all licenses. If the resulting allocation displaces the sole remaining bidder for any of the licenses, the auction is restarted and the clock prices on those licenses are raised to let those bidders have the chance to reclaim them.

C.4 Experimental Instructions for the Simultaneous Multi-Round Auction with Package Bidding (SMRPB)

The instructions were presented in the form of a powerpoint-presentation in all sessions. The following figures (figure – figure) correspond to the slides used for the simultaneous multi-round auction with package bidding (SMRPB). We are only displaying the full set of slides for this particular auction since we used the same instructions with appropriate simplifications for all other auction formats. All other auction formats used a different activity rule, which was described in the slides corresponding to figures C.35 and C.36. Also, the combinatorial clock auction used a substantially different pricing rule, which is described in the slides corresponding to figures C.37 and C.38.

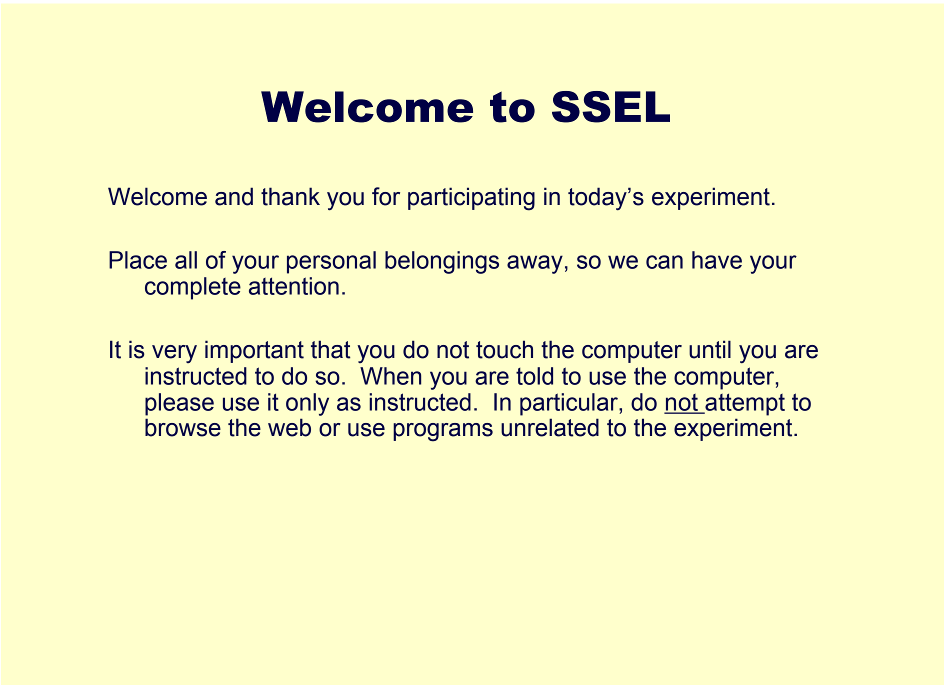


Figure C.1. Slide 1.

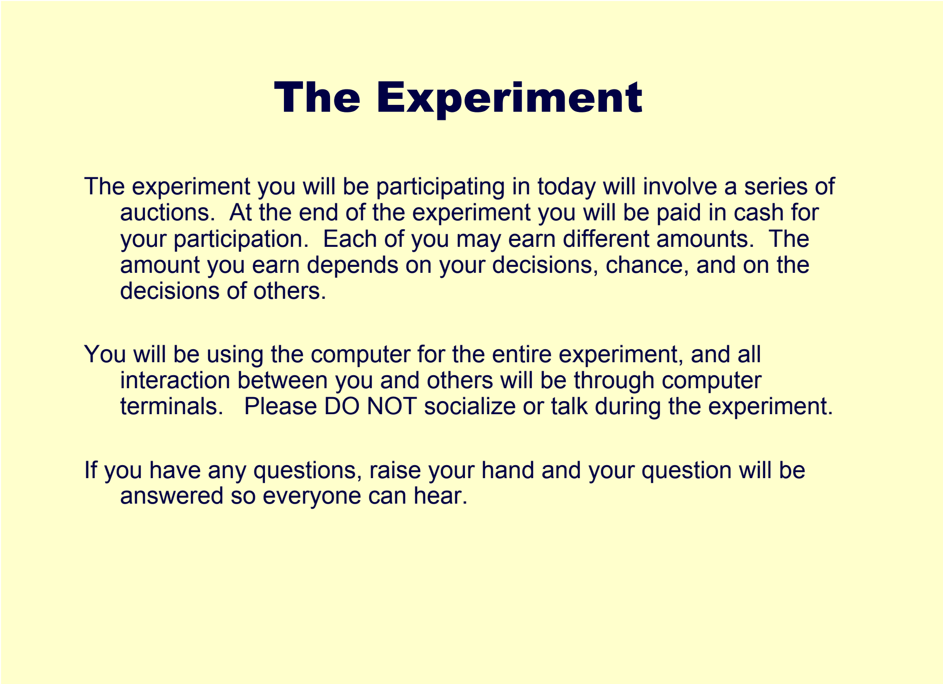


Figure C.2. Slide 2.

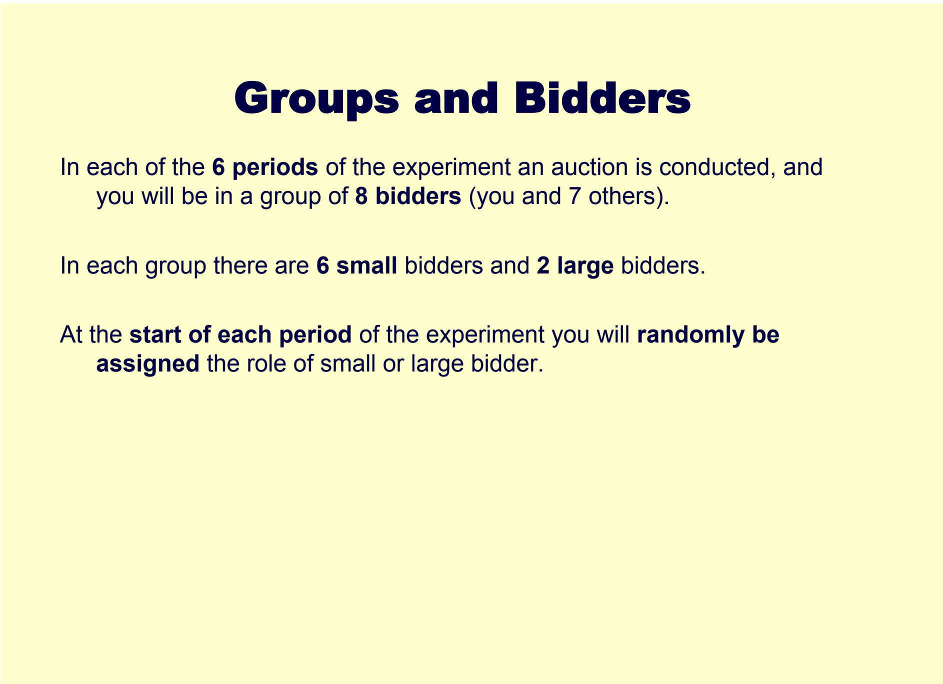


Figure C.3. Slide 3.

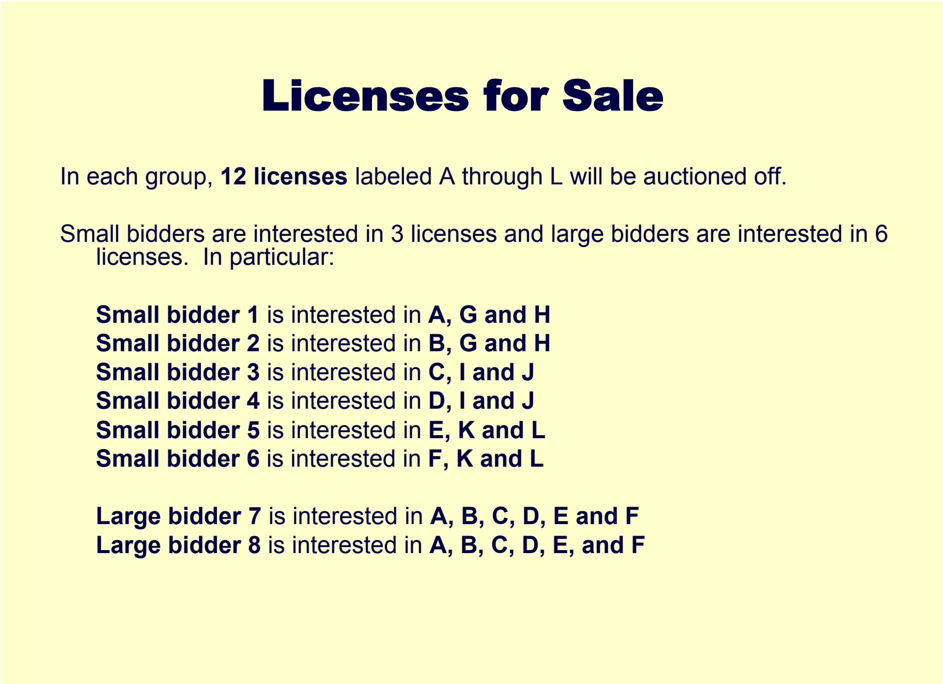


Figure C.4. Slide 4.

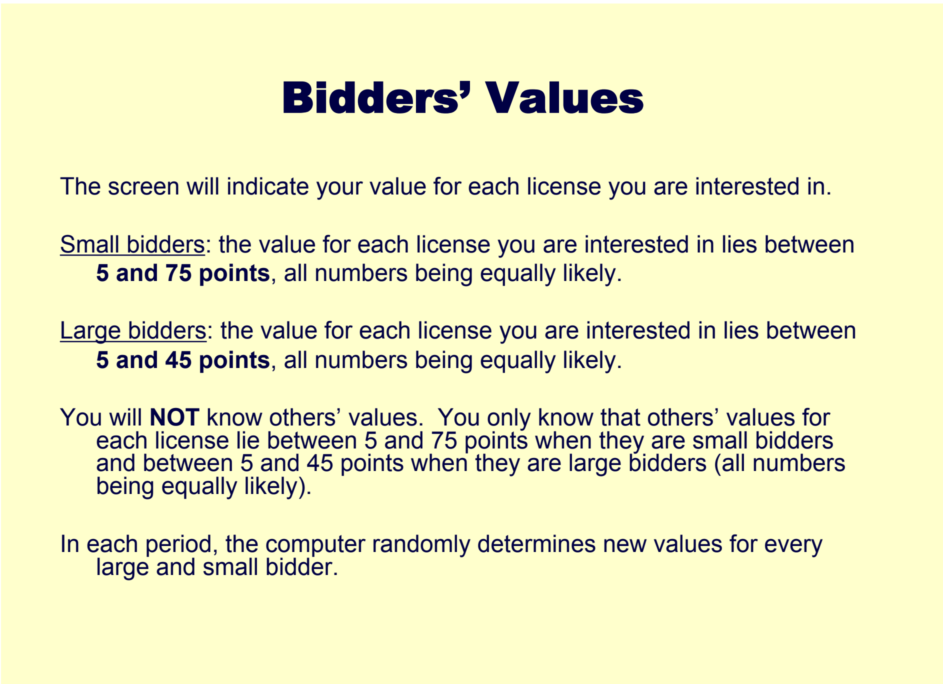


Figure C.5. Slide 5.

Small Bidders' Values

If you win a single license then the **amount of points** you receive is your value for the license.

If you win:

2 licenses then the value of every license won goes up by **12.5%**
3 licenses then the value of every license won goes up by **25%**

Examples: values are 10 for A, 40 for G, and 70 for H.
The value of winning both A and H is 1.125*(10+70) = 90.
The value of winning A and G and H is 1.25*(10+40+70) = 150.

Figure C.6. Slide 6.

Large Bidders' Values

If you win a single license then the **amount of points** you receive is your value for the license.

If you win:

2 licenses then the value of every license won goes up by 20%
3 licenses then the value of every license won goes up by 40%
4 licenses then the value of every license won goes up by 60%
5 licenses then the value of every license won goes up by 80%
6 licenses then the value of every license won goes up by 100%

Examples: values are 5 for A, 10 for B, 15 for C, 20 for D, 25 for E and 30 for F.
The value of winning both A and B is 1.2*(5+10) = 18.
The value of winning A, B, C, and D is 1.6*(5+10+15+20) = 80.
The value of winning A, B, C, D, E and F is 2*(5+10+15+20+25+30) = 210.

Figure C.7. Slide 7.

Packages and Values

A **package** is a combination of licenses, e.g. the package JK contains licenses J and K, the package ABCDEF contains licenses A through F etc.

A package is worth more than the sum of its parts because each license goes up in value when more than one license is won.

At the start of every auction period, you have **60 seconds** to create packages (as explained next) and see how much they are worth.

It is very important that you use these 60 seconds to create packages! Bidding on a package reduces the risk of bidding high on two or more licenses and only winning one of them. **If your bid for a package wins, you get all licenses in that package**. If a package bid for ABC does not win, you get neither A nor B nor C. The only way that you can obtain A or B or C separately is if you have winning bids for A or for B or for C separately.

Figure C.8. Slide 8.

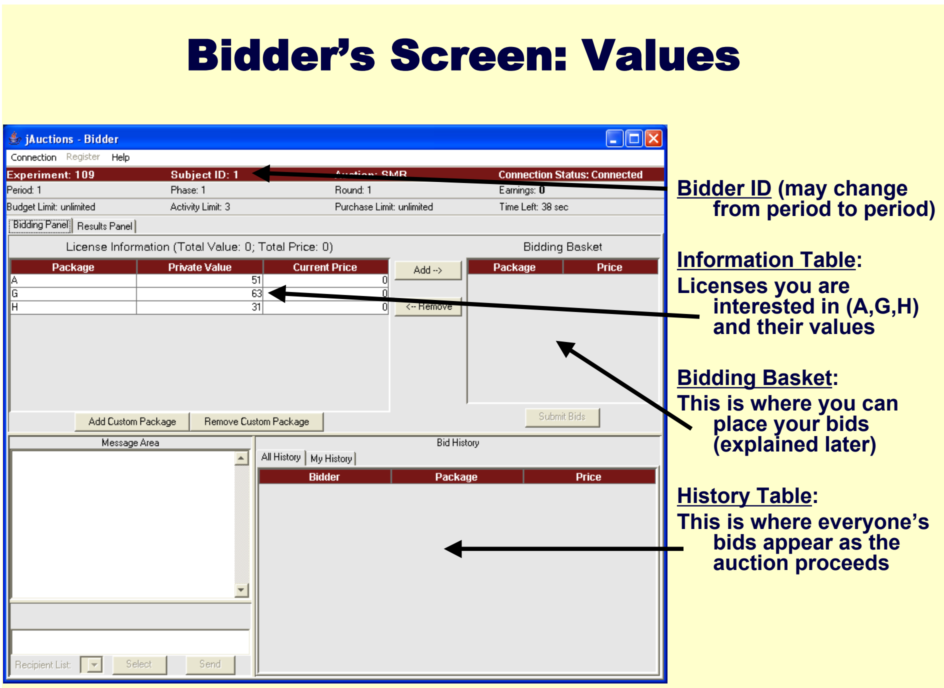


Figure C.9. Slide 9.

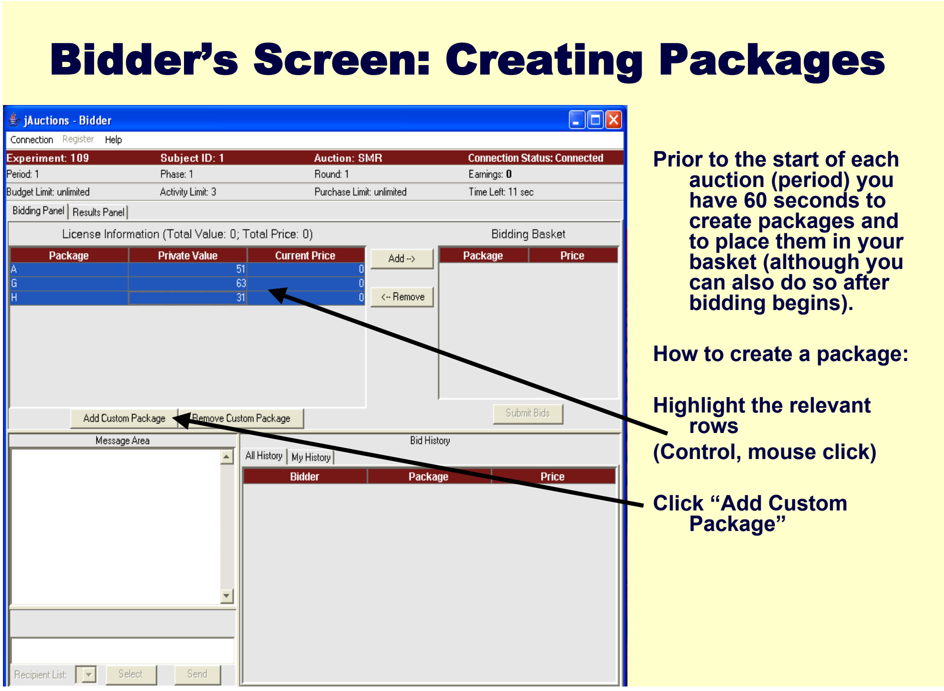


Figure C.10. Slide 10.

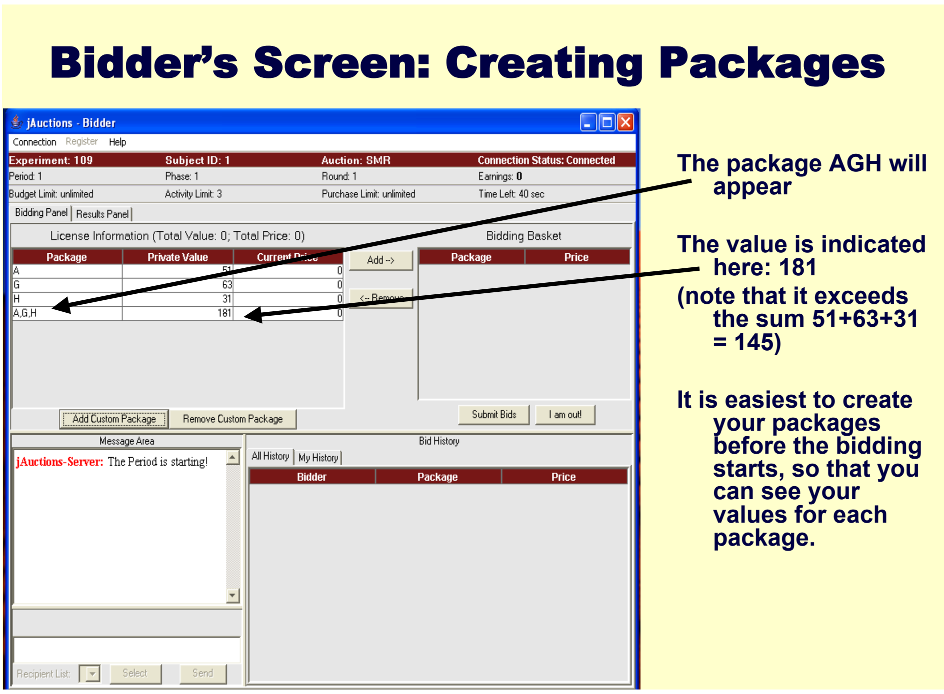


Figure C.11. Slide 11.

Periods and Rounds

In each of the **6 periods** of the experiment an auction is conducted.

Each auction consists of **multiple rounds:** each round consists of **at most 60 seconds** in which bidders can submit bids on any license they are interested in. **You can only submit bids once per round.**

There will be a **new round** of the auction as long as one or more bidders in your group submit a new bid. You can speed things up by submitting your bids early in the round. **If bidders submit late or forget to submit a basket then the entire group may have to wait the full 60 seconds.**

If prices are such that you are no longer interested in bidding you can submit an empty basket at the start of the round or you can click the **"I am out"** button. Please keep in mind that when you click the "I am out" button, you will no longer be able to bid in that period, that is, you will not be able to bid in any of the remaining rounds of the auction.

Figure C.12. Slide 12.

Bidder’s Screen: Periods and Rounds

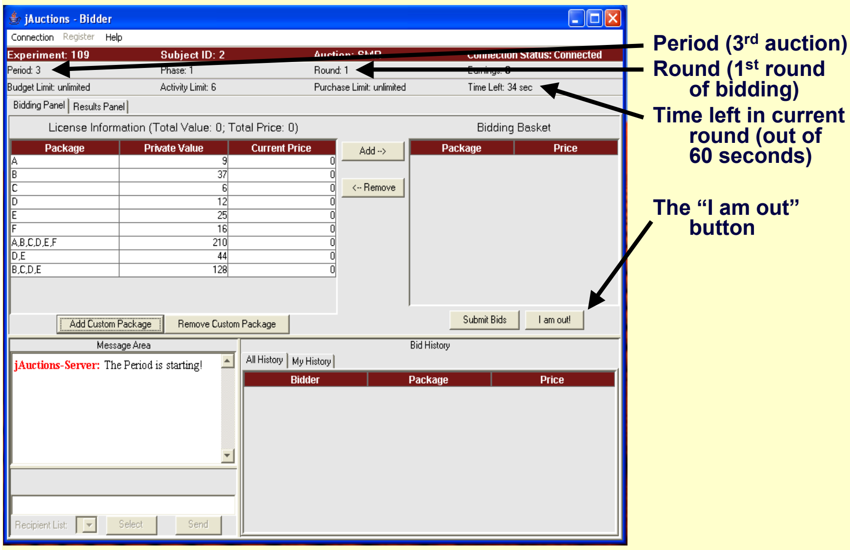


Figure C.13. Slide 13.

Bids and Bid Increments

New bids have to improve on the current price for a license by a minimum of **5 points**.

New bids have to improve on the current price for a package by a minimum of **5 points times the number of licenses in the package**.

The software automatically adds a bid increment of 5 points to the current price when you move a license to your bidding basket, and it adds 5 points times the number of licenses when a package is moved to the basket.

Once the license or package is in your basket, you can further raise your bid by clicking on its price.

Figure C.14. Slide 14.

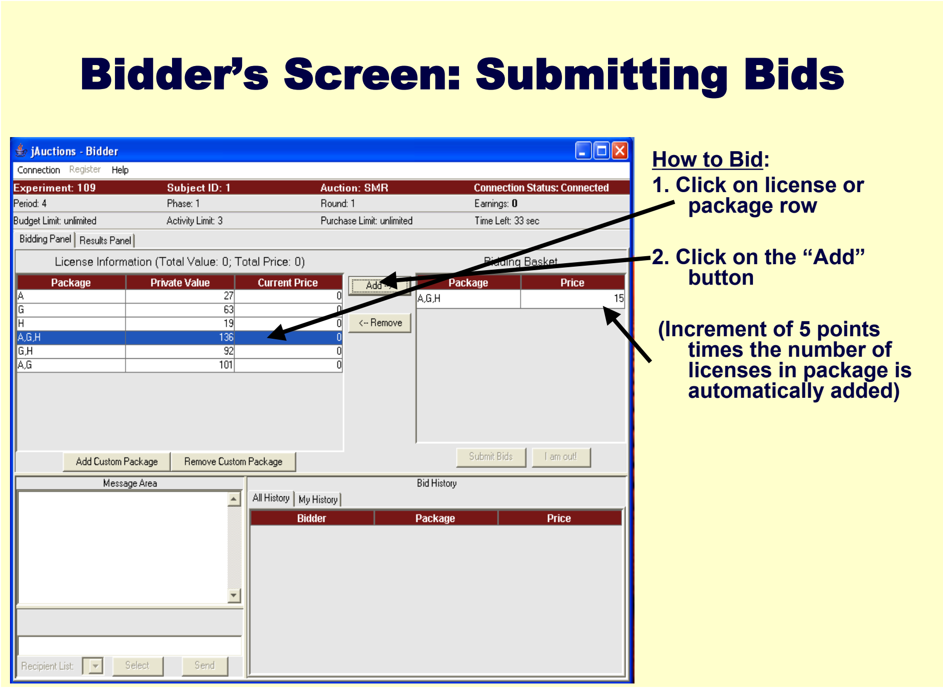


Figure C.15. Slide 15.

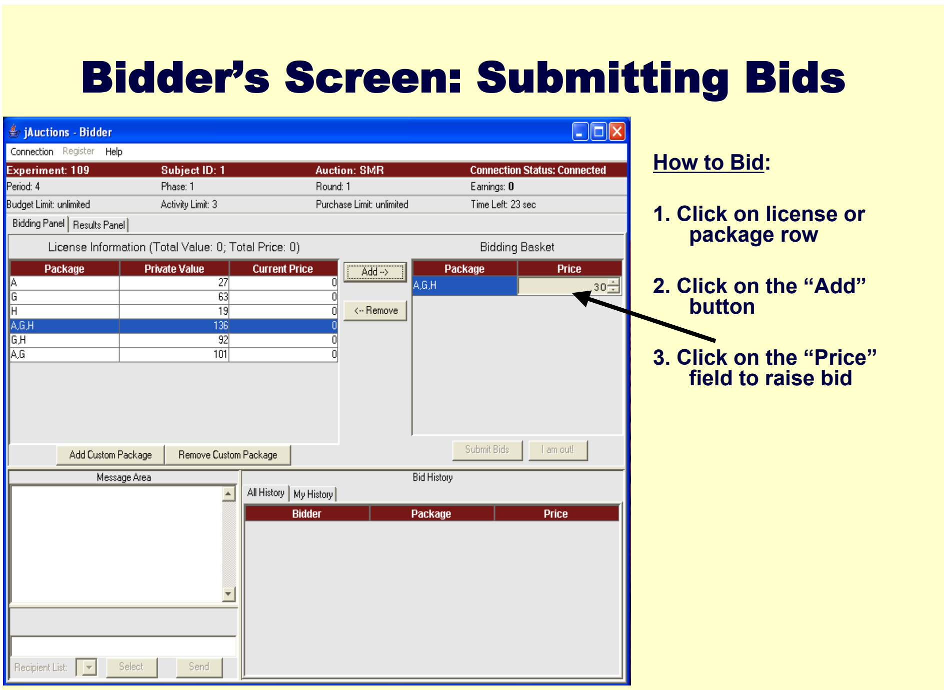
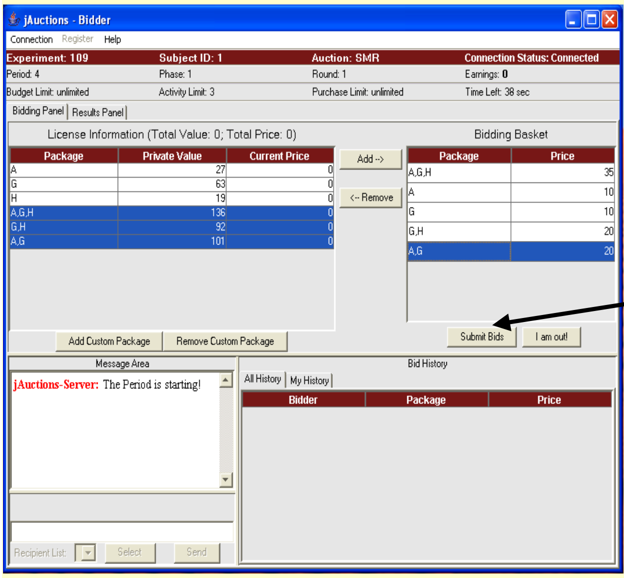


Figure C.16. Slide 16.

Bidder's Screen: Submitting Bids



<u>How to Bid</u>:

1. Click on license's row
2. Click on the "Add" button
3. Click on the "Price" field to raise bid
4. Once all desired bids are in the basket click "Submit"

Please Note: Since you only submit bids once per round, be sure that you have placed all desired bids in the basket before pressing "Submit".

Figure C.17. Slide 17.

Provisional Winners

At the end of each round (after bids are submitted or time runs out), the computer calculates the combination of winning bids that would **maximize the sales revenue** for all licenses combined (more details given later). Ties are broken by a random choice between the tied bidders.

The winning bids are announced as **"provisional winners"** for the round.

In the final round (when no bids are submitted), the provisional winning bids become the final sales prices for each license. Only these final bids are used to calculate earnings for the auction (period).

Figure C.18. Slide 18.

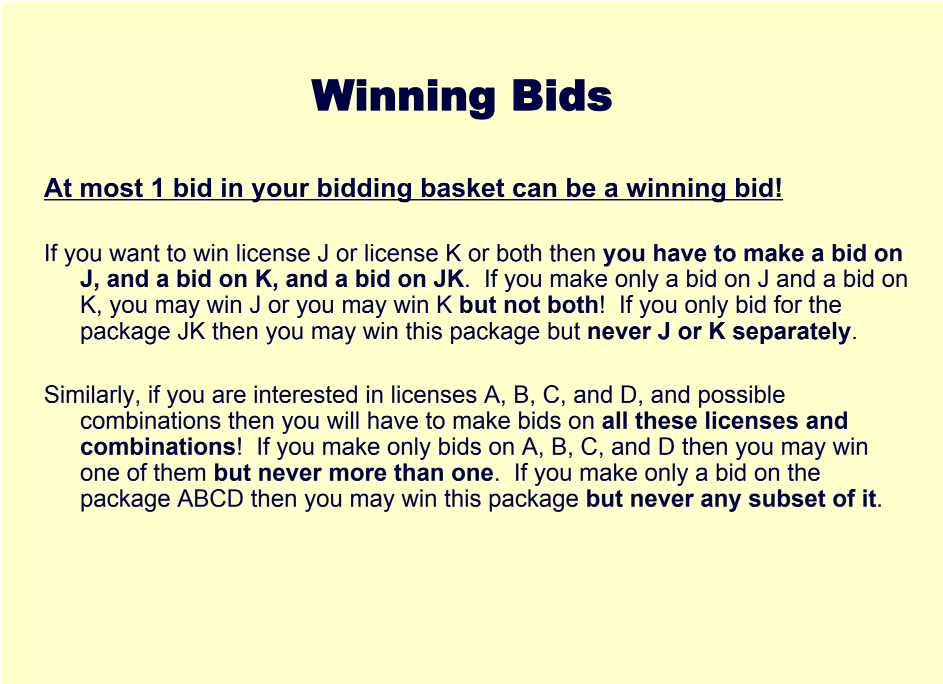


Figure C.19. Slide 19.

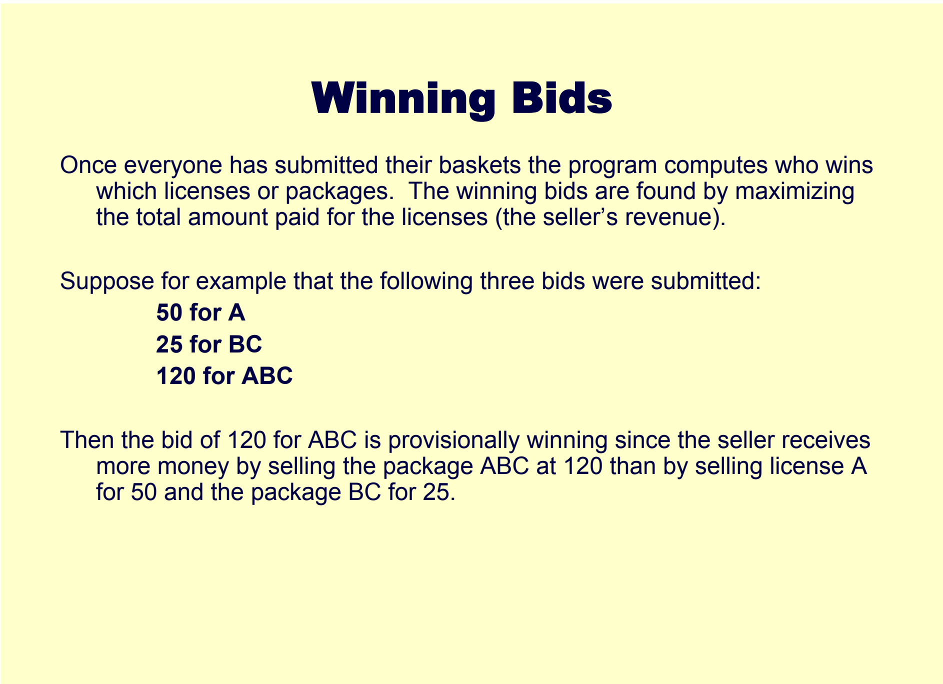


Figure C.20. Slide 20.

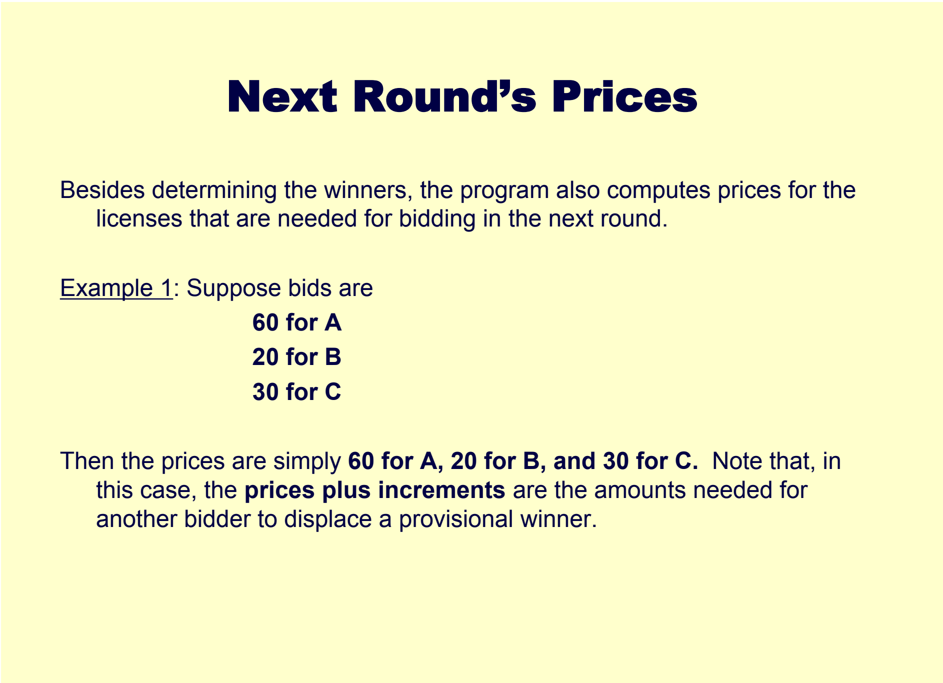


Figure C.21. Slide 21.

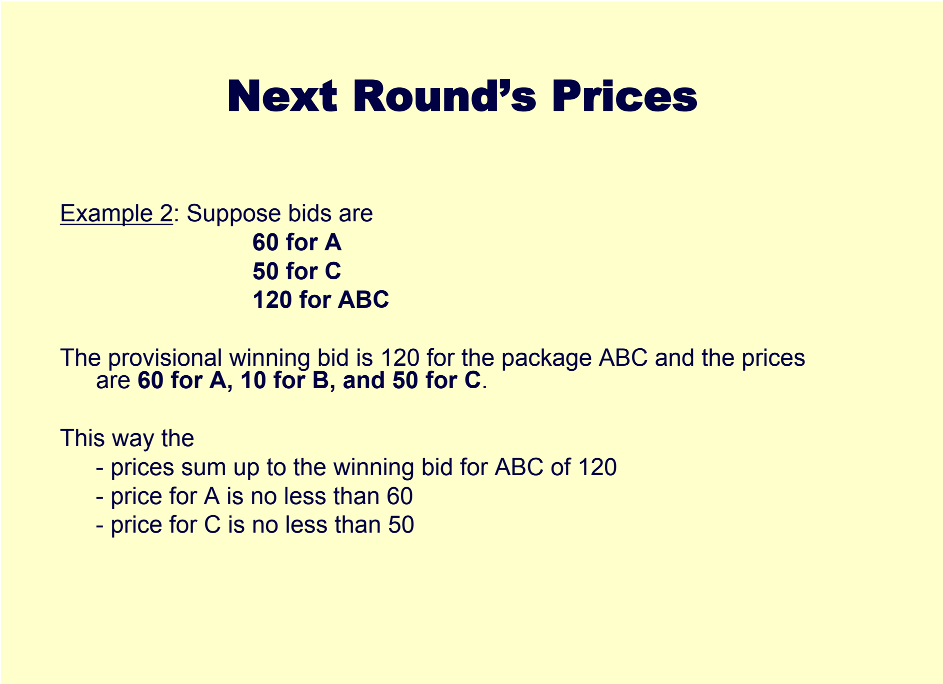


Figure C.22. Slide 22.

Next Round's Prices

Example 3: Suppose bids are

100 for AB
100 for BC
100 for AC
120 for ABC

Again, the bid of 120 for ABC would be provisionally winning since there is no way to allocate licenses in a manner that generates a sales revenue above 120. In this case, the computer would assign prices of **40 for A, 40 for B, and 40 for C**.

Now it is impossible for the three prices to sum up to 120 and for any pair of prices to sum up to at least 100. In this case, the program computes prices that come as close as possible.

So your bid may exceed the minimum price and yet not be winning, and in this sense, prices may only provide a rough idea of how high one must bid to "get into the action" for a particular license or package.

Figure C.23. Slide 23.

Next Round's Prices

Example 3 (continued): Suppose bids are

100 for AB
100 for BC
100 for AC
120 for ABC

The computer assigns prices of **40 for A, 40 for B, and 40 for C**.

Note that in this case the minimum bids for the packages AB, BC, and AC would be their current prices plus 2 times the bid increment = 40 + 40 + 5 + 5 = 90. These minimum bids are lower **than** the bids of 100 that were submitted.

To keep prices moving upward the software will **increase the bid increment on each license** (from 5 to 10 to 15 to …) after a round in which the new incoming bids exceed the minimum bids **but do not displace the current winners**.

So when adding licenses or packages to your basket **you may sometimes notice that the bid increment is higher than 5.**

Figure C.24. Slide 24.

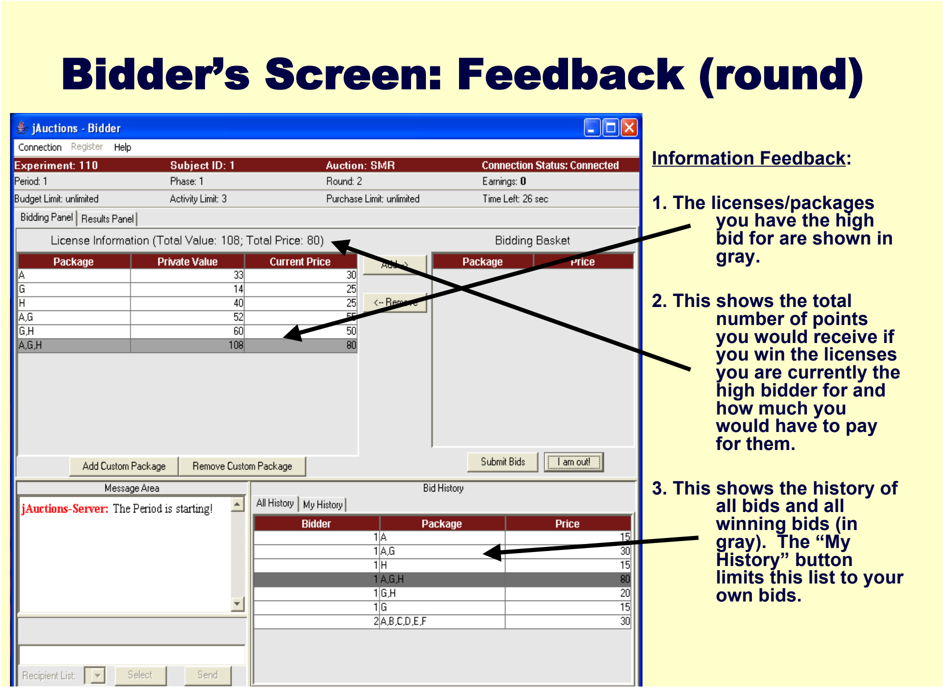


Figure C.25. Slide 25.

Activity: "Use It or Lose it"

At the start each period bidders get an **activity limit equal to the number of licenses they are interested in.**

Bidders' activity limits in a round indicate the **maximum size of the packages they can bid for**. For example, if the activity limit is 3 a bidder can bid on ABC but not on ABCD.

Your activity limit for the next round is equal to:
the number of licenses you are currently the provisional winner for
+ the number of licenses in the largest package you are bidding for but not provisionally winning

So to keep your activity you do **NOT** have to bid on the licenses you are already provisionally winning.

If the number of licenses you bid on plus the number of licenses you are currently the provisional winner for is less than your current activity then **your activity limit will drop in the next round and subsequent rounds**.

Figure C.26. Slide 26.

Activity: "Use It or Lose it"

Example:

a bidder currently has an activity limit of 4 units
the bidder is provisionally winning license A
the bidder is bidding on the package BC, and on the license D

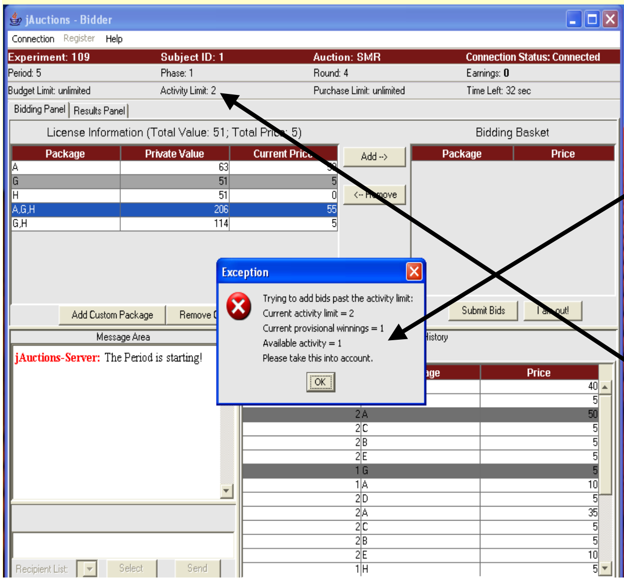
In the next round the bidder will have:

+1 activity unit for provisionally winning license A
+2 activity units for bidding on the package BC
= total of 3 activity units

This means that in the next round the bidder's activity limit will drop to 3 meaning that the bidder can bid at most for 3 licenses

Figure C.27. Slide 27.

Activity: "Use It or Lose it"

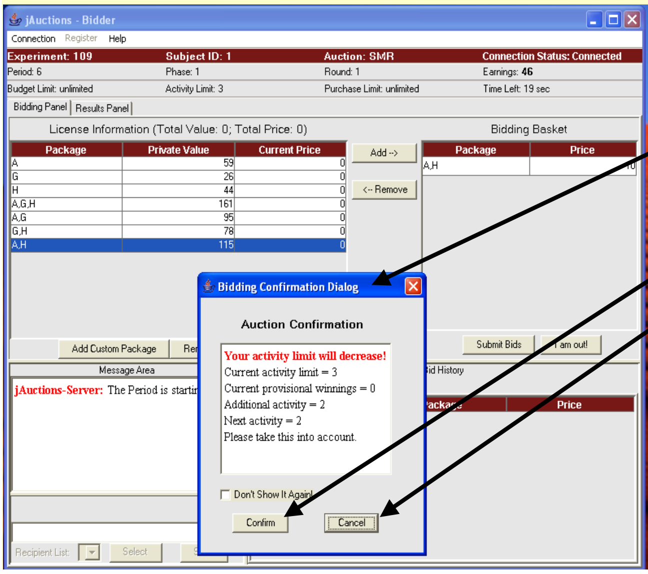


When you try to bid on more licenses than your current activity limit allows for, you will get a warning message.

Here the current activity limit is 2, and the bidder is already provisionally winning 1 license, so can bid only for 1 other license.

Figure C.28. Slide 28.

Activity: "Use It or Lose it"



When you submit a basket that will lower your current activity limit you will get a warning message. You have to click "confirm" to submit your bids or "cancel".

You can disable this option if you prefer.

Your activity limit will always be shown at the top.

Figure C.29. Slide 29.

Your Earnings

Once no more bids are submitted (either all bidders submit empty baskets or the clock ticks down with no bids made)**, the auction closes** and your earnings for the period are determined. Your earnings equal the total value of the licenses or packages you won minus the total bids you made for them.

earnings = sum of values for licenses/packages you won
– sum of your bids for licenses/packages you won

Figure C.30. Slide 30.

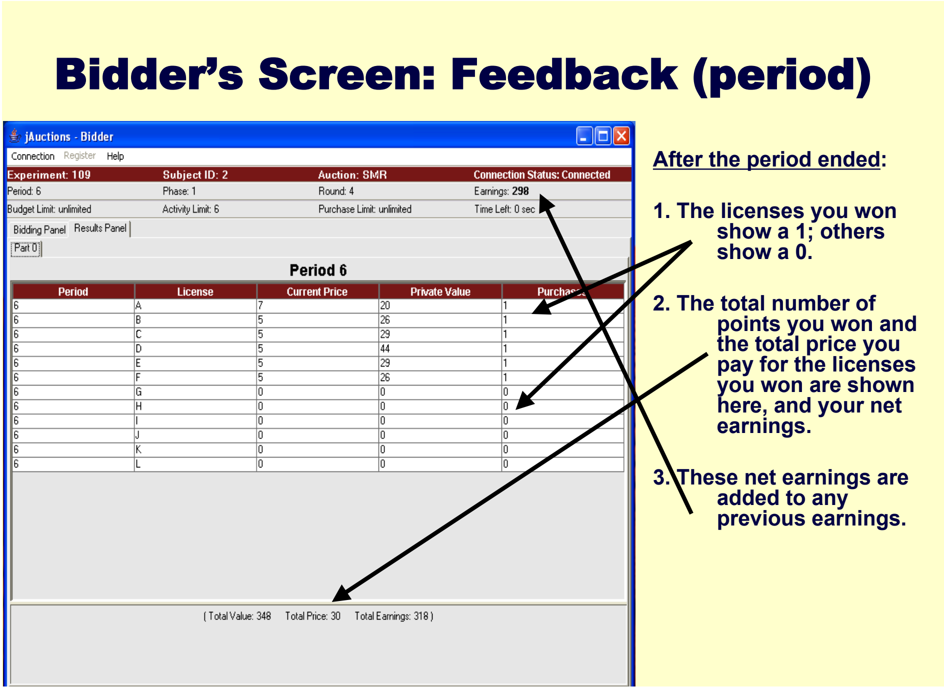


Figure C.31. Slide 31.

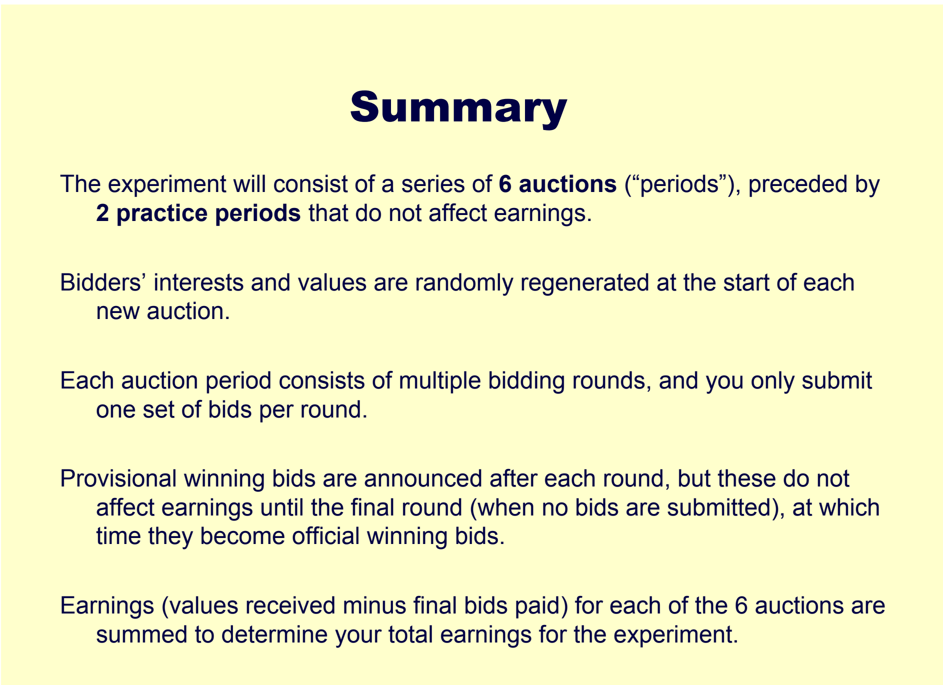


Figure C.32. Slide 32.

Ready to Begin

In a minute, the first of two practice periods will begin. Please raise your hand if you have any questions during these periods, and someone will come to your position to help. These periods will not affect your earnings.

After the practice periods there will be 6 auction periods, and the number of rounds in each is not fixed, so to speed things up, **please make your bids early in a round**. If you are not interested in raising your bids, press “Submit Bids” with an empty basket or click on the **“I am out”** button so that people do not have to wait the full 60 seconds. Please keep in mind that when you click the “I am out” button, you will no longer be able to bid in that period, that is, you will not be able to bid in any of the remaining rounds of the auction.

When all bidders submit empty baskets, the auction ends and a results screen showing your winning bids and earnings will appear.

Figure C.33. Slide 33.

Concluding Remarks

The exchange rate used in the experiment is $0.40 per point, so **10 points = $4.**

You will be paid at the end of the experiment the total amount you have earned in all of the periods (auctions). You need not tell any other participant how much you earned.

To ensure your privacy and that of others in the experiment, please pull out the dividers as far as they will go.

Figure C.34. Slide 34.

C.5 Experimental Instructions for the Activity Rule Used in the Simultaneous Multi-Round Auction, Resource Allocation Design, and the Combinatorial Clock

As mentioned earlier, the instructions for all auction formats differed from SMRPB in their description of the activity rule. The according slides are shown in figures C.35 and C.36.

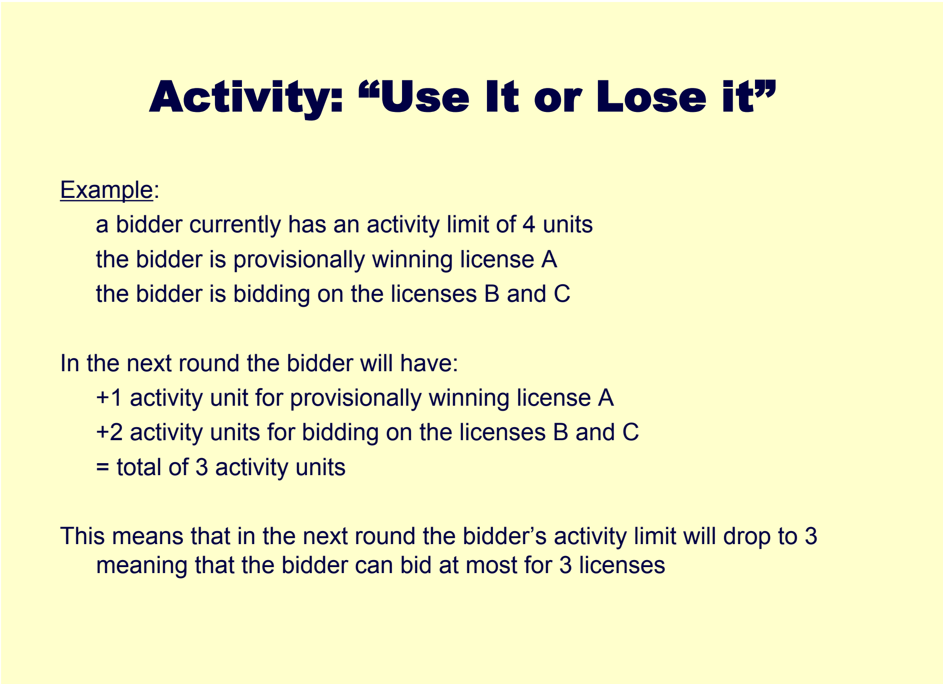


Figure C.35. Activity Rule I.

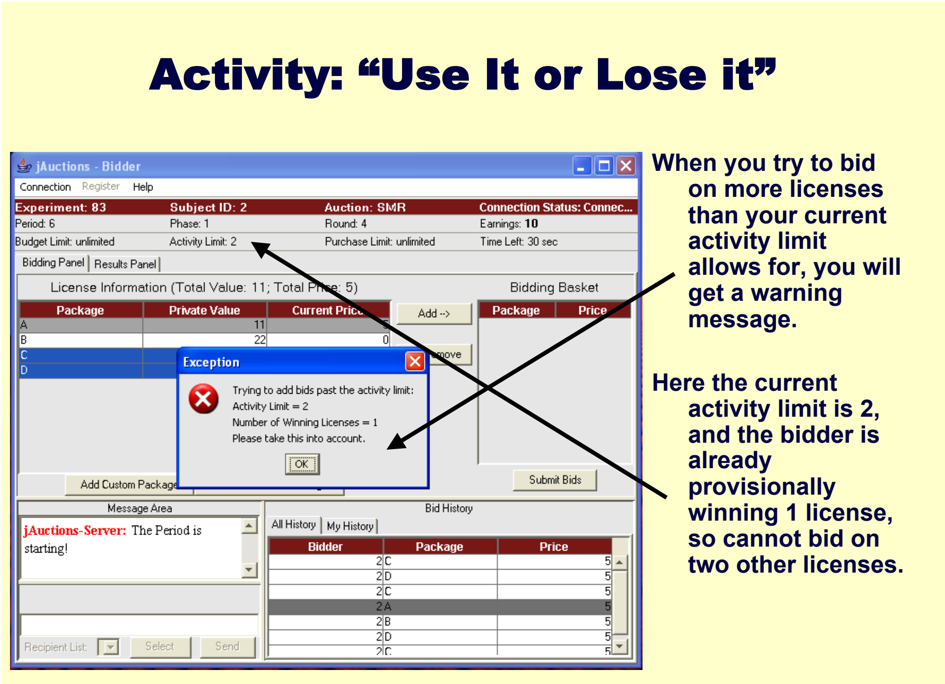


Figure C.36. Activity Rule II.

C.6 Experimental Instructions for the Pricing Rule Used in the Combinatorial Clock Auction

In the combinatorial clock auction, prices are determined in a completely different way than in SMRPB. Figures C.37 and C.38 correspond to the according slides.

Prices

In the first round of an auction, prices for all licenses and packages are 0. In the second round, the price of a license **increases to 5 points if more than one bidder has the license in his bidding basket** at the end of the first round. Otherwise, the price of the license remains the same as in the previous round.

Prices in subsequent rounds are determined in the same manner. The price of a license increases by 5 points if more than one bidder has the license in his bidding basket at the end of the previous round.

Example 1:

- Bidder 1 bids for package AB and license A
- Bidder 2 bids for license B
- No further bids

Therefore, the price of license B increases by 5 points while the price for license A remains the same. Note: If you are the only bidder bidding on a certain license, its price will not go up even if you bid for that license separately as well as in a package.

Figure C.37. Pricing for the Combinatorial Clock I.

Prices

Example 2:

- Bidder 1 bids for package AB and license C
- Bidder 2 bids for package AC
- No further bids

Therefore, the price of licenses A and C increases by 5 points while the price of license B remains the same in the next round.

The price of a **package** is equal to the **sum of the prices of the individual licenses** contained in the package. Therefore, in example 2, the prices for packages ABC and AC increase by 10 points while the prices for packages AB and BC increase by 5 points.

Figure C.38. Pricing for the Combinatorial Clock II.

Table C.1. Efficiency, Revenue, and Profits by Auction Format.

	HOHC				LOHC				HOLC				LOLC			
	SMR	CC	RAD	SMRPB	SMR	CC	RAD	SMRPB	SMR	CC	RAD	SMRPB	SMR	CC	RAD	SMRPB
Efficiency																
Wave 1	86%	92%	89%	88%	86%	89%	93%	89%	96%	92%	95%	88%	96%	82%	94%	93%
Wave 2	85%	94%	89%	93%	79%	91%	89%	91%	96%	93%	95%	84%	97%	93%	93%	86%
Wave 3	84%	86%	94%	93%	85%	87%	94%	87%	98%	96%	98%	93%	96%	96%	99%	92%
Average	**85%**	**91%**	**90%**	**91%**	**83%**	**89%**	**92%**	**89%**	**97%**	**94%**	**96%**	**88%**	**96%**	**90%**	**95%**	**90%**
Revenue																
Wave 1	54%	70%	56%	46%	40%	59%	50%	49%	30%	46%	40%	21%	30%	47%	29%	38%
Wave 2	55%	71%	52%	44%	42%	56%	38%	48%	32%	42%	33%	15%	28%	40%	41%	36%
Wave 3	42%	51%	43%	41%	38%	45%	42%	33%	30%	35%	23%	28%	24%	39%	34%	23%
Average	**50%**	**64%**	**50%**	**44%**	**40%**	**53%**	**43%**	**43%**	**31%**	**41%**	**32%**	**21%**	**27%**	**42%**	**35%**	**32%**
Profit Nationals																
Wave 1	3%	3%	10%	10%	4%	1%	1%	-3%	4%	2%	-3%	5%	3%	2%	0%	2%
Wave 2	-4%	3%	6%	7%	-2%	3%	4%	2%	4%	2%	3%	1%	2%	1%	4%	4%
Wave 3	1%	5%	9%	6%	1%	2%	1%	2%	3%	0%	3%	4%	1%	1%	4%	2%
Average	**0%**	**4%**	**8%**	**8%**	**1%**	**2%**	**2%**	**0%**	**3%**	**2%**	**1%**	**3%**	**2%**	**1%**	**3%**	**3%**
Losses	2%	0%	0%	0%	0%	0%	0%	2%	0%	0%	1%	0%	0%	0%	0%	0%
Profit Regionals																
Wave 1	29%	19%	23%	32%	48%	29%	41%	43%	62%	44%	58%	63%	63%	32%	64%	53%
Wave 2	33%	19%	31%	42%	38%	32%	47%	42%	60%	49%	59%	68%	66%	52%	48%	46%
Wave 3	41%	31%	41%	46%	46%	40%	51%	52%	65%	60%	72%	60%	71%	56%	61%	67%
Average	**34%**	**23%**	**32%**	**40%**	**44%**	**33%**	**47%**	**46%**	**63%**	**51%**	**63%**	**64%**	**67%**	**47%**	**58%**	**55%**
Losses	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	1%	1%	1%
# Rounds	**15.2**	**22.8**	**24.2**	**17.1**	**16.3**	**22.8**	**22.2**	**18.7**	**13.7**	**18.1**	**16.4**	**17.9**	**14.9**	**18.8**	**15.8**	**21.8**
Unsold Licenses	**4%**	**1%**	**0%**	**0%**	**7%**	**0%**	**0%**	**0%**	**0%**	**1%**	**0%**	**0%**	**0%**	**0%**	**0%**	**0%**

Bibliography

[1] Alevy, J. E., M. S. Haigh, and J. A. List (1997) "Information cascades: Evidence from a field experiment with financial market professionals," *The Journal of Finance*, 62(1), 151–180.

[2] Anderson, L. R. and C. A. Holt (1997) "Information cascades in the laboratory," *American Economic Review*, 87, 847–862.

[3] Ausubel, L. M., P. Cramton, and P. Milgrom (2006) "The clock-proxy auction: A practical combinatorial design," *Combinatorial Auctions*, Cramton, P., R. Steinberg, and Y. Shoham, eds., MIT Press, 115–138.

[4] Avery, C. and P. Zemsky (1998) "Multidimensional uncertainty and herd behavior in financial markets," *American Economic Review*, 88, 724–748.

[5] Banerjee, A. V. (1992) "A simple model of herd behavior," *Quarterly Journal of Economics*, 107 (3), 797–817.

[6] Banks, J., J.O. Ledyard, and D. Porter (1989) "Allocating uncertain and unresponsive resources: An experimental approach," *Rand Journal of Economics*, 20(1), 1–25.

[7] Banks, J., M. Olson, D. Porter, S. Rassenti, and V. Smith (2003) "Theory, experiment and the federal communications commission spectrum auctions," *Journal of Economic Behavior and Organization*, 51, 303–350.

[8] Barner, M., F. Feri, and C. R. Plott (2005) "On the microstructure of price determination and information aggregation with sequential and asymmetric information arrival in an experimental asset market," *Annals of Finance*, 1, 73–107.

[9] Bikhchandani, S., D. Hirschleifer, and I. Welch (1992) “A theory of fads, fashion, custom, and cultural change as informational cascades,” *Journal of Political Economy*, 100(5), 992–1026.

[10] Bose, S., G. Orosel, M. Ottaviani, and L. Vesterlund (2009) “Monopoly pricing in the binary herding model,” *Economic Theory*, forthcoming.

[11] Callander, S. and J. Hörner (2009) “The wisdom of the minority,” *Journal of Economic Theory*, 144(4), 1421–1439.

[12] Camerer, C. F. (2003) *Behavioral game theory: Experiments on strategic interaction*. Princeton: Princeton University Press.

[13] Camerer, C. F., T.-H. Ho, and J.-K. Chong (2004) “A cognitive hierarchy model of games,” *Quarterly Journal of Economics*, 119(3), 861–898.

[14] Capra, C. M., R. Gomez, J. K. Goeree, and C. H. Holt (2002), “Learning and noisy equilibrium behavior in an experimental study of imperfect price competition,” *International Economic Review*, 43(3), 613–636.

[15] Çelen, B. and S. Kariv (2004a) “Distinguishing informational cascades from herd behavior in the laboratory,” *American Economic Review*, 94(3), 484–498.

[16] Çelen, B. and S. Kariv (2004b) “Observational learning under imperfect information,” *Games and Economic Behavior*, 47(1), 72–86.

[17] Cipriani, M. and A. Guarino (2005) “Herd behavior in a laboratory financial market,” *American Economic Review*, 95(5), 1427–1443.

[18] Crawford, V. P. and N. Iriberri (2007a) “Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games,” *American Economic Review*, 97, 1731-1750.

[19] Crawford, V. P. and N. Iriberri (2007b) “Level-k auctions: Can a non-equilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions?” *Econometrica*, 75, 1721–1770.

[20] Drehmann, M., J. Öchssler, and A. Roider (2005a) "Herding and contrarian behavior in financial markets—an internet experiment," *American Economic Review*, 95(5), 1403–1426.

[21] Drehmann, M., J. Oechssler, and A. Roider (2005b) "Herding with and without payoff externalities—an internet experiment," University of Heidelberg Discussion Paper No. 420.

[22] Fama, E. (1970) "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, 25, 383–417.

[23] Fligner, M. A. and G. E. Policello (1981) "Robust rank procedures for the Behrens-Fisher problem," *Journal of the American Statistical Association*, 76(373), 162–168.

[24] Garber, P. M. (1990) "Famous first bubbles," *Journal of Economic Perspectives*, 4 (2), 35–54.

[25] Goeree, J. K. and C. A. Holt (2002) "Learning in economics experiments," *Encyclopedia of Cognitive Science*, volume 2, L. Nagel, ed., London: Nature Publishing Group, McMillan, 1060–1069.

[26] Goeree, J. K. and C. A. Holt (2004) "A model of noisy introspection," *Games and Economic Behavior*, 46(2), 2004, 365–382.

[27] Goeree, J. K. and C. A. Holt (2005) "Comparing the FCC's combinatorial and non-combinatorial simultaneous multi-round auctions: Experimental design report," `http://wireless.fcc.gov/auctions/default.htm?job=papers_studies#experiments`, prepared for the wireless communications bureau of the Federal Communications Commission.

[28] Goeree, J. K. and C. A. Holt (2008) "Hierarchical package bidding: A paper & pencil combinatorial auction," *Games and Economic Behavior*, forthcoming.

[29] Goeree, J. K., C. A. Holt, and T. R. Palfrey (2003) "Risk averse behavior in generalized matching pennies games," *Games and Economic Behavior*, 45(1), 97–113.

[30] Goeree, J. K., T. R. Palfrey, and B. Rogers (2006) "Social learning with private and common values," *Economic Theory*, 28(2), 245–264.

[31] Goeree, J. K., T. R. Palfrey, B. Rogers, and R. McKelvey (2007) "Self-correcting information cascades," *Review of Economic Studies*, 74(3), 733–762.

[32] Hey, J. D. and A. Morone (2004) "Do markets drive out lemmings—or vice versa?" *Economica*, 71, 637–659.

[33] Hung, A. A. and C. R. Plott (2000) "Information cascades: Replication and an extension to majority rule and conformity rewarding institutions," *American Economic Review*, 2001, 91(5), 1508–1520.

[34] Krämer, C., M. Nöth, and M. Weber (2006) "Information aggregation with costly information and random ordering: Experimental evidence," *Journal of Economic Behavior and Organization*, 59, 423–432.

[35] Kübler, D. and G. Weizsäcker (2004) "Limited depth of reasoning and failure of cascade formation in the laboratory," *Review of Economic Studies*, 71, 425–441.

[36] Kwasnica, A. M., J. O. Ledyard, D. Porter, and C. DeMartini (2005) "A new and improved design for multi-object iterative auctions," *Management Science*, 51(3), 419–434.

[37] Ledyard, J. O., D. Porter, and A. Rangel (1997) "Experiments testing multiobject allocation mechanisms," *Journal of Economics and Management Strategy*, 6(3), 639–675.

[38] Lee, I. H. (1998) "Market crashes and information avalanches," *Review of Economic Studies*, 65, 741–759.

[39] McAfee, P. R. and J. McMillan (1996) "Analyzing the airwaves auction," *Journal of Economic Perspectives*, 10(1), 159–175.

[40] McCabe, K., S. Rassenti, and V. Smith (1989) "Designing smart computer assisted markets," *European Journal of Political Economy*, 5, 259–283.

[41] McKelvey, R. D., T. R. Palfrey, and R. A. Weber (2000) "The effects of payoff magnitude and heterogeneity on behavior in 2×2 games with a unique mixed strategy equilibrium," *Journal of Economic Behavior and Organization*, 42(4), 523–548.

[42] Milgrom, P. (2004) *Putting auction theory to work*, New York: Cambridge University Press.

[43] Milgrom, P. and N. Stokey (1982) "Information, trade and common knowledge," *Journal of Economic Theory*, 26, 17–27.

[44] Nöth, M. and M. Weber (1993) "Information aggregation with random ordering: Cascades and overconfidence," *The Economic Journal*, 113, 166–189.

[45] Ochs, J. (1995) "Games with a unique mixed-strategy equilibrium: An experimental study," *Games and Economic Behavior*, 10, 202–217.

[46] Ottaviani, M. and P. Sørensen (2001) "Information aggregation in debate: Who should speak first?" *Journal of Public Economics*, 81, 393–421.

[47] Porter, D. (1999) "The effect of bid withdrawal in a multi-object auction," *Review of Economic Design*, 4(1), 73–97.

[48] Porter, D., S. Rassenti, A. Roopnarine, and V. Smith (2003) "Combinatorial auction design," *Proceedings of the National Academy of Sciences*, 100(19), 11153–11157.

[49] Plott, C. R. (2000) "Markets as information gathering tools," *Southern Economic Journal*, 67(1), 1–15.

[50] Plott, C. R. and S. Sunder (1982) "Efficiency of experimental security markets with insider information: An application of rational-expectations models," *Journal of Political Economy*, 90(4), 663–698.

[51] Plott, C. R. and S. Sunder (1988) "Rational expectations and the aggregation of diverse information in laboratory security markets," *Econometrica*, 56(5), 1085–1118.

[52] Rassenti, S., V. Smith, and R. Bulfin (1982) "A combinatorial auction mechanism for airport time slot allocation," *Bell Journal of Economics*, 13, 402–417.

[53] Rothkopf, M. H., A. Pekec, and R. M. Harstad (1998), "Computationally manageable combinational auctions," *Management Science*, 44, 1131–1147.

[54] Selten, R. and T. Chmura (2008) "Stationary concepts for experimental 2×2 games," *American Economic Review*, 98(3), 938–966.

[55] Smith, L. and P. N. Sørensen (2000) "Pathological outcomes of observational learning," *Econometrica*, 68(2), 371–398.

[56] Surowiecki, J. (2004) *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations*, Little: Brown.

[57] Vuong, Q. H. (1989) "Likelihood ratio tests for model selection and non-nested hypotheses" *Econometrica*, 57(2), 307–333.