The Double Auction Market Institutions, Theories, and Evidence

PROCEEDINGS OF THE WORKSHOP ON DOUBLE AUCTION MARKETS
HELD JUNE, 1991
IN SANTA FE, NEW MEXICO

Editors

Daniel Friedman
Department of Economics
University of California, Santa Cruz

John Rust
Department of Economics
University of Wisconsin

Proceedings Volume XIV

Santa Fe Institute
Studies in the Sciences of Complexity

Addison-Wesley Publishing Company
The Advanced Book Program

Reading, Massachusetts Menlo Park, California New York
Don Mills, Ontario Wokingham, England Amsterdam Bonn
Sydney Singapore Tokyo Madrid San Juan
Paris Seoul Milan Mexico City Taipei

Publisher: *David Miller*
Production Manager: *Michael Cirone*

Director of Publications, Santa Fe Institute: *Ronda K. Butler-Villa*
Publications Assistant, Santa Fe Institute: *Della L. Ulibarri*
Publications Secretary, Santa Fe Institute: *Joanne Carlsen*

Library of Congress Cataloging-in-Publication Data

Workshop on Double Auction Markets (1991 : Santa Fe, N.M.)
The double auction market : institutions, theories, and evidence / editors, Daniel Friedman, John Rust.
p. cm. — (Proceedings volume, Santa Fe Institute studies in the sciences of complexity ; 14)
"Proceedings of the Workshop on Double Auction Markets, held June, 1991 in Santa Fe, New Mexico."
Includes bibliographical references and index.
ISBN 0-201-62263-7
ISBN 0-201-62459-1 (pbk.)
1. Program trading (Securities)—Congresses. 2. Stock-exchange—Data processing—Congresses. 3. Securities industry—Data processing—Congresses. 4. Commodity exchanges—Data processing—Congresses. I. Friedman, Daniel. II. Rust, John.
III. Title. IV. Series: Proceedings volume in the Santa Fe Institute studies in the sciences of complexity ; v. 14.
HG4515.5.W67 1991 92-34829 CIP
332.63'0285—dc20

This volume was typeset using TEXtures on a Macintosh II computer. Camera-ready output from a NEC Laser Printer.

1 2 3 4 5 6 7 8 9 10 - MA - 95 94 93 92
First printing, January 1993

About the Santa Fe Institute

The *Santa Fe Institute* (SFI) is a multidisciplinary graduate research and teaching institution formed to nurture research on complex systems and their simpler elements. A private, independent institution, SFI was founded in 1984. Its primary concern is to focus the tools of traditional scientific disciplines and emerging new computer resources on the problems and opportunities that are involved in the multidisciplinary study of complex systems—those fundamental processes that shape almost every aspect of human life. Understanding complex systems is critical to realizing the full potential of science, and may be expected to yield enormous intellectual and practical benefits.

All titles from the *Santa Fe Institute Studies in the Sciences of Complexity* series will carry this imprint which is based on a Mimbres pottery design (circa A.D. 950–1150), drawn by Betsy Jones.

Santa Fe Institute Studies in the Sciences of Complexity

PROCEEDINGS VOLUMES

Volume	Editor	Title
I	David Pines	Emerging Syntheses in Science, 1987
II	Alan S. Perelson	Theoretical Immunology, Part One, 1988
III	Alan S. Perelson	Theoretical Immunology, Part Two, 1988
IV	Gary D. Doolen et al.	Lattice Gas Methods for Partial Differential Equations, 1989
V	Philip W. Anderson, Kenneth Arrow, & David Pines	The Economy as an Evolving Complex System, 1988
VI	Christopher G. Langton	Artificial Life: Proceedings of an Interdisciplinary Workshop on the Synthesis and Simulation of Living Systems, 1988
VII	George I. Bell & Thomas G. Marr	Computers and DNA, 1989
VIII	Wojciech H. Zurek	Complexity, Entropy, and the Physics of Information, 1990
IX	Alan S. Perelson & Stuart A. Kauffman	Molecular Evolution on Rugged Landscapes: Proteins, RNA and the Immune System, 1990
X	Christopher Langton et al.	Artificial Life II, 1991
XI	John A. Hawkins & Murray Gell-Mann	Evolution of Human Languages, 1992
XII	Martin Casdagli & Stephen Eubank	Nonlinear Modeling and Forecasting, 1992
XIII	Jay E. Mittenthal & Arthur B. Baskin	Principles of Organization in Organisms, 1992
XIV	Daniel Friedman & John Rust	The Double Auction Market: Institutions, Theories, and Evidence, 1993

LECTURES VOLUMES

Volume	Editor	Title
I	Daniel L. Stein	Lectures in the Sciences of Complexity, 1989
II	Erica Jen	1989 Lectures in Complex Systems, 1990
III	Lynn Nadel & Daniel L. Stein	1990 Lectures in Complex Systems, 1991
IV	Lynn Nadel & Daniel L. Stein	1991 Lectures in Complex Systems, 1992

LECTURE NOTES VOLUMES

Volume	Author	Title
I	John Hertz, Anders Krogh, & Richard Palmer	Introduction to the Theory of Neural Computation, 1990
II	Gérard Weisbuch	Complex Systems Dynamics, 1990

REFERENCE VOLUMES

Volume	Author	Title
I	Andrew Wuensche & Mike Lesser	The Global Dynamics of Cellular Automata: An Atlas of Basin of Attraction Fields of One-Dimensional Cellular Automata, 1992

Contributors to This Volume

Tim Bollerslev, Northwestern University
Colin Camerer, University of Chicago
Tim Cason, University of Southern California
Laura Clauser, Occidental College
Ian Domowitz, Northwestern University
David Easley, Cornell University
Daniel Friedman, University of California, Santa Cruz
Dhananjay Gode, Carnegie-Mellon University
John Kagel, University of Pittsburgh
John Ledyard, California Institute of Technology
John Miller, Carnegie-Mellon University
Kevin McCabe, University of Minnesota
John O'Brien, Carnegie-Mellon University
Richard Palmer, Duke University
Charles Plott, California Institute of Technology
Stephen Rassenti, University of Arizona
John Rust, University of Wisconsin
Mark Satterthwaite, Northwestern University
Vernon Smith, University of Arizona
Sanjay Srivastava, Carnegie-Mellon University
Shyam Sunder, Carnegie-Mellon University
William Vogt, Stanford University
Keith Weigelt, University of Pennsylvania
Steven Williams, University of Illinois
Robert Wilson, Stanford University

Contents

Preface

This book is a collection of papers focused on markets organized as double auctions. In a double auction (DA), both buyers and sellers can submit bids (offers to buy) and asks (offers to sell) for standardized units of well-defined commodities and securities. The DA is a general name for a broad class of trading institutions surveyed by Dan Friedman in chapter 1. A classic example of a DA market (known by experimentalists as the *continuous DA* and by practitioners as an *open-outcry market*) is the commodity trading pit at the Chicago Board of Trade. A related institution is the *clearinghouse* (also known as a *call market* or *sealed-bid auction*), where traders submit bids and asks to a clearinghouse which periodically "crosses" the list of bids and ask to determine a market-clearing price, i.e., a price which equates the revealed supply and demand. A call market is used to determine opening prices on the New York Stock Exchange.

Already the predominant trading institution for financial and commodities markets, the DA has many variants and is evolving rapidly in the present era of advancing computer technology and regulatory reform. DA markets are of intense theoretical as well as practical interest in view of the central role these institutions play in allocating resources. Although the DA has been studied intensively in the laboratory, and practitioneers have considerable experience in the field, only recently have tools started to become available to provide the underpinnings of a behavioral theory of DA markets.

Many of the ideas for this book originated in the late 1980s when the DA institution was the subject of increased scrutiny by regulators and practitioners at the New York and Chicago exchanges, and by economic and financial theorists, experimental economists, and even computer scientists. However, the investigations remained disjoint and interchange among the various groups was limited. The Santa Fe Insitute sponsored a small informal conference in May 1990 to bring together active investigators from the diverse groups. Participants learned a great deal from the conference discussions and agreed that the time was ripe for a coordinated attack on some of the unsolved issues raised by DA markets. A second conference on DA markets was held at the Santa Fe Institute in June 1991: this book is a collection of the papers presented at that meeting. The papers have been rewritten, very substantially in some cases, to reflect the lively and multifaceted discussions provoked by the presentations. In order to get a perspective on how the work in this volume contributes to our understanding of DA markets, it is useful to provide a brief historical discussion of the theory and practice of DA markets.

The origins of DA markets are not well documented. Trading in bazaars and marketplaces can be traced back at least as far as ancient Egypt and Mesopotamia.[4,8,10] Undoutedly the process of "haggling" between buyers and sellers that occurred in these markets bears many similarities to the multilateral bargaining that occurs in modern DA markets. The increase in international commerce in medieval Italy lead to the invention of double-entry bookkeeping, bills of exchange, limited liability, and the first modern banks. These developments paved the way for trading of more abstract objects, stocks and bonds, in places we now recognize as organized exchanges. By the seventeenth century, government war bonds and shares of joint stock companies were regularly traded in London coffeehouses, culminating in the foundation of the London Stock Exchange in the late 1700s and the construction of the first exchange building in 1802.[16] By the early 1800s similar institutions existed in many other countries including the United States. While we have much better information about the trading rules in these exchanges, many of the trading rules might be better regarded as commonly understood "customs" rather than precisely codified regulations. We do know that the New York Stock exchange operated as a call market until the 1860s, at which time the rapid growth in trading volume prompted a move to continuous trading with *specialists* serving as *market makers* holding inventories of particular securities.[11] The survey by Daniel Friedman in chapter 1 elaborates on the distinctions among current market institutions.

There are a great variety of different trading institutions around the world but, unfortunately, there is no clear record documenting the evolution of trading rules and practices. We do know that technological developments have spurred many changes in market organization. For example the invention of the telegraph and telephone in the mid 1800s lead to rapid integration of regional and international exchanges. Technological advances have not always been openly embraced by the dominant exchanges. New technologies have often been perceived as threats and were ultimately adopted only when forced by increasing competition:

> "With the appearance of the telephone in the late 1870s, the management of the stock exchange faced a new threat: those trading on the floor could now be directly connected with interested parties outside. The first application to introduce telephone service between the London Stock Exchange and outside subscribers was made in November 1879. In contrast to the London Stock Exchange, the provincial stock exchanges accepted the telephone almost immediately. Faced with the continued refusal to provide proper facilities for telephones, in July 1888 the members threatened to find an alternative to the facilities provided by the stock exchange unless the managers gave way. The management backed down, putting only minimal obstacles in the way of further expansion; the year 1888 marked the end of any serious resistance by the management to the changes created by the communications revolution." (Mitchie,[11] p. 20)

In the last 15 years the computer revolution has precipitated another wave of changes in market integration and organization, including the development of computerized trading algorithms and automated trading systems. By all accounts some of the dominant exchanges (such as the New York Stock Exchange and the Chicago Board of Trade) have resisted these developments in a way reminiscent of the London Stock Exchange's resistence to the telephone, although in this case it is the floor traders who have actively opposed the changes.[1] However, competition from new exchanges may ultimately force the NYSE to capitulate and adopt automated trading practices. In chapter 2, Ian Domowitz surveys the great variety of automated trading systems that have recently appeared. Much of the growth appears to be a response to the need for 24-hour trading in a global marketplace, but there has also been entry of entirely new exchanges in which new trading rules are embodied in innovative computerized trading systems designed to lower trading costs and improve market efficiency and "price discovery." An example is the Wunsch Auction System (WAS) that recently opened in Arizona. This market has aspects of both a call market and an open-outcry auction, except that trading is done via a network of computer terminals rather than verbally in the traditional "trading pit." In chapter 11, Kevin McCabe, Steven Rassenti, and Vernon Smith conduct an experimental analysis of a "uniform price double auction" (UPDA) that is similar in many respects to the WAS. They find that UPDA compares favorably to the continuous DA both in terms of price volatility and market efficiency. In chapter 8 Tim Bollerslev and Ian Domowitz analyze markets with an electronic "order book" that stores incoming limit orders to buy or sell a security. The price and time priority rules for executing orders from the book are modeled closely after the new "GLOBEX" trading system that recently became operational for night trading on the Chicago Mercantile Exchange. They find that increases in the length of the order book from one (which corresponds to the continuous DA) to four or five can

[1] Note that unlike its previous opposition to the telephone, the London Stock Exchange was one of the first to adopt a computerized trading system, and within weeks of its adoption, the old open-outcry trading floor had been completely abandoned.

significantly reduce transaction price volatility without compromising market efficiency, but that further increases in book length do not seem to have significant additional effects.

With the advent of systems like WAS and GLOBEX, we are beginning to see a more abstract competition at the level of trading rules, providing investors and traders with a menu of alternative organized exchanges on which they can trade. This competition will surely lead to lower transaction costs, greater access, and faster execution. It is not yet clear, however, whether slightly modified versions of existing trading institutions will maintain their dominant position, or whether they will be displaced by entirely new trading institutions.

This book is motivated by the limitations of current economic theory, which offers few insights about the pros and cons of different trading systems and makes no predictions about which types of institutions might emerge from this higher level competition in trading rules. Indeed, we are only just beginning to attain a theoretical understanding of trading strategies and price formation in the simplest market institutions. Development of a general theory of trading under alternative market institutions will require major intellectual advances. The existing "general equilibrium theory" of Kenneth Arrow and Gerard Debreu represented a major intellectual advance of the 1950s and 1960s, but is inadequate as a tool for analyzing the great variety of specific trading institutions we observe in the field. The Arrow-Debreu theory provides an elegant mathematical model of general equilibrium of an economy, employing topological fixed-point theorems to establish the existence of a *competitive equilibrium* (CE), i.e., a set of prices which equate the demands of utility-maximizing consumers to the supplies of profit-maximizing firms. The theory also elucidated the conditions under which a CE allocation is *Pareto efficient*, providing a rigorous justification for the intuitive notion of the desirability of competition that dates back to Adam Smith.[2] The shortcoming of the Arrow-Debreu theory is that it attempts to be *institution free:* although it proves that a CE exists, it does not specify the dynamic process by which an economy actually gets to a CE. The typical textbook story relies on a "Walrasian auctioneer" who adjusts prices in various markets to equilibrate supply and demand. Of course, in most markets no such auctioneer exists, and in the early 1970s a large literature arose on alternative price-quantity adjustment or *tâtonnement* processes.[3] The key feature of these processes is that agents are not permitted to trade "out of equilibrium," i.e., until the adjustment process has discovered a CE price vector p. This literature seems to have died out, partly due to the ad hoc and unrealistic nature of these processes, and partly due to Herbert Scarf's 1960 counterexample that these processes (which are essentially deterministic dynamical systems) can get trapped in "limit cycles" that never converge to CE. Scarf did develop a combinatorial fixed-point algorithm that is guaranteed to find a CE, but the algorithm requires a single agent (a "central planner") to have complete knowledge of the structure of the economy.

[2] An allocation is said to be *Pareto efficient* if there is no alternative allocation that can make all consumers and producers better off.

[3] *Tâtonnement* is French for "groping."

However, as Hayek pointed out in 1945,[7] a key economic problem is that information is *decentralized:* "the problem is to show how a solution is produced by the interactions of people each of whom has partial knowledge" (p. 530). Conceptually, one possible way a CE might be computed is for each agent to report their preferences, endowments, and production technologies to a central planner, who would then use Scarf's fixed-point algorithm to compute a CE. However, such a procedure is clearly unrealistic, since it is not clear that people would be able to communicate their preferences as mathematical "utility functions" needed by Scarf's algorithm. Even if they could, there is a further problem with such a procedure first noted by Leonid Hurwicz[9]: the reporting procedure is not *incentive compatible,* i.e., agents will generally have an incentive to misrepresent their preferences and endowments to the central planner to distort the computation of CE to their advantage.

Much of what we know about DA markets is not due to theory but due to an accumulated body of experimental research dating back to the early 1960s by researchers such as Vernon Smith, Charles Plott, and others. The advantage of studying experimental markets rather than direct observation of field institutions is that we can observe human behavior in a simplified, controlled environment and can modify the institution in various ways to test specific hypotheses. The market institution studied most extensively in the laboratory is the continuous DA. By assigning subjects to play the roles of buyers and sellers with fixed redemption values and costs, the experimenter can construct well-defined supply and demand curves for which the CE can be readily calculated. Individual traders in this market only know their own private token valuations, so they face a complicated problem deciding how much to bid or ask and whether to accept an outstanding bid or ask. Since these markets run for fixed amount of time, typically several minutes, there is a clear tradeoff between waiting sufficiently long to learn what a "fair price" might be versus waiting too long and getting preempted by other traders. Despite the inherent difficulty of traders' learning/decision problem, the nearly universal finding of nearly 30 years of experimental research on the continuous DA is that prices and quantities typically converge to the CE and result in allocations that are nearly 100% efficient. One of the remarkable findings is that we see competitive behavior in these markets even with very small numbers of traders, confounding the conventional economic wisdom that competitive behavior emerges only in large markets where single individuals do not have much influence on prices. Why DA markets work so well has been viewed by Vernon Smith as something of a "scientific mystery." In the language of computer science, convergence to CE can be viewed as an "emergent computation" resulting from the loosely coordinated interactions of the individual traders.[5]

By the 1980s developments in game theory, especially the concept of *Bayesian Nash equilibrium,* lead to the possibility that economists might be able to model a competitive market more realistically as an institution with an explicit set of *trading rules*, and "explain" the apparent disequilibrium behavior of uncoordinated traders in such a market as the equilibrium path of a a well-defined *game of incomplete information.* Such an approach might result in a theory of convergence to CE that deals with Hayek's problem of informational decentralization and Hurwicz's problem of incentive compatibility. An excellent example of this approach is Satterthwaite and Williams' analysis of the call market or k-double auction institution presented in chapter 4. Their analysis (and indeed most of the analysis in this book) is conducted in the context of the simplest possible general equilbrium model where traders exchange two goods: an indivisible good, *tokens,* and a divisible good, *money.* Satterthwaite and Williams analyze a particular type of DA market called the k-DA in which buyers and sellers submit sealed bids and asks which are arrayed into revealed supply and demand curves and a market-clearing price determined. In general, buyers will have an incentive to bid below their true valuations and sellers have an incentive to ask more than their true costs. However, Satterthwaite and Williams show that in a large market "truthtelling" is nearly a dominant strategy, a result that bears many similarities to Roberts and Postlewaite's[15] 1976 result that pricetaking behavior is nearly a dominant strategy in large complete-information exchange economies. The innovation of the Satterthwaite-Williams result is that it recognizes that traders have incomplete information and models trader behavior within the context of an explicit trading institution. Their results imply that in a k-DA prices and quantities approach the CE and market efficiency tends to 100% as the number of traders increases, providing a game-theoretic "explanation" of convergence to CE and the efficiency of the k-DA institution. To complement these asymptotic results, Satterthwaite and Williams numerically calculated equilibria and showed that even in markets with very small number of traders, the deviations from truthtelling are not very large. This result provides an explanation of the high efficiency observed in experimental markets with small numbers of traders.

The successes of game theory lead to an even more ambitious theory of *mechanism design* in which the institutional rules governing trading are treated as variables to be optimized. One of the first applications of mechanism design was by Roger Myerson[12] in 1981 who used the *revelation principle* to show that among all trading rules a second price auction with an appropriately determined reservation price maximizes the expected revenue to the seller in a single-sided auction. Robert Wilson[19] used this approach to show that when the number of traders N is sufficiently large, a clearinghouse similar to the k-DA is *incentive efficient:* i.e., no other trading mechanism would be preferred by all traders even after knowing their private information about their token valuations. A limitation of the mechanism design approach is that results typically depend on specific assumptions about the particular environment. For example, both Myerson's and Wilson's results depend on the assumption of *independent private values,* i.e., traders' token valuations are *IID* draws from a common distribution F.[4] In chapter 5 Wilson develops a much more general framework that can be used to characterize efficient forms of investment and risk-sharing arrangements as well as multilateral trading mechanisms in the more realistic situation where traders' private information is correlated. He

[4] Myerson allowed token draws to be *INID,* that is, independent draws from possibly different distributions for each trader. Thus, the independence assumption is what is really key to these results.

shows how the general framework can be applied to the problem of optimal auction design problems originally studied by Myerson, but in a much more general setting with multilateral trading, multiple token units, and correlation in token valuations.

It is now generally recognized that the standard static "revelation principle" approach to mechanism design is inadequate if we admit the possibility that players in a game can communicate with each other sequentially (Meyerson,[13] chapter 6). For example, the fundamentally static sealed-bid mechanism studied by Wilson may not be the most efficient mechanism if we explicitly model all the communication possibilities that are present, say, in an open outcry auction. This may explain why experimental studies have found that allocations in the continuous DA are significantly more efficient than allocations in the one-shot sealed-bid institutions. To our knowledge the only available game-theoretic analysis of the continuous DA is due to Wilson[20] who derived necessary conditions for a sequential equilibrium of this game. The key insight of Wilson's analysis is that traders will delay making "serious bids" to avoid signalling the value their tokens to the other traders. However, in equilibrium the buyers with the highest values (and sellers with the lowest costs) will be most impatient to trade because they stand the most to lose by waiting too long and being preempted. Wilson's theory thus predicts that in equilibrium, trade will occur in the "efficient order," i.e., the buyer with the highest token value trades with the seller with the smallest token cost, and so on. Wilson also showed that with high probability, transaction-price trajectories will converge to CE, yielding allocations that are nearly 100% efficient. Note that unlike the k-DA where all trading is done at a uniform price p_e, there will usually be considerable variation in successive transaction prices in the continuous DA, although the prices will follow a martingale to preclude intertemporal arbitrage. Thus, while price volatility is necessarily higher in the continuous DA, it is an open theoretical question whether it necessarily yields more efficient allocations than the static sealed-bid DA's such as the k-DA.

While the game-theoretic approach has lead to very important insights into DA markets, there is increasing awareness that the results are critically dependent on a host of assumptions about the rationality of traders, details of the distribution of traders' token values and characteristics, and especially the very strong requirement that traders have *common knowledge* about the trading environment, each others' trading strategies, and the joint distribution of their private information.[5] Game theorists' assumption of perfect rationality has been widely criticized as a model of human behavior on both practical and logical grounds.[3] A characteristic of most game-theoretic models is their exponential computational complexity: even the world's top game theorists are unable to explicitly solve any but the simplest dynamic games. The continuous DA is a case in point: while Wilson's model provides a general characterization of the form of equilibrium trading strategies, nobody has yet been able to calculate the exact timing and size of bids and asks. In view of this, it seems incredible to postulate that "ordinary" humans are making

[5] A fact is said to be *common knowledge* if each trader knows it, each trader knows that every other knows it, each trader knows that every other traders knows it, and so on in infinite regression.

such calculations in real-time trading. Indeed, many of the predictions of Wilson's "waiting game" equilibrium are inconsistent with experimental evidence; in chapter 9 Timothy Cason and Dan Friedman find that less sophisticated strategies (the BGAN and ZI strategies discussed below) appear better able to explain several aspects of human behavior in the continuous DA. Even the simpler static k-DA model analyzed by Satterthwaite and Williams does not appear to yield good predictions of human bidding behavior. In chapter 10 John Kagel and William Vogt conduct an experimental evaluation of a particular case of the k-DA, the 1-DA, (also known as the "buyer's-bid double auction" since the clearing price p_e is typically determined by the buyer's bid). A strong prediction of the Satterthwaite-Williams theory is that in a 1-DA sellers' dominant strategy is to be truthful and submit asks equal to their true costs. Contrary to the prediction of game theory, Kagel and Vogt found that their subjects' asks exceeded their costs over 50% of the time, and asked *below* cost between 18% and 40% of the time. They conclude that "the most significant failing of the theory in our minds is that efficiency fails to increase significantly as the number of traders increase, rather it increases moderately and the direction of change is erratic across experimental sessions" (p. 302).[6]

The common knowledge assumption has also been widely criticized. As Rust, Miller, and Palmer note in chapter 6, "the common-knowledge underpinnings of game theory presume a high degree of implicit coordination amongst the traders, begging Hayek's question of how decentralized coordination is achieved in the first place" (p. 159). And, as Wilson[21] has observed, "Only by repeated weakening of common-knowledge assumptions will the theory approximate reality. But game theory as we presently know it cannot proceed without the fulcrum of common knowledge" (p. 34). At present it appears that there is no way for game-theoretic models to avoid their dependence on a host of particular assumptions that may or may not be true of a given market environment. This implies that conclusions drawn from these models may not be very robust:

> "The optimal trading rule for a direct revelation game is specialized to a particular environment. For example the rule typically depends on the agents' probability assessments about each other's private information. Changing the environment requires changing the trading rule. If left in this form, therefore, the theory is mute on one of the basic problems challenging the theory. I refer to the problem of explaining the prevalence of a few simple trading rules in most of the commerce conducted via organized exchanges. A short list—including auctions, double auctions, bid-ask markets, and specialist trading—accounts for most organized exchange. Indeed, [DA] markets (such as those conducted in the commodities pits) have long been economists' paradigms for the nearly perfect markets addressed by the Walrasian theory of general equilibrium. The rules of these markets are not changed daily as the environment changes; rather they persist

[6] Satterthwaite and Williams offer some insightful comments on the potential causes of these discrepancies in the last section of chapter 4.

> as stable, viable institutions. As a believer that practice advances before theory, and that the task of the theory is to explain how it is that practitioners are (usually) right, I see a plausible conjecture: These institutions survive because they employ trading rules that are efficient for a wide class of environments." (Wilson,[21] p. 37)

These problems have motivated new approaches to studying the DA in which theorists have chosen to discard the assumption that traders are perfectly rational optimizers and instead model traders as employing simple but plausible "rules-of-thumb." The first influential study of the sort is due to Easley and Ledyard, the final version of which is presented in chapter 3. They establish that if traders adopt relatively simple learning procedures, price trajectories in the continuous DA will eventually converge to the CE after sufficiently many replications of the DA game (using the same traders and token values). Their results are consistent with many of the "stylized facts" observed in experimental DA markets, particularly the ability of the market to "learn" the CE price and quantity. The Easley and Ledyard result suggests that the convergence to CE that we observe in experimental markets may be consistent with a wide range of trading strategies, not just the rational game-theoretic delay strategies studied by Wilson. Friedman[6] confirms this result in his model where traders are rational Bayesian decisionmakers who treat the DA as a "game against nature" (BGAN), ignoring the impact of their choice of strategy on the possible strategic choices of their opponents. Friedman showed that a collection of these traders will also generate price trajectories that converge to CE. Perhaps the most dramatic confirmation of the power of the "invisible hand" implicit in the DA trading rules is due to Dhananjay Gode and Shyam Sunder, who showed that continuous DA markets populated by a collection of "zero intelligence" (ZI) traders are highly efficient and price trajectories typically converge to CE. In chapter 9 Gode and Sunder show that efficiency in DA markets has a lower bound "independent of the motivations or abilities of the traders who participate in them" (p. 213). This result is clearly dependent on the nature of the trading institution, however. In chapter 10 Kagel and Vogt show that allocations in sealed-bid auctions such as the 1-DA are much more sensitive to deviations from rationality: ZI traders succeed in exploiting only 30-50% of the surplus in these markets.

A limitation of these studies is that they all depend on a level of implicit co-ordination via the assumption that all traders use identical trading strategies. As noted above, we do observe convergence to CE in continuous DA markets with human subjects who are presumably using different trading strategies, although we can never be sure since experiments only allow us to observe traders' *behavior*. One way to deal with this problem is to hold a tournament in which traders codify their "market intuition" in a standard computer programming language. In chapter 6 John Rust, John Miller, and Richard Palmer report the results of a computerized DA tournament in which a collection of 30 heterogenous computer programs played the role of buyers and sellers vying for a pool of $10,000 prize money. They found that the collection of computer programs behaved very similarly to human subjects, with price trajectories converging to CE and allocations that were nearly 100% efficient. Surprisingly they found that a very simple "wait in the background" trading strategy emerged as the winner of the tournament, beating out many more complex algorithms that used statistically based predictions of future transaction prices, explicit optimizing principles, and sophisticated learning principles. The winning strategy required remarkably little information beyond its private token values, current time, and the current bid and ask. Whether or not good human traders really behave according to a few simple decision rules is the subject of ongoing investigations, but the results do seem to confirm Hayek's conjecture about the economy of information needed to take the right action in a competitive market.

Chapters 10 to 13 present the results of new experimental research on DA and related market institutions. In chapter 10 Kagel and Vogt conduct an experimental test of the predictions of Satterthwaite and Williams' buyer's-bid DA market already described above. In chapter 11 Kevin McCabe, Steven Rassenti, and Vernon Smith use experimental methods to evaluate a new type of trading institution, the *uniform-price double auction* and find that it compares favorably to the standard continuous DA both in terms of market efficiency and price volatility. In chapter 12 Laura Clauser and Charles Plott conduct an experimental test designed to isolate the observed "conspiracy breaking" feature of DA markets: it is much more difficult for a small number of traders to exploit monopoly or oligopoly power in the continuous DA than in other institutions such as posted-offer markets or sealed-bid auctions.

In chapters 13 and 14 we start to see how the nice properties of DA markets can start to break down in more complicated environments were token valuations are not specified exogenously by the experimenter but are determined endogenously and must be inferred by traders over the course of the trading process. In chapter 13 Colin Camerer and Keith Weigelt study a stochastically lived asset that pays a dividend each period. The probability that the asset survives from period t to $t+1$ plays the role of a discount factor, so their framework can be regarded as a experimental re-creation of a stock market. A security's "fundamental value" in this framework is the discounted value of its dividends. However, the possibility of speculation and bubbles can lead traders to value the stock at more than its fundamental value in hopes of reaping capital gains. Price bubbles have been observed in previous stock market experiments such as Smith, Suchanek, and Williams.[18] Camerer and Weigelt also observe a tendency of the market prices to diverge from fundamental values, but conclude that "The double auction does not fail completely at generating convergence to competitive equilibrium prices for stochastically lived assets—convergence occurs with experienced traders—but the auction mechanism performs much more slowly and erratically than in simpler settings" (p. 389). In chapter 14 John O'Brien and Sanjay Srivastava use the DA market to study the of role of arbitrage in setting prices in an experimental option market. They show that while simple arbitrage relations succeed in determining market prices when there is no bid-ask spread, arbitrage by itself is not sufficient once we recognize that DA markets typically have a positive bid-ask spread. While they rarely find traders foregoing clear arbitrage opportunities, they conclude that "market dynamics *do* *not* appear to be characterized by arbitrage opportunities disappearing over the time interval of our typical experiment" (p. 418).

We organized the book into three parts. The first part, which includes this preface as well as chapters 1 and 2, surveys current knowledge and practice. The second part, chapters 3 through 5, offers new contributions on the underlying theory. The third and largest part examines evidence from the laboratory and computer simulations. (As the Kagel-Vogt and Satterthwaite-Williams chapters illustrate, the separation of theory from evidence is artificial, but some arbitrary ordering is inescapable.) Taken as a whole, we think these contributions have significantly improved our understanding of the DA and related market institutions, and illustrate innovative and interdisciplinary methods that promise further returns. However, some key questions remain unresolved and many new questions arise from the work presented here. Our hope is that this book will stimulate further research into some of these deep unsolved questions.

We have several debts to acknowledge. This book and much of the research it contains would not have been possible without the active support of the Santa Fe Institute. We are particularly grateful to directors of the Santa Fe Institute's economics program, Brian Arthur, John Geanakoplos, and David Lane, for providing funding and helping to organize the first and second DA conferences. Other participants include Marek Fludzinski, Pete Kyle, Warren Langley, Jean Lequarre, Mark Olsen, Richard Palmer, and Skip Sorensen; their thoughtful comments indirectly contributed to the work of this volume. Finally, we very much appreciate the expert assistance of SFI staff on this project including Ronda Butler-Villa, Andi Sutherland, and Della Ullibarri who helped organize the conferences and typeset this volume.

Dan Friedman
University of Californa
Santa Cruz, CA 95064

John Rust
University of Wisconsin
Madison, WI 53706

August 7, 1992

REFERENCES

1. Arrow, K., and G. Debreu. "Existence of an Equilibrium for a Competitive Economy." *Econometrica* **22** (1954): 265–290.
2. Arrow, K., and F. Hahn. *General Competitive Analysis*. San Fransico: Holden-Day, 1971.
3. Binmore, K. *Essays on the Foundations of Game Theory*. Oxford: Basil Blackwell, 1991.
4. Carney, T. F. *The Economies of Antiquity*. Lawrence, KS: Coronado Press, 1971.
5. Forrest, S. *Emergent Computation*. Cambridge, MA: MIT Press, 1992.
6. Friedman, D. "A Simple, Testable Model of Double Auction Markets." *J. Econ. Behav. Org.* (1991): 47–50.
7. Hayek, F. A. "The Use of Knowledge in Society." *Amer. Econ. Rev.* **34(4)** (1945): 519–530.
8. Hodges, R. *Primitive and Peasant Markets*. Oxford: Basil Blackwell, 1988.
9. Hurwicz, L. "On Informationally Decentralized Systems." In *Decision and Organization*, edited by R. Radner and B. McGuire, 297–336. North Holland: Amsterdam, 1972.
10. Lévy, J. *The Economic Life of the Ancient World*. University of Chicago Press, 1967.
11. Mitchie, R. C. *The London and New York Stock Exchanges 1850–1914*. London: Allen & Unwin, 1987.
12. Myerson, R. "Optimal Auction Design." *Math. Oper. Resh.* **6** (1981): 58–73.
13. Myerson, R. *Game Theory: Analysis of Conflict*. Cambridge: Harvard University Press, 1992.
14. Plott, C. R. "Industrial Organization Theory and Experimental Economics." *J. Econ. Lit.* **20** (1982): 1485–1527.
15. Roberts, J., and A. Postlewaite. "The Incentives for Price-Taking Behavior in Large Exchange Economies." *Econometrica* **44** (1976): 115–1128.
16. Reader, W. J. *A House in the City*. London: B. T. Batsford, 1979.
17. Smith, V. L. "Microeconomic Systems as an Experimental Science." *Amer. Econ. Rev.* **72** (1982): 923–955.
18. Smith, V. L., G. L. Suchanek, and A. Williams. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica* **56** (1988): 1119–1151.
19. Wilson, R. "Efficient Trading." In *Issues on Contemporary Microeconomics and Welfare*, edited by G. Feiwell, 169–208. New York: MacMillan, 1985.
20. Wilson, R. "On Equilibria of Bid-Ask Markets." In *Arrow and the Ascent of Modern Economic Theory*, edited by G. Feiwell, 375–414. New York: MacMillan, 1987.
21. Wilson, R. "Game Theoretic Analysis of Trading." In *Advances in Economic Theory: Fifth World Congress*, edited by T. F. Bewley. Econometric Society Monographs. Cambridge: Cambridge University Press, 1987.

THE DOUBLE AUCTION MARKET INSTITUTIONS, THEORIES, AND EVIDENCE

I. Institutions

Daniel Friedman
Economics Department, University of California, Santa Cruz, California 95064;
e-mail:dan@cash.ucsc.edu

ONE
The Double Auction Market Institution: A Survey

1. INTRODUCTION

Trade in actual markets (in the field or in the laboratory) is governed by a *market institution*, that is, a set of rules specifying which sorts of bids and other messages are legitimate, and how and when specific traders transact, given their chosen messages. The continuous double auction (DA) is an important type of market institution. It allows traders to make offers to buy or sell and to accept other traders' offers at any moment during a trading period.

Interest in the DA institution arises from two facts:

1. For more than 100 years, trade in the most important field markets for homogeneous goods has been governed primarily by DA rules. For example, the New York Stock Exchange and the major Chicago exchanges all use versions of the DA institution, and the traded goods include stocks, bonds, agricultural commodities, metals, and derivative securities. Laboratory markets also have relied primarily on variants of the DA since Smith.[79]

The Double Auction Market, Eds. D. Friedman and J. Rust,
SFI Studies in the Sciences of Complexity,
Proc. Vol. XIV, Addison-Wesley, 1993

2. The laboratory experiments demonstrate that even with rather small numbers of traders who have very imperfect information on supply and demand, the DA institution consistently produces very efficient allocations and prices, much more so than traditional theory would suggest. Smith[83] calls this finding a "scientific mystery." The mystery remains unsolved to this day, despite the central importance of the issues it raises for theorists and practitioners.

The purpose of this chapter is to survey recent developments in academic research which may help us better understand the double auction institution. (The practitioner's perspective is equally important and will be emphasized in the next chapter.) Here I report work by theorists and by experimentalists which addresses questions such as:

- How do outcomes in laboratory DA markets compare to relevant theoretical benchmarks?
- How does the DA institution fare in different laboratory environments and in different field environments?
- How sensitive are the results to apparently minor variations in the DA institution?
- How does the DA compare to alternative market institutions?

Section 2 standardizes terminology, places the DA institution and its variants in the taxonomy of market institutions, and distinguishes the different sorts of trading environments. Section 3 surveys several strands of the theoretical literature relevant to the DA. Section 4 surveys some of the relevant empirical literature. Concluding remarks are offered in Section 5.

2. MARKET INSTITUTIONS AND MARKET ENVIRONMENTS

Unfortunately no standard terminology for market institutions has yet emerged in the literature. For example, the DA institution sometimes is called a double oral auction or a bid-ask market, while a quite different institution (referred to below as the CH) is sometimes called a call market and sometimes called a (static) double auction! Sometimes a market institution is called dynamic because it is repeated over time, and sometimes because it allows traders to chose the timing of their messages continuously. To avoid confusion, I begin by standardizing my terminology.

Institutions I consider deal with at least two agents, called *traders*, and at least two commodities allocated to (owned by) the traders. Unless otherwise noted, I assume exactly two commodities, one of which is called the *good* and the other of which is called *money*. Money is the numeraire and is assumed divisible. *Exchange* refers to any non-coercive process whereby traders alter the allocation of commodities, total quantities of all commodities remaining constant (i.e., abstracting from production and consumption). A *market institution* defines an exchange process by specifying the set of admissible *messages* (i.e., traders' actions, usually price and/or quantity offers), and by specifying a final commodity allocation given any combination of messages chosen by the traders and any initial allocation.

A *net trade* vector is the difference between a final and an initial allocation. A *bilateral transaction* is a minimal net trade, involving non-zero components of money and the good for two traders. For any trader whose net trade component for the good is non-zero, the *transaction price* is the absolute value of her money component divided by her good component. If the good component is positive (negative), I refer to her as a *buyer* (*seller*). For example, if trader A's allocation of the good increases by 2 units and his allocation of money (in dollars) decreases by $1.50 while trader B's allocation of the good decreases by 2 and her money increases by $1.50, other traders' allocations constant, then we have a 2-unit bilateral transaction between buyer A and seller B at transaction price $0.75. Sometimes the institution specializes all traders, either as buyers who can never sell or as sellers who can never buy; I refer to this as a case of *one-way traders.*

2.1 A PARTIAL TAXONOMY OF MARKET INSTITUTIONS

Bilateral search processes are minimal sorts of market institutions in which traders seek partners for mutually beneficial bilateral transactions. Of greater interest for present purposes are *unified* or consolidated market institutions which provide more information on trading opportunities and preclude simultaneous transactions at different prices. A (non-discriminatory) *auction* is a unified market institution in which traders' messages include an offered price (called a *bid* for an offer to buy and an *ask* for an offer to sell), and which gives higher priority in transactions to better offers (higher bids and lower asks). An auction is *one-sided* if only bids or only asks are permitted, and *two-sided* if both are permitted. Institutions with most but not all features of an auction will be called *quasi-auctions.*

For example, the (sellers') *posted offer* institution, in which traders simultaneously announce ask prices and then each trader can choose quantity in a bilateral transaction at his partner's announced ask price, is a one-sided quasi-auction; it is not a proper auction because it does not guarantee priority to lower asks. The primary market for U.S. Treasury securitites is conducted (at least through 1991–reform is under serious consideration) as a one-sided quasi-auction; it is not a proper auction, according to present terminology, because successful bidders can, and do, buy simultaneously at different prices.

Auctions (and other trading institutions) can be either *one-shot* or *repeated.* A repeated auction consists of several *trading periods*: agents receive new initial allocations at the beginning of each trading period, presumably engaging in production and consumption between trading periods.

In a *discrete-time* auction, all traders move in a single step from initial allocation to final allocation. By contrast, a *continuous-time* auction permits exchange at any moment during a trading period, and the overall net trade typically is composed of many bilateral transactions. The clearinghouse (CH) is the prime example of a discrete-time two-sided auction. Its key feature is that bid and ask messages are collected or "batched" during the trading period, and then "cleared" at the end of the period. Given the supply and demand revealed in the messages, a maximal trade vector is selected subject to a unified price constraint.

The focus in this chapter is the basic *continuous double auction* (DA). It is a continuous-time, two-sided auction in which messages consist of bids and asks for single units of the good, and of acceptances of the current best bid or ask. Net trades consist of the bilateral transactions triggered by an acceptance of the best bid or ask. Messages are admissible whenever the resulting transaction would not violate any specified non-negativity constraint on traders' money or good allocation.

2.2 DA VARIETIES AND HYBRID INSTITUTIONS

Many variants of the basic continuous double auction have been employed in field and laboratory environments. For example, inferior offers (e.g., bids lower than the current best bid) may or may not be made public or even be admissible. The identity of traders making offers may or may not be revealed. Bids and asks may be for stated quantities and not necessarily for a single unit. Numerous details must be specified; e.g., are unaccepted offers queued or must they be renewed following displacement by (and/or acceptance of) a better offer? Laboratory studies suggest that DA market performance can be affected by whether traders are one- or two-way, and whether the institution implemented on a computer or as oral "open outcry."

Computerized implementations of the DA are rapidly displacing oral implementations in the laboratory primarily because of the more direct control they offer over traders' information and their greater efficiency in data capture and data analysis. The main disadvantage (besides higher laboratory set-up costs) is that subjects seem to adapt to the environment more slowly in the computerized DA.[90] As Domowitz shows in the next chapter, there is also a trend towards the computerized DA in field markets, spurred by the promise of broader service at lower transactions cost and a better audit trail, but the trend is strongly resisted by traders whose "open outcry" skills would lose value in a computerized DA.

One can imagine auction market institutions with some, but not all, of the DA features. Several such hybrid institutions recently have been tested in the laboratory. In some respects a DA is like a long series of very short period repeated CH markets, most of which have zero transactions, and almost none with two or more. This observation (plus some arbitrary tie-breaking conventions, etc.) leads fairly directly to the "synchronized double auction" institution used in the double auction tournament hosted by the Santa Fe Institute, reported in chapter 6 of this volume. From another perspective, the key aspect of the DA is the continuous information on the "order flow" (traders' offers or messages). The Uniform Price Double Auction (UPDA) institution (presented in Chapter 11) also provides such information, allows recontracting (improvement of orders) during the trading period, and ends in a single unified price for all transactions. Friedman[26] reports versions of the CH which allow continuous order flow (or "order book") information. The UPDA institution is conceptually equivalent to some versions although there are some differences in implementation.

At the University of Arizona, work is also underway on a series of hybrid institutions which feature a "price clock" reminiscent of the clock used in Dutch flower auctions.[10] For example, in the Double Dutch institution,[55] a buyer price clock starts at a prohibitively high price and ticks down until a trader stops it, thereby purchasing a unit at a price at least as favorable to him as the price displayed on the clock. At this point, a seller price clock ticks upward from a very low initial price until stopped by a trader who thus sells a unit. Then the buyer clock resumes its descent until stopped again, etc. The trading period is over when the buyer clock price crosses the seller clock price, at that point all purchases and sales are transacted at the crossing price.

One observes a bewildering variety of market institutions in the field. In well-developed markets for homogeneous commodities and financial assets, variants of the DA seem most prevalent and CH variants are not unusual. DA field markets appear to differ largely in their assignment of specialized trader roles. Few, if any, field DAs have one-way traders, but most give some traders *privileges* such as a bigger message space and/or better access to other traders' messages than available to unprivileged traders. For example, the New York Stock Exchange (NYSE) offers at least three levels of privilege: for each good (traded security), there is a single "specialist" trader who alone has immediate access to all bids and asks (limit orders); there are "floor traders" who can periodically check the current list of unaccepted bids and asks ("the specialist's order book"); and there are unprivileged traders ("the public") with slightly delayed access only to the current best bid and ask. At the major Chicago exchanges, full participants in the DA (the "pit" traders) are usually agents for unprivileged traders who do not have direct access to current offers on the trading floor. Privileges are typically transferrable property rights—you can "buy a seat" at an exchange—and these property rights may help cope with the public-goods problem inherent in maintaining an organized market. See Friedman[29] for further discussion.

Field markets sometimes combine different institutions over time. For example, the NYSE and many financial markets begin trading (to start a new day or after a suspension of trade) with a CH and continue until the close with a DA. The newly launched Wunsch auction for after-hours trading in NYSE-listed securities is a hybid institution—in essence an UPDA or CH variant. See chapter 2 of Schwartz[75] and chapter 2 of the present volume for a more extended discussions of current financial market institutions.

The fact that so many DA variants and hybrids coexist in the field might suggest to some economists that market efficiency is insensitive to institutional details, because otherwise all markets would have adopted the most efficient variant. An equally plausible explanation with very different implications is that efficiency is very sensitive to institutional details, but also is very sensitive to environmental

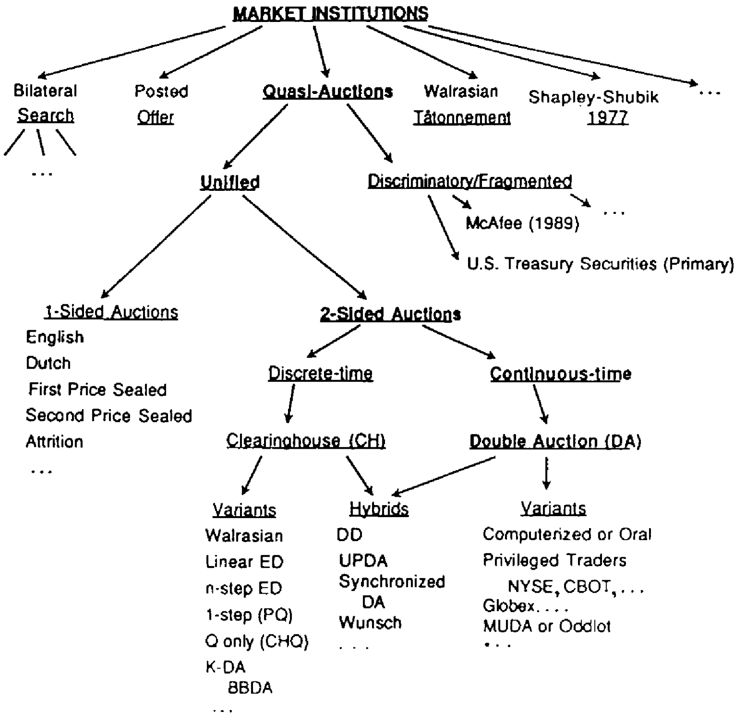


FIGURE 1 A Family Tree for the Double Auction.

TABLE 1 Some Trading Environments[1]

1	Goods	2 or >2
2	Traders	One-way or *two-way* 2 (bargaining) or *# buyers, # sellers* ≥ 2 or *$n \geq 3$ traders* or "large numbers"
3	Endowments	*single* or multiple units divisible or *indivisible* units (tokens)
4	Preferences	*reservation prices* (value or cost) for indivisible units, or smooth classical preferences for divisible units
5	Information	complete symmetric incomplete: private values, common values, affiliated values asymmetric ("insiders,"...)

[1] Note: Italics indicate the defaults; e.g., unless otherwise mentioned, 2 goods and at least 3 two-way traders are assumed, and efficiency refers to *ex post* (no remaining gains from trade).

details (e.g., to the particular type of good traded, the national payments system, etc.), and the efficient variant is different in different circumstances. Another alternative explanation is that liquidity considerations induce strong increasing returns to market size ("order flow attracts order flow") so that the first institution to become viable in some market will become entrenched. In this case, the relative efficiency and the prevalence of an institution need not be closely linked.[4]

Evidently informal theorizing is insufficient to understand the prevalence of the DA and its variants. Therefore, I turn next to the relevant formal theory and empirical research. The literature is quite diverse in its maintained assumptions as well as in its terminology, but Table 1 and Figure 1 may provide some guidance. Table 1 lists the main dimensions of the assumed environment; italicized entries indicate the most commonly encountered cases. Figure 1 indicates the relation of the DA and its variants to most other market institutions I will discuss.

3. THEORETICAL APPROACHES TO THE DOUBLE AUCTION

3.1 WALRASIAN TRADITIONS

Leon Walras' discussion of price formation in 1874 was the first influential treatment of a market institution. In the modern theoretical literature, one can distinguish two different interpretations of Walras' institution. The tâtonnement interpretation regards the institution as a repeated quasi-auction in which some disinterested agent (or automaton) called the Walrasian auctioneer sends a price message, and all traders respond with quantity messages specifying their desired net trades at the given price. The auctioneer adjusts the price according to some algorithm (the tâtonnement or "groping") until the quantities sum to zero. At this point, traders announced net trades are transacted at the "market clearing" price. Although inspired by the trading arrangements on the Paris bourse of Walras' day, this tâtonnement institution is now very rare in the field; to the best of my knowledge, it resembles only the procedures used for many years in the London daily gold price "fixing" and in some Japanese wholesale rice markets.[73] The institution has occasionally been studied in the laboratory[6] with mixed results. Arrow and Hahn[3] summarize an extensive (and largely negative) literature regarding the stability properties of this institution under the assumption that traders are numerous and are non-strategic pricetakers. One gets the impression that the theoretical properties of the tâtonnement institution are even worse when strategic behavior is allowed, but relevant theoretical studies are rare.

A second interpretation, more commonly used in the literature but less often explicitly discussed, is that the Walrasian institution is really a one-shot clearinghouse. Traders' messages are their announced excess demand correspondences (or just their announced demand or supply curves), and the final allocation arises from maximal net trade at a single market-clearing price. Perhaps because of its familiarity and because of its "direct revelation" nature, this Walrasian CH institution seems very natural to most theorists—so natural indeed, that work employing it is often called "institution free"! Unfortunately, truthtelling in the Walrasian CH is not a Nash equilibrium strategy, except in the "large numbers" limit in which each trader's transactions are always a negligible fraction of aggregate transactions; it follows that the institition is theoretically inefficient.[69] Its presumed inefficiency together with its huge information requirements (the dimension of each trader's message space potentially is infinite) probably account for the non-existence of the Walrasian CH in field environments. Only greatly simplified versions have been tested in the laboratory. (For example, McCabe et al.[53] restrict excess demand messages to two-parameter linear functions.) Still, the large numbers limit of the Walrasian CH institution defines the benchmark of perfect competition in mainstream theoretical literature.

3.2 GAMES OF INCOMPLETE INFORMATION

Perhaps the most natural way for a modern theorist to think about a market institution is as a game of incomplete information, since traders generally do not know each others' preferences (e.g., sellers' costs and buyers' true valuations). Vickrey[88] pioneered this approach in his classic analysis of auctions. He considers a first-price sealed-bid auction, a one-sided one-shot buyer's auction for a single indivisible unit. Traders' messages are simultaneously chosen bids, and the resulting net trade is a bilateral transaction for the unit between the (passive) seller and the highest bidder at her bid price.

Vickrey's analysis assumes each trader i has prior knowledge regarding (a) the IID uniform $[0,1]$ distribution of reservation prices ν_j and (b) the bid function $b(\nu_j) = (n-1)\nu_j/n$ of the other $n-1$ traders. He verifies that a risk-neutral trader with private knowledge of her own reservation price ν_i will maximize expected utility by bidding $b(\nu_i)$, i.e., by using the same bid function. Vickrey's analysis anticipates *Bayesian Nash equilibrium* (BNE), the general solution concept for games of incomplete information introduced by Harsanyi.[38] In BNE, players use Bayes theorem to form expectations given individual private signals and given common prior information regarding (a) the distribution of players' (buyers') relevant characteristics ("type") and (b) players', (buyer's) type-contingent strategies. Relative to these expectations each player's (buyers') strategy maximizes expected utility; i.e., the strategies are simultaneous best responses.

As documented in recent surveys such as McAfee and McMillan[50] and Wilson,[92] the BNE approach has provided very satisfying analyses of one-sided one-shot auctions which extend Vickrey in allowing risk aversion, general distributions of reservation prices, and multiple units. Two other sorts of extensions require brief discussion at this juncture. First, Vickrey assumed *private values* in that each trader's private signal is precisely her own reservation value ν_i. One can also consider private signals which provide each trader useful but imperfect information regarding ν_i, and perhaps also regarding other traders' values ν_j. The most general useful case assumes a condition called *affiliated values.* Another very important special case, in some sense polar to private values, is called *common values*. Here all traders have the same true reservation value and receive independent, unbiased signals regarding that uncertain value. See the surveys cited above for more precise definitions and results.

The second sort of extension is to different auction institutions. Vickrey himself discussed at least four institutions: single unit, first- and second-price sealed-bid auctions, and English (ascending) and Dutch (descending) continuous auctions. The literature now covers many other one-sided auction institutions, not all of which yield efficient outcomes even *ex ante*, e.g., "the war of attrition."[68] Some that are *ex ante* efficient have surprising transactions patterns, e.g., the one-sided multi-unit Dutch seller's auction.[7]

The extension of the Bayesian Nash equilibrium (BNE) approach to two-sided auctions begins with Chatterjee and Samuelson,[12] who consider the bilateral monopoly (1 buyer + 1 seller) single-unit case with private values drawn independently from known uniform distributions. The institution is a CH with the transaction price at the midpoint of the interval of market-clearing prices when that interval is non-empty; otherwise, no transaction occurs. They find linear BNE bidding strategies which miss mutually beneficial transactions with probability $1/6$.

A series of papers by Satterthwaite and Williams, summarized in chapter 4 of this volume, extend the results to an environment of one-way traders (usually m sellers and m buyers) with independent private values for single indivisible units. They first analyzed a version of the CH institution, called BBDA, which takes the upper endpoint of the interval of market-clearing prices as the transaction price in order to give one side, the sellers, a dominant truthtelling strategy. Later papers analyze versions of the CH, called k-DA's, which use other points along the interval of market-clearing prices. They show that in BNE the difference between buyers' bids and true values is $O(1/m)$ and foregone gains from trade are $O(1/m^2)$, so *ex post* inefficiency vanishes reasonably fast as the market gets larger. See also McAfee[51] for a slightly stronger result concerning two-sided quasi-auctions which give lower transaction prices to sellers than to buyers.

Incomplete information game theory indicates that some *ex post* inefficiency is inevitable. For example, Myerson and Satterthwaite,[59] and Makowski and Ostroy[48] use the revelation principle to show that BNE of *any* one-shot market institution cannot be both individually rational and *ex post* efficient for a finite number of players with general preferences. Since a dynamic trading process may gradually disseminate private information, it is unclear whether these inefficiency results apply to a continuous trading institution such as the DA. See Myerson[60] for an explanation of some difficulties in applying the revelation principle to extensive form games with communication.

Wilson[91] is the only serious attempt of which I am aware to analyze the basic DA (or any other continuous auction) as a game of incomplete information. He proposes a strategy selection in which, roughly speaking, traders play a waiting game for making serious bids and asks, with each buyer's (seller's) impatience arising from possible preemption of gains by other buyers (sellers). Once a serious bid (ask) is made, that buyer (seller), in essence, conducts a Dutch auction until a seller (buyer) accepts. The waiting game resumes after the resulting transaction. These messages fully reveal traders' private values and produce final allocations that are nearly *ex post* efficient—at worst, a few of the least valuable trades are missed. Wilson verifies that the proposed strategy selection satisfies the necessary conditions for BNE.

3.3 PROBLEMS WITH THE INCOMPLETE INFORMATION APPROACH

Modeling the DA as a game of incomplete information is natural, since in both laboratory and field applications, traders do not know each others' reservation prices, and often do not even know their own final valuations (e.g., in markets for risky assets). Moreover, this approach has been very fruitful in analyzing one-sided auction institutions. Nevertheless, it may be appropriate to note some problems in applying it to the DA, and to suggest some alternative approaches.

On the theoretical side, the incomplete information approach relies heavily on prior common-knowledge assumptions. Even very simple auctions involve type-contingent strategies—your bid in a first-price auction depends on your reservation price. In continuous auctions, such as the DA, one must also take history-contingent and time-contingent actions—you care about the sequence of bids, asks, and acceptances observed so far as well as about the time remaining when you decide on your next bid, ask, or acceptance message. In BNE each trader in a DA must have prior knowledge of other traders' type-contingent strategies (themselves time- and history-contingent messages) and must be able to compute expected utility-maximizing messages at every moment as the DA unfolds. Such prior knowledge and computational ability is literally incredible. As Vernon Smith has remarked in private correspondence, this approach pushes the real action off stage: one leaves unformalized the acquisition of prior knowledge and computational algorithms. It is possible to model information acquisition itself as a game of incomplete information, but this seems only to compound the problem.[24]

On the empirical side, the evidence so far does not seem favorable. As I will note in the next section, DA outcomes in the laboratory seem quite insensitive to the number of traders beyond a minimal two or three active buyers and two or three active sellers. Moreover, as discussed at length in Friedman and Ostroy[27] parameter choices which, according to an incomplete information analysis, should greatly reduce efficiency in CH (and presumably in DA) markets had no such effect in recent laboratory tests. This empirical issue is discussed more broadly in McCabe, Rassenti, and Smith.[52]

3.4 ALTERNATIVE APPROACHES

Models of trading institutions as games of incomplete information have dominated the recent theoretical literature, but there are alternative approaches. In no particular order, I will review the market microstructure literature, some other literature based primarily on decision theory (i.e., ignoring strategic interaction), and some models using games of complete information.

A large body of finance literature studies two-sided markets for a common-value good. Unfortunately, most of it assumes a Walrasian CH and "large numbers"(e.g., Grossman and Stiglitz[36]). Other market institutions are studied in the market microstructure literature, so named by Garman.[32] The classic microstructure articles largely ignore game-theoretic considerations, but the more recent articles often implictly or explicitly model games of incomplete information. I am unaware of any survey of this literature more recent than Cohen et al.[13]

The market microstructure literature begins with Stigler[86] and Demsetz,[17] who use a simple utility-maximizing model to analyze a monopoly specialist DA institution: only one trader can send bids and asks, the other traders being restricted to acceptances. Recent analyses of this institution, in Copland and Galai[14] and Glosten and Milgrom,[34] impose a zero-profit condition on (competitive) specialists, and study how the bid-ask spread in BNE responds to the presence of traders with superior information and to liquidity-motivated ("noise") traders. In essentially the same setting, Kyle[43] finds an optimal strategy for a trader with sole access to superior information, and shows that the strategy gradually but profitably disseminates the information.

Ho and Stoll[39] study the DA institution using dynamic programming techniques. They seek to characterize equilibrium bid-ask spreads as a function of order-unit size, risk tolerance, and uncertainty. They argue that the bid-ask spread is essentially independent of the number n of active traders as long as $n \geq 2$; the intuition is that the best price wins in a form of Bertrand competition which requires only two traders.

The microstructure literature also considers CH institutions. For example, Mendelson[56] analyzes the performance of a repeated CH market given an exogenous stochastic flow of offers. More recently, Kyle[44] adopts the game of incomplete information framework to study the effect of inside information in a one-shot CH market. Under his parametric assumptions, traders with heterogeneous superior information (i.e., less noisy signals about the common value of the good) optimally will submit linear excess demand functions which in BNE will reveal some but not all of their information.

Apart from an obscure precursor (Garcia[31]), Easley and Ledyard[20] were the first theorists to attempt to explain the efficiency of laboratory DA markets. They emphasize convergence to competitive equilibrium across successive trading periods in a DA experiment. Easley and Ledyard postulate plausible behavioral rules on how specialized traders set and adjust "reservation prices" (the links to true value/cost parameters left unspecified) and bids or asks across and within trading periods. They also test three implications of the model against some laboratory data, with favorable results. The final version of their paper appears as chapter 3 of this volume.

A recent decision-theoretic approach to laboratory DA markets, focussing on behavior within a single trading period, appears in Friedman.[25] A special case of the NCE model discussed below, it makes the strong simplifying assumption that traders disregard the impact that their current bid or ask may have on other traders subsequent offers. Thus traders act as if they are playing a game against nature, but otherwise are good Bayesian decision makers. Such traders are shown to engage in Bertrand competition with respect to reservation prices which decay to true cost and value as trading time runs out. Therefore final outcomes are nearly 100% efficient.

Perhaps the most obvious alternative to the incomplete information approach is to regard the DA and other auction institutions as games of complete information. This terminology need not be taken literally; it is only in the last 20 years or so that theorists have supposed that ordinary NE requires each agent to know (as common knowledge) the preferences of all agents. Luce and Raiffa[47] and probably most other theorists of that time preferred to regard NE as the rest point of some unformalized groping process. Of course, theorists will require a satisfactory formal model of the process before accepting this interpretation again, but it is possible that current work (e.g., Fudenberg and Kreps,[30] Gilboa and Matsui,[33] and Friedman[28]) could lead to such a model. In the meantime, one can at least entertain the thought that stationary repetition in the laboratory, or long experience in the field, may be a practical substitute for direct knowledge of other traders' preferences and strategies and, therefore, might lead traders to choose messages which constitute an NE in the complete information game.

Shapley and Shubik[76] were probably the first to use explicit complete information game-theoretic models for exchange institutions. Like Roberts and Postlewaite,[69] they find that only in the large numbers limit do NE strategies produce efficient outcomes for their (non-auction) institution. Much different theoretical results are reported in Dubey,[19] Simon,[77] and Benassy[5] for multi-commodity CH auctions and in Friedman and Ostroy[27] for simple CH auctions. Here each trader's message consists of a single price-quantity (PQ) pair. (Recall that Satterthwaite and Williams[72] assume price-only messages for single indivisible units, so their results do not apply.) In the PQ version of the CH institution, we have the surprising result that "active" NE (at least two buyers and two sellers) coincide with Walrasian equilibria. (See also Schmeidler.[74]) Thus small numbers are compatible with 100% efficiency in this institution. Intuitively, we have a form of Bertrand competition.

I am not aware of any satisfactory model of the continuous DA as a game of complete information. Friedman[24] contains a partial model, where NE for the extensive game is replaced by a renegotiation-proofness concept called NCE (for no-congestion equilibrium; the idea is that traders are not failing to transact at the close of trade because they all waited to the last second and the market got congested). The main result is that if strategies satisfy NCE, then the DA will produce (a) 100% efficient outcomes in smooth economies (i.e., divisible goods and classical preferences) with at least three traders, and (b) outcomes just one (least valuable) trade shy of 100% efficiency in economies with indivisible units and at least two "buyers" and two "sellers." Again, it is a form of Bertrand competition (plus time pressure) that is responsible for small-numbers efficiency.

My conjecture that games of complete information will be useful in solving Smith's "scientific mystery" now can be broken into two parts. First, that competitive (Walrasian) equilibrium coincides with ordinary (complete information) NE in interesting environments for the DA institution. Second, that the DA promotes some plausible sort of learning process which eventually guides both clever and not-so-clever traders to behavior which constitutes an "as-if" complete-information NE. The work cited in the previous three paragraphs lends some plausibility to these conjectures but does not begin to prove them.

4. EMPIRICAL WORK

In studying double auction markets, empirical researchers can tap three general data sources: field data from large-scale ongoing markets, laboratory data from small-scale DA and comparison markets, and computer simulation data. The strengths of field data are its availability and its relevance. Its main weakness is that unobserved and/or uncontrolled variables may preclude valid inferences.[45] Laboratory data permits control of many variables and renders others observable, but it can be expensive (in time and money) to acquire. Computer simulations now are relatively cheap and are completely controllable and observable; e.g., the researcher can confidently detect a mixed strategy in a computer simulation but not in laboratory data. The main disadvantage is that traders' strategies are not endogenously chosen, but rather must be specified exogenously in a computer simulation.

A moment's reflection discloses strong complementarities among the three data sources. This section, and the volume as a whole, will draw on all three. Here I emphasize comparisons of DA market outcomes to theoretical benchmarks and to outcomes under alternative market institutions. For these purposes, it turns out that the laboratory data are most informative, so my treatments of field data and computer simulation data will be a bit cursory.

4.1 FIELD DATA

The primary theoretical benchmark for the DA (or for any market institution) is competitive equilibrium; to what extent do actual allocations and transactions prices resemble competitive equilibrium allocations and prices? When some traders have private information, the efficient benchmark becomes fully revealing rational expectations equilibrium (FRREE): to what extent do actual allocations and transactions prices resemble those of a competitive equilibrium in which all private information was made public?

Preferences and private information are not observable in field data, so these questions can't be answered directly. The largest body of empirical field literature follows Fama[22] in testing indirect consequences of FRREE in asset markets: do (suitably adjusted) transactions prices follow a martingale? Do they immediately and fully respond to new information? LeRoy[46] provides a skeptical survey of martingale tests and concludes that, although most previous studies reached favorable conclusions, the support for martingales (relative to interesting alternatives) actually is quite weak. The arrival of private information precludes direct answers to the second question, but most of the numerous event studies following Fama et al.[21] seem favorable. A more controversial approach to this question is to compare transaction price volatility to price volatility in some appropriate FRREE model—again see LeRoy,[46] for a recent account. It may also be worth mentioning that O'Brien and Srivastava[61] conclude that their laboratory asset markets are quite inefficient even though they pass the standard martingale and volatility tests.

Field comparisons of market institutions are much less common than efficiency studies. For example, Cohen et al.[13] compare two DA markets with monopoly "specialists" (NYSE and AMEX) to two non-specialist DA markets in Tokyo and Rio de Janeiro. Their results are consistent with reduced price volatility in specialist markets for thinly traded issues. Amihud and Mendelson[2] point out that conclusions are ambiguous in this sort of study because it is "...hard to discern differences [in market performance] resulting from the trading mechanism itself from differences due to dissimilarities of securities and environments." They try to finesse the problem by comparing NYSE close-to-close price changes with NYSE open-to-open changes, noting that opening price is set in a CH market. They find that close-to-close returns have greater variance and kurtosis, and greater deviations from a random walk in ARMA (1,1) estimates. Stoll and Whaley[87] reach similar conclusions in a more recent and thorough study of the NYSE data. Neither study considers the alternative hypothesis that, given the difference in information conditions between opening and mid-day, the CH institution was chosen to reduce returns variance at opening which otherwise might be even greater. In the absence of a clean institutional comparison, one must make a leap of faith to attribute measured differences market performance to differences in the market institution. Field data unfortunately rarely permit such clean comparisons.

4.2 EFFICIENCY IN LABORATORY MARKETS

Laboratory data are especially useful for present purposes because the experimenter can observe private information and preferences and can control the market institution. Consequently, one can directly test efficiency and directly compare market institutions. The small scale and simplicity of laboratory markets suggests some caution in generalizing results to large, complex field markets, but small scale and simplicity are actually advantages in testing explicit theory.[64,83]

Chamberlin[11] is the first published report of (not very well controlled) laboratory market experiments. He induced cost and value parameters for one-way traders of single indivisible units, and employed a bilateral search institution with public announcement of transaction prices. His outcomes were not very efficient. Smith,[79] by contrast, reported highly efficient outcomes in a set of otherwise similar repeated DA experiments. Since then, numerous other studies have abundantly confirmed what Smith[82] calls the Hayek Hypothesis: "Strict privacy [i.e., induced private values] together with the trading rules of a market institution [viz., the DA repeated several times] suffice to produce competitive outcomes at or near 100% efficiency" (page 167). This result even seems to hold with as few as two or three buyers and two or three sellers[64,83] and even when very unfavorable value parameters are chosen, such as the box and swastika designs in Smith and Williams.[85]

Asset-market experiments as introduced by Plott and Sunder,[65] and Forsythe, Palfrey, and Plott[23] feature DA markets with two-way traders and a good (the "asset") whose marginal value is constant for each trader and state. In these experiments, states are uncertain, private information is present, and/or the asset pays a dividend over several trading periods. Traders' endowments and (more importantly) contingent valuations and information typically differ across trader types, an environment of affiliated values rather than private values. In some asset-market experiments, there is only one trader type, in which case we have common values and all allocations are vacuously *ex post* efficient.

The main question addressed in asset-market experiments is usually informational efficiency—to what extent do transactions prices approach the fully revealing rational expectations equilibrium (FRREE) benchmark? In DA experiments with assets that live only one period and two or more trader types, the results generally support FRREE. Even when traders have differential information or some have superior information, the asset price does a surprisingly good job of aggregating or disseminating the information. I should note that Camerer and Weigelt,[8] Plott and Sunder,[66] and Copeland and Friedman[15] report some interesting anomalies, but in my view the basic lesson is that DA prices are very informationally efficient in this environment.

The lesson from the DA asset-market experiments of Smith, Suchanek, and Williams[84] seems rather different. They report frequent large bubbles—episodes where transaction price rises well above the fundamental (the FRREE value) for an extended time period, usually ending in a sudden price crash to or below the fundamental. Besides using common values, their experiments differ from most other asset-market studies in that they use very long lived assets ("dividends" paid over 15 or 30 periods rather than the usual 1, 2, or 3) and much less stationary repetition (only 1 to 4 reinitializations with a given groups of traders, rather than the usual 5 to 20). Despite some useful follow-up work (e.g., Porter and Smith[67]), it is not yet clear whether the observed market inefficiencies should be attributed to common values, to learning and experience effects, or to other differences in design. Camerer and Weigelt present new results and a fuller discussion in chapter 13 of this volume.

A few remarks on one-sided auctions may also be in order. Large numbers of one-sided private values auctions have been conducted in the laboratory in the last ten years or so, e.g., Cox, Smith, and Walker.[16] My view is that the results are generally consistent with the incomplete information theory, although there are some apparent anomalies whose interpretation recently has become controversial.[37] Of greater relevance to present concerns are one-sided auction experiments with induced common values by Kagel and Levin.[42] Their outcomes are less consistent with BNE and they conclude that the winner's curse (overbidding) is a relatively persistent phenomenon which is eventually overcome by slow and environment-specific learning.

In chapter 10 of this volume, Kagel and Vogt extend the same basic setup—independent draws from known uniform distributions—to the BBDA and a variant. Again they find substantial deviations from BNE. Their results suggest (to me at least) that these institutions do not encourage rapid learning of equilibrium strategies.

4.3 COMPARISONS OF LABORATORY MARKET INSTITUTIONS

Comparisons of market institutions prior to 1980 established that the DA produced more efficient outcomes than many other trading institutions such as bilateral search, posted offer, and one-sided auctions (e.g., Plott and Smith[63]; see also Williams[90]). Smith et al.[81] report that a computerized DA generally produces more efficient final allocations and more rapid price convergence than various versions of the CH, except that allocations are perhaps a bit more efficient in a recontracting version of the CH which allows each trader to send multiple price-quantity messages. Equally important, most of the CH variants yield more efficient allocations than incomplete information game theory would suggest. Friedman and Ostroy[27] confirm these findings for divisible goods and very unfavorable "box like" parameters, and also report inefficient allocations under a modified CH institution which permits underrevelation in quantities but not in prices.

Institutional comparisons are currently a very active area of laboratory research. Preliminary results at Arizona suggest that the double Dutch institution is more efficient than the DA in some simple environments.[13] Friedman[26] reports that some hybrid institutions (continuous-information versions of the CH) have efficiencies comparable to the DA in several asset-market environments. In chapter 11 of this volume, McCabe, Rassenti, and Smith present evidence that their UPDA institution also achieves efficiencies comparable to the DA in one-way trader environments. The patterns of inefficiencies differ across the two institutions; in the UPDA one occasionally sees substantial unrealized gains when a block of traders underbids a bit too much. In chapter 12 of this volume, Plott and Clauser compare institutions in an unusual laboratory environment which encourages seller conspiracies. They find that a DA variant called ISMDA shares the efficiency (or conspiracy-thwarting) properties of the DA.

Institutional comparisons of this sort eventually should produce well-documented stylized facts about the absolute and comparative advantage of many market institutions in various environments. Such facts will be invaluable to theorists and to practitioners.

4.4 COMPUTER SIMULATIONS AND TRADING STRATEGIES IN THE DA

Some empirical studies look at individual behavior as well as market-wide outcomes. Benchmarks provided by computer simulations are particularly useful for this purpose. For example, in chapter 8 of this volume, Domowitz and Bollerslev use computer-generated bidding to take a first look at the effect of a DA variant and maximum length of the order book queue. In Chapter 6 of this volume, Rust, Miller, and Palmer, look at both individual behavior and market outcomes in a tournament of computer-simulated trading strategies in a hybrid DA-CH institution called the synchronized double auction. Follow-up work will allow direct comparisons of computer-simulated traders and human traders.

At a 1990 Santa Fe Institute workshop, Sunder introduced another performance benchmark which already has become quite influential. His zero-intelligence (ZI) algorithm prescribes random advantageous bids and asks. When all (simulated) traders employ the ZI algorithm in a DA market, the outcomes are surprisingly efficient, comparable to those achieved by human subjects in early trading periods. See chapter 7 of this volume by Gode and Sunder for some follow-up work which raises the question of whether DA efficiency is attributable to trader rationality or inherent in the market institution itself.

The theoretical models of Wilson[91] and Friedman[25] as well as the ZI algorithm provide alternative characterizations of individual trader behavior and, therefore, make distinct predictions regarding transactions price sequences, bid and ask sequences, transactions partners, and efficiency within a DA trading period. In chapter 9 of this volume, Cason and Friedman begin the task of comparing the predictions to experimental data. The results so far are mixed; for example, the Friedman[25] model best predicts the bid and ask sequences, and ZI best predicts the price sequences, at least for inexperienced traders.

5. CONCLUDING REMARKS

The rest of this volume documents the acceleration of research in DA markets in recent years. Theoretical and empirical researchers both have recently developed promising new approaches. Evidence accumulates on which features of the DA are consequential and which are not, and which sorts of environments are conducive to efficient outcomes and which are not. Yet Smith's mystery remains: why is the DA (and some of its hybrids and variants) so efficient in most environments?

One can now distinguish three main lines of attack, in some respects complementary and in other respects rival.

1. The BNE approach extends the Vickrey-Harsanyi modeling strategy to increasingly complex market institutions, and tests the extensions empirically as they become available. This approach influences virtually all chapters in this volume and dominates chapters 4, 5, and 10.

2. The Nash equilibrium learning approach regards trader behavior as changing in response to experience with a market institution, and eventually settling down (if at all) in an as-if complete-information Nash equilibrium. The influence of this approach can be detected in several chapters, but it is explicit only in the present chapter.
3. The zero intelligence (ZI) approach uses computer simulations of very simple trader strategies to compare outcomes across market institutions. To the extent that simulated outcomes resemble outcomes with human traders, and to the extent that ZI strategies are not easily exploited by humans, more complex analysis becomes unnecessary. Many of the empirical chapters of this volume are influenced by the ZI approach.

All three approaches promise good returns to further work. I close with some conjectures about which activities will generate the highest returns in the next few years. In my view, the biggest gap in laboratory investigations of two-sided auctions is in environments which closely implement the prior knowledge assumptions of incomplete information games. Variants of the DA and CH institutions should be examined in environments with single-sided or with non-specialized traders whose value/cost parameters are drawn independently *each period* from known distributions. Such random value experiments should better establish the domains of applicability for incomplete versus complete information theories.

A second neglected area in the experimental literature is in comparing outcomes in two-sided auctions across common values and private values (and affiliated values) environments. Previous discussion foreshadows my conjecture that the lack of induced trading incentives and the "winner's curse" problem may make some or all market institutions less efficient in common values environments.

On the theoretical side, I hope to see more attempts to compare market institutions, with the goal of predicting which institutions will be most efficient in different environments. I suspect that in practice, the most efficient market institutions are those which best promote rapid learning of efficient equilibrium strategies. If so, we may have to supplement economists' usual equilibrium theories with positive theories of learning.

ACKNOWLEDGMENTS

I am grateful to the National Science Foundation for support under grant IRI-8812798, to John Rust for editorial suggestions, and to SFI conference participants for many useful comments.

REFERENCES

1. Amihud, Y., T. S. Y. Ho, and R. A. Schwartz, eds. *Market Making and the Changing Structure of the Securities Industry.* New York: Lexington Books, 1985.
2. Amihud, Y., and H. Mendelson. "Trading Mechanisms and Stock Returns: An Empirical Investigation." *J. Fin.* **42(3)** (1987): 533–553.
3. Arrow, K., and F. Hahn. *General Competitive Analysis.* San Francisco: Holden-Day, 1972.
4. Arthur, W. B. "Self-Reinforcing Mechanisms in Economics." In *The Economy as an Evolving Complex System,* edited by P. Anderson, K. Arrow and D. Pines, 9–32. Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume 5. Redwood City, CA: Addison-Wesley, 1988.
5. Benassy, J. P. "On Competitive Market Mechanisms." *Econometrica* **54(1)** (1986): 95–108.
6. Bronfman, C., K. McCabe, D. Porter, S. Rassenti, and V. Smith. "Tâtonnement Auctions: Theory and Behavior." Manuscript, University of Arizona, 1992.
7. Bulow, J., and P. Klemperer. "Frenzies and Crashes with Rational, Efficient, Market Clearing." SITE Manuscript, Stanford University, 1991.
8. Camerer, C., and K. Weigelt. "Testing for Information 'Mirages' in Experimental Asset Markets." Working Paper, Department of Decision Sciences, University of Pennsylvania, 1986.
9. Cason, T., and D. Friedman. "An Empirical Analysis of Price Formation in Double Auction Markets." This volume.
10. Cassady, R. *Auctions and Auctioneering.* Berkeley, CA: University of California Press, 1967.
11. Chamberlin, E. H. "An Experimental Imperfect Market." *J. Pol. Econ.* **56** (1948): 95–108.
12. Chatterjee, K., and W. Samuelson. "Bargaining Under Incomplete Information." *Operations Research* **31** (1983): 835–51.
13. Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb. *The Microstructure of Securities Markets.* New Jersey: Prentice-Hall, 1986.
14. Copeland, T. E., and D. Galai. "Information Effects on the Bid-Ask Spread." *J. Fin.* (1983): 763–797
15. Copeland, T. E., and D. Friedman. "Partial Revelation of Information in Experimental Asset Markets." *J. Fin.* **46(1)** (1991): 265–295
16. Cox, J. C., V. L. Smith, and J. M. Walker. "Theory and Individual Behavior of First-Price Auctions." *J. Risk & Uncertainty.* **1(1)** (1988): 61–99
17. Demsetz, H. "The Cost of Transacting." *Quart. J. Econ.* **82** (1968).
18. Dresher, M. L., L. S. Shapley, and A. W. Tucker. *Advances in Game Theory.* New Jersey: Princeton University Press, 1964.
19. Dubey, P. "Price-Quantity Strategic Market Games." *Econometrica* **50(1)** (1982): 111–126.

20. Easley, D., and J. O. Ledyard. "Theories of Price Formation and Exchange in Oral Auctions." This volume.
21. Fama, E., L. Fisher, M. C. Jensen, and R. Roll. "The Adjustment of Stock Prices to New Information." *Intern'l. Econ. Rev.* **10** (1969).
22. Fama, E. "Efficient Capital Markets: A Review of Theory and Empirical Work." *J. Fin.* (1970): 383–417.
23. Forsythe, R., T. R. Palfrey, and C. R. Plott. "Asset Valuation in an Experimental Market." *Econometrica.* **50(1)** (1982): 537–568.
24. Friedman, D. "On the Efficiency of Experimental Double Auction Markets." *Amer. Econ. Rev.* **74(1)** (1984): 60–72.
25. Friedman, D. "A Simple Testable Model of Double Auction Markets." *J. Econ. Behav. & Org.* (1991): 47–70.
26. Friedman, D. "How Trading Institutions Affect Financial Market Performance: Some Laboratory Evidence." Working Paper 221, Department of Economics, University of California at Santa Cruz, 1991.
27. Friedman, D., and J. Ostroy. "Competitivity in Auction Markets: An Experimental and Theoretical Investigation." Working Paper #633, Department of Economics, University of California at Los Angeles, 1991.
28. Friedman, D. "Evolutionary Games in Economics." *Econometrica* **59(3)** (1991): 637–666.
29. Friedman, D. "Privileged Traders and Asset Market Efficiency: A Laboratory Study." Working Paper 232, Department of Economics, University of California at Santa Cruz, 1991.
30. Fudenberg, D., and David M. Kreps. "A Theory of Learning, Experimentation, and Equilibrium in Games." Unpublished, 1988.
31. Garcia, G. B. "On the Dynamics of Prices to Equilibrium in a Double-Auction Market." Working Paper, University of Chicago Business School, 1980.
32. Garman, M. B. "Market Microstructure." *J. Fin. Econ.* **3** (1976).
33. Gilboa, I, and A. Matsui. "Social Stability and Equilibrium." *Econometrica* **59(3)** (1991): 859–860.
34. Glosten, L., and Paul Milgrom. "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders." *J. Fin. Econ.* (1983): 71–100.
35. Gode, D., and S. Sunder. "Human and Artificially Intelligent Traders in Computer Double Auctions." GSIA Manuscript, Carnegie-Mellon University, 1991.
36. Grossman, S., and J. Stiglitz. "Information and Competitive Price Systems." *Amer. Econ. Rev.* **66(2)** (1976): 246–253.
37. Harrison, G. W. "Theory and Misbehavior of First-Price Auctions." *Amer. Econ. Rev.* **79(4)** (1989).
38. Harsanyi, J. C. "Games with Incomplete Information Played by Bayesian Players." *Management Sci.* **14** (1967): 159–182, 320–334, 486–502.
39. Ho, T. S. Y., and H. R. Stoll. "The Dynamics of Dealer Markets Under Competition." *J. Fin.* (1983): 1053–1074.

40. Ho, T. S. Y., R. A. Schwartz, and D. K. Whitcomb. "The Trading Decision and Market Clearing Under Transaction Price Uncertainty." *J. Fin.* **40(1)** (1985): 21–42.
41. Hurwicz, L. "On Informationally Decentralized Systems." In *Decision and Organization*, edited by R. Radner and M. McGuire. 1972.
42. Kagel, J. H., and Dan Levin. "The Winner's Curse and Public Information in Common Value Auctions." *Amer. Econ. Rev.* **76(5** (1986): 894–920.
43. Kyle, A. S. "Continuous Auctions and Insider Trading." *Econometrica* **53(5)** (1985): 1315–1335.
44. Kyle, A. S. "Informed Speculation with Imperfect Competition." *Rev. Econ. Stud.* **56(3)** (1989).
45. Leamer, E. E. "Let's Take the Con Out of Econometrics." *Amer. Econ. Rev.* **73** (1983): 31–43.
46. LeRoy, S. "Efficient Captial Markets and Martingales." *J. Econ. Lit.* **27(4)** 1989: 1583–1621.
47. Luce, R. D., and H. Raiffa. *Games and Decisions.* New York: John Wiley & Sons, 1957.
48. Makowski, L., and J. M. Ostroy. "Efficient and Individually Rational Bayesian Mechanisms Only Exist on Perfectly Competitive Environments." Working Paper 556, Department of Economics, University of California at Los Angeles, 1989.
49. Marimon, R., and S. Sunder. "Price Dynamics in an Overlapping Generations Model." Working Paper, Carnegie-Mellon University, 1991.
50. McAfee, R. P., and J. McMillan. "Auctions and Bidding." *J. Econ. Lit.* **25** (1987): 699–738.
51. McAfee, R. P. "A Dominant Strategy Double Auction." Mimeograph copy, California Institute of Technology, 1989.
52. McCabe, K. A., S. J. Rassenti, and V. L. Smith. "Lakatos and Experimental Economics." Manuscript, University of Arizona, 1989.
53. McCabe, K. A., S. J. Rassenti, and V. L. Smith. "Auction Design for Composite Goods: The Natural Gas Industry." Manuscript, University of Arizona, 1989.
54. McCabe, K. A., S. J. Rassenti, and V. L. Smith. "Designing a Uniform Price Double Auction: An Experimental Evaluation." This volume.
55. McCabe, K. A., S. J. Rassenti, and V. L. Smith. "Designing Auction Institutions: Is Double Dutch the Best?" *Econ. Jour.* (1992).
56. Mendelson, H. "Market Behavior in a Clearing House." *Econometrica* **50(6)** (1982): 1505–1524.
57. Miller, J., R. Palmer, J. Rust, and V. Smith. "Comparison of the Behavior of Human and Computer Traders in a Simplified Double Auction Market." Manuscript, 1992.
58. Myerson, R. "Incentive Compatibility and the Bargaining Problem." *Econometrica* **47** (1979): 61–74.
59. Myerson, R., and M. Satterthwaite. "Efficient Mechanisms for Bilateral Trading." *J. Econ. Theor.* **29** (1983): 265–281.

60. Myerson, R. *Game Theory: Analysis of Conflict.* Massachusetts: Harvard University Press, 1991.
61. O'Brien, J., and J. Srivastava. "Dynamic Stock Markets with Experimental Assets." *J. Fin.* (1992): forthcoming.
62. O'Brien, J., and J. Srivastava. "Liquidity and Persistence of Arbitrage in Experimental Options Markets." This volume.
63. Plott, C. R., and V. L. Smith. "An Experimental Examination of Two Exchange Institutions." *Rev. Econ. Studies* **45** (1978).
64. Plott, C. R. "Industrial Organization and Experimental Economics." *J. Econ. Lit.* **20** (1982): 1485–1527.
65. Plott, C. R., and S. Sunder. "Efficiency of Experimental Security Markets with Insider Information." *J. Pol. Econ.* **90(4)** (1982): 663–698.
66. Plott, C. R., and S. Sunder. "Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets." *Econometrica* **56** (1988): 1058–1188.
67. Porter, D., and V. Smith. "The Scope of Bubbles and Crashes in Experimental Asset Markets." Discussion paper, University of Arizona, 1990.
68. Riley, J. G. "Strong Evolutionary Equilibria in the War of Attrition." *J. Theor. Biol.* **82** (no date):383–400.
69. Roberts, D. J., and A. Postlewaite. "The Incentive for Price-Taking Behavior in Large Exchange Economies." *Econometrica* **454** (1976): 115–128.
70. Rust, J., J. Miller, and R. Palmer. "Behavior of Trading Automata in a Computerized Double Auction market." This volume.
71. Rustichini, A., M. Satterthwaite, and S. Williams. "Convergence to Price-Taking Behavior in a Simple Market." CMSEMS Discussion Paper 914, Northwestern University, 1990.
72. Satterthwaite, M., and S. Williams. "The Rate of Convergence to Efficiency in the Buyer's Bid Double Auction as the Market Becomes Large." *Rev. Econ. Stud.* **56** (1989): 477–498.
73. Schaede, U. "The Development of Organized Futures Trading: The Osaka Rice Bill Market of 1730." In *Japanese Financial Market Research*, edited by Ziemba, Yasushi, and Bailey. North-Holland, 1991.
74. Schmeidler, D. "Walrasian Analysis via Strategic Outcome Functions." *Econometrica* **48(7)** (1980): 1585–1593.
75. Schwartz, R. A. *Equity Markets: Structure, Trading, and Performance.* New York: Harper & Row, 1988.
76. Shapley, L., and M. Shubik. "Trading Using One Commodity as a Means of Payment." *J. Pol. Econ.* **85(5)** (1977): 937–968.
77. Simon, L. "Bertrand, the Cournot Paradigm, and the Theory of Perfect Competition." *Rev. Econ. Stud.* **51** (1984): 209–230.
78. Smidt, S. "Continuous Versus Intermittent Trading in Auction Markets." *J. Fin. & Quant. Analysis* **14(4)** (1979): 837–866.
79. Smith, V. L. "An Experimental Study of Competitive Market Behavior." *J. Pol. Econ.* **70** (1962): 111–137.

80. Smith, V. L. "Experimental Economics: Induced Value Theory." *Amer. Econ. Rev.* **66** (1976): 274–279.
81. Smith, V. L., A. Williams, K. Bratton, and M. Vannoni. "Competitive Market Institutions: Double Auctions Versus Sealed Bid-Offer Auctions." *Amer. Econ. Rev.* **72** (1982): 58–77.
82. Smith, V. L. "Markets as Economizers of Information: Experimental Examination of the Hayek Hypothesis." *Economic Inquiry* **20** (1982): 165–170.
83. Smith, V. L. "Microeconomic Systems as an Experimental Science." *Amer. Econ. Rev.* **72** (1982): 923–955.
84. Smith, V. L., G. L. Suchanek, and A. W. Williams. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica* **56(6)** (1988): 1119–1152.
85. Smith, V. L., and A. W. Williams. "The Boundaries of Competitive Price Theory: Convergence, Expectations, and Transactions Costs." In *Advances in Behavioral Economics*, edited by L. Green and J. H. Kagel, vol. 2. Norwood, NJ: Ablex, 1989.
86. Stigler, G. J. "Public Regulation of the Securities Markets." *J. Bus.* **37** (1964): 117–134.
87. Stoll, H. R., and R. E. Whaley. "Stock Market Structure and Volatility." *Rev. Fin. Stud.* **3(1)** (1990): 37–71.
88. Vickrey, W. "Counterspeculation, Auctions and Sealed Tenders." *J. Fin.* **16** (1961): 8–37.
89. Walras, L. *Elements of Pure Economics*, translated by W. Jaffe. Homewood, IL: Irwin Press, 1973. Original French edition published 1874.
90. Williams, A. "Computerized Double Auction Markets: Some Initial Experimental Results." *J. Bus.* (1980): 235–258.
91. Wilson, R. "Equilibrium in Bid-Ask Markets." In *Arrow and the Ascent of Economic Theory: Essays in Honor of Kenneth J. Arrow*, edited by G. Feiwel. London: McMillan, 1987.
92. Wilson, R. "Strategic Analysis of Auctions." Manuscript, Graduate School of Business, Stanford University, 1990.
93. Zabel, E. "Competitive Price Adjustment Without Market Clearing." *Econometrica* **49(5)** (1981): 1201–1221.

Ian Domowitz
Department of Economics, Northwestern University, Evanston, IL 60208

TWO
Automating the Continuous Double Auction in Practice: Automated Trade Execution Systems in Financial Markets

We do feel that there is an area under discussion which is not only a clear and present threat to the efficient working of our markets, but also adverse to the public interest and one which carries with it tremendous adverse economic consequences. This is the area of computerized floor trading.

—Steven Greenberg, past Chairman
Board of Managers of the New York Cotton Exchange
Remarks made at the CFTC Conference on
Automation in the Futures Industry, June 15, 1977.

Automated. . .order execution systems are firmly established in our securities markets. These developments have improved the informational efficiency of our markets, increased the speed with which customer orders are executed, and expanded market capability.

—David Ruder, past Chairman
United States Securities and Exchange Commission
Remarks made at the Annenberg Forum on Technology
and Financial Markets, February 27, 1989.

1. INTRODUCTION

It is rarely the case that economists have a precise description of how trade is conducted in their examination of market activity. The leading exception to this statement is the auction institution. Auction markets with well-specified rules of trading have been observed for centuries, and the continuous double auction under the rubric of open-outcry floor trading has dominated United States financial markets for over 140 years. A major contribution of the theoretical and experimental literature on auctions is the observation that the form of the trading institution matters a great deal in the analysis of agent behavior, the properties of transactions prices, and welfare. This finding also is emphasized in the related, but largely disconnected, literature on financial market microstructure.

Technological advances now have introduced a new form of market institution: the computerized auction mechanism. The advent of automated trade execution is beginning to transform financial markets with respect to the building of exchanges, the conduct of the trading of financial instruments, and regulation of the markets. The scope of this revolution has not been fully documented, and may be surprising to many. It will not be long before a substantial amount of transactions data emanating from such automated systems becomes part of major data banks, upon which both theoretical observations and empirical work are based.

The purpose of this paper is to provide a survey of automated auction systems and to discuss some of the issues surrounding their adoption as market mechanisms. Any such survey must start with some description of the extent of the market with respect to automated exchange operations. Over 45 such systems are listed in section 2, spanning financial centers in 14 countries.

Automated trade execution systems are basically mathematical algorithms enabling trade matching, combined with information display and transmission mechanisms. The survey proceeds along such lines. A system may be thought of as a communications technology for passing messages between traders together with a set of rules that transform traders' signals into price and allocations. Rules also may include restrictions on which individuals may actively trade on the system, as well as on the times at which trading occurs. The exposition here is focused on continuous auctions, in which traders may submit messages in the form of bids and offers at any time for immediate execution should market conditions permit.

Priority rules are the heart of any trade execution algorithm. The priority assigned to any bid or offer, conditional on the state of the system at any given time, governs the place of the order in the queue awaiting execution. The relative distribution of such bids and offers in the queue determines the distributional properties of transactions prices, given the order placement strategies of traders. A listing of priority rules used in existing trade-matching algorithms is given and discussed in section 3. Characterizations of some systems in terms of the priority rules and their ordering incorporated in the execution algorithm are provided as examples.

Transparency in financial markets may be defined as the extent to which trading information is made available after each discrete event. Regulatory concerns generally are focused on the degree and amount of information provided to the investing public and to the regulators themselves. In this survey of systems, however, attention is limited to the information provided to and by the traders, i.e., to the actual users of the system. Although the flow of information to outside investors presents interesting public policy issues and affects order flow, it is the data available to the traders on the system itself that is most relevant in discussions of trading in automated auctions. Automation of the auction mechanism, as opposed to the face-to-face trading of securities on an open floor, presents alternatives in terms of information flow and display. This is discussed in section 4, and examples of screen design serving different purposes and types of markets are given.

The global nature of the development of computerized trading has motivated debate on an international scale relating to regulation of automated financial markets in particular. Some discussion of progress in this area is appropriate, as regulatory pressures will have an inevitable influence on the design and scope of automated exchange operations. Regulatory issues pertaining to order execution algorithms, transparency, system access, and operational problems are briefly outlined in section 5.

The question of whether automated auction mechanisms will dominate the open outcry floor institution in the competition for exchange services is the subject of section 6. Advocates of both modes of trading claim the same desirable set of market characteristics in any debate on the issue. The position taken in this paper is that opposition to the spread of automated markets largely stems from inertia brought about by trading heritage and tradition, as well as vested financial and human capital interests in maintaining the status quo. The development of automated exchanges in fast-growing markets lacking a floor-trading tradition, and surveys of exchange members on an exchange which embodies both modes of trading are cited in support of this proposition.

This last section raises an issue which is related, but not treated in this paper, that of optimal market structure. It is certainly true that many, if not all, of the systems in place today may not have optimal properties in one sense or another. On a theoretical basis, the prevalence of a wide variety of auction variants in financial markets presents a puzzle which has not been fully resolved. Friedman[12] notes that such variety may suggest that market efficiency is not sensitive to institutional details, but that environmental issues such as type of instrument traded may account for differences in market design. Striking systematic differences in design for alternative financial instruments are not observed in the case of automated execution systems, however, nor across jurisdictions with different national payment systems, for example. Variations that do exist sometimes can be traced to differences in regulatory structure, which certainly is part of the trading environment.[1]

Automated trade execution systems typically are not designed by economists backed up by theory and laboratory evidence, but by engineers, computer scientists, and market practitioners. There are some good reasons for this, relating in part to

[1]See, for example, Domowitz[7] for discussion of differences in the law and regulatory attitude that affect the differential treatment of systems for stocks and futures in the United States.

the politics of trading and to the idea of an exchange for the trading of financial instruments as a product itself.

There exists a large pool of traders who have human capital invested in the open-outcry paradigm, for example. These traders own pieces of existing exchanges in the form of seats, and, therefore, have a vested financial interest in the form of the market. In order to found a marketplace based on computers and automated execution, such ideas must literally be sold to the trading community. This community understands and embraces tradition in terms of market practice, and distrusts change on the grounds of theoretical optimality results. Thus, automated trading algorithms are not necessarily optimal, but they are adaptive, in the sense that they appeal to the community as a small move away from traditional trading, with benefits, say, in terms of cost and market exposure over the long run. An extreme example is the AURORA system proposed by the Chicago Board of Trade. The promotional literature for exchange members stresses that AURORA is not even a trade matching system, but rather a system that replicates pit trading on a video screen, with the loss only of the physical trading floor.

Existing systems merit serious study, however. The amount of physical and human capital investment in the current generation of automated trade execution systems will dictate a reasonably long life for designs now in place.

2. THE EXTENT OF AUTOMATION IN FINANCIAL MARKETS

Automation of the double auction institution in financial markets is taking place on a large scale. Automation of a market can include computerization of information dissemination services, order routing, and clearance and settlement procedures. In the context of this paper, automation of the trading mechanism refers specifically to the technology of trade execution: computerization of the trade matching, quantity allocation, and in most cases, price discovery mechanisms.

In this context, it is important to distinguish between automated trade execution and "automated trading." Automated trading is the practice of automatically transmitting orders to an exchange for execution of trades mandated by computerized contingent order strategies. So-called program trading and portfolio insurance are examples. Computerized trading was made feasible by advances in information dissemination and order routing, and certainly existed before much of the growth in automated trade execution. Program trades are not always represented and executed on the floor of the exchange as quickly as their designers might desire. Automated trade execution systems offer the potential of speeding up the process by providing computer-to-computer interfaces. Not all systems allow this, however.[2] Further, existing automated auctions do not allow complicated contingent orders

[2] The NSTS system permits computer-to-computer interface, for example, but the GLOBEX system does not.

at present. The SOFFEX system provides the possibility of trading an order contingent on the execution of a single additional trade. This represents the state of the art, and most systems do not even provide such a simple feature.[3]

Trading floors, where they exist, are being superseded or complemented by automated trade execution systems on a worldwide basis. Considerations of cost, market efficiency, and competition between exchanges for order flow, abetted by a lack of cooperation among international financial exchanges and growth in off-exchange trading, all have contributed to this growth in the utilization of technology. Not all such automated exchanges have succeeded. Citicorp and McGraw-Hill failed with the GEMCO electronic commodity trading system some years ago. The World Energy Exchange and the International Futures Exchange of Bermuda did not succeed in converting open-outcry traders to screen-based trading in the futures market.[4] It is difficult to pinpoint the reasons for the lack of success on a system-by-system basis. Most failed designs were introduced very early in the process of trade execution automation. This suggests both the possibility of flawed technological design and the lack of effort to integrate new mechanisms into existing market architecture in such a way as to attract order flow to the point of being self-sustaining. The Toronto CATS system is a counter example to the technology argument. Introduced in 1977, it survives today. It was introduced gradually, however, in conjunction with existing floor market operations. Only recently has there been a firm commitment to the idea of replacing floor trading completely by the automated system.[5] Some failure is attributable to forms of competitive pressure. Liquidity considerations create increasing returns with respect to market size, and overcoming this barrier to entry can be difficult. Sometimes just the perception of this barrier can be enough. The AURORA futures trading system appears to be stillborn at the end of its design stage as a result of the perceived gains from the first-mover advantage of the GLOBEX system.

Such failed ventures have not deterred others from building new systems or converting old ones. Tables 1 through 3 contain a listing of almost 50 automated trade execution systems in use or planned over the next couple of years; full names of systems and their associated exchanges are given with their acronyms in the appendix. Most of these efforts are very new. Over 25 systems have come on line between 1988 and 1991, with several more scheduled to start operation between 1991 and 1993. The vast majority of systems date from 1985 or later.

[3]This does not mean that the possibility does not exist. See Amihud and Mendelson[1] for suggestions with respect to an auction design integrating choices among alternative auction mechanisms with a computerized portfolio management system that produces and submits many forms of contingent orders.

[4]See Office of Technology Assessment.[18]

[5]See *The Wall Street Journal*, January 9, 1992.

TABLE 1 Automated Futures and Options Exchanges

System	Exchange	Start-up	Country	Global
APT	LIFFE	1989	UK	no
ATS	NZFOE	1985	New Zealand	no
ATS/2	IFOX	1989	Ireland	no
AUTOEX	AMEX	1985	US	no
AUTOM	PHLX	1990	US	no
CORES-F	TSE	1988	Japan	no
CORES-0	TSE	1989	Japan	no
DTB	GFOE	1990	Germany	planned
FACTS	TIFFE	1989	Japan	no
FAST	LFOX	1990	UK	yes
GLOBEX	CME	1992	US	yes
MOFEX	MOFF	1990	Spain	no
OTS	OSE	1989	Japan	no
POETS	PSE	1991	US	no
RAES	CBOE	1985	US	no
SFTS	OSE	1988	Japan	no
SOFFEX	SOFFE	1988	Switzerland	no
S-MART	MEFF	1990	Spain	no
SYCOM	SFE	1989	Australia	no
TGE	TGE	1988	Japan	no

Tables 1 and 2 contain information on futures/options systems and stock/bond systems, respectively, operating as a formal exchange. This means that the market is regulated as an exchange in its domestic market, and definitions for such treatment vary from country to country.[6] Table 3 covers proprietary systems, which enable the trading of stocks for the most part, but includes one options system.[7] Proprietary systems are not registered as exchanges, although they are subject to many of the same trade reporting requirements. Automated systems are included in

[6] In the United States, for example, an exchange is defined within the context of the Securities Exchange Act under Section 3(a)(1). The definition is so broad that virtually anything could be considered an exchange. Regulatory history has shown, however, that merely having a communication technology for bringing together buyers and sellers is necessary, but not sufficient, for a securities market to be classified as an exchange.

[7] Delta Government Options ("Delta") is operated by RMJ Securities, a registered clearing agency, and RMJ Options, a registered broker-dealer. The system trades options on underlying United States Treasury bills, bonds, and notes. Participants are primarily large banks and securities firms.

TABLE 2 Automated Stock and Bond Exchanges

System	Exchange	Start-up	Country	Global
ABS	NYSE	1976	US	no
BEACON	BSE	1987	US	yes
CAC	PSE	1986	France	no
CATS	TSE	1977	Canada	no
CLOB	SSE	1987	Malaysia	no
CORES	TSE	1982	Japan	no
GTB	MSE	1991	Italy	no
HKTS	SEHK	1993	Hong Kong	no
IBIS	FSE	1991	Germany	no
MAX	MSE	1981	US	no
MORRE	ME	1990	Quebec	no
NSTS	CSE	1985	US	no
OHT	NYSE	1991	US	no
PACE	PHLX	1976	US	no
SAEF	LSE	1989	UK	no
SCOREX	PSE	1969	US	no
SEATS	ASX	1987	Australia	no
SIB	SSE	1991	Spain	no
SOES	NASD	1985	US	no
STS	OSE	1991	Japan	no

TABLE 3 Proprietary Automated Trading Systems

System	Company	Start-up	Country	Global
BEST	KB	1986	UK	no
DELTA	RMJ	1988	US	no
INSTINET	Reuters	1985	US	yes
NORDEX	Transvik	1990	UK	yes
POSIT	Jefco	1987	US	no
TRADE	BZW	1986	UK	no
WAS	WASI	1991	US	no

these lists only if the trading protocol explicitly includes an "electronic handshake." In other words, person-to-person interaction for the purpose of trade execution is ruled out. There is a variety of proprietary off-exchange systems, in particular, that act mainly as electronic bulletin boards, requiring that trades actually be consummated by telephone.[8]

Automated markets are classified with respect to the system sponsor (exchange or company), date of inception, location, and global reach. A total of 14 countries are represented here. Hours of operation vary widely, and may even differ with respect to individual products traded on a given system. The main distinction is between systems which operate during the regular trading day and those that operate after-hours, usually supplementing a conventional trading floor. Examples of the latter include GLOBEX and SYCOM for futures, and OHT for stocks.

The vast majority of automated systems operate during regular trading hours. In many cases, the automated system is the main trading system of the exchange; i.e., all trades in a financial product are processed through the automated execution mechanism. Exceptions generally involve systems which are designed to handle only small retail customer orders or trading in very illiquid issues. Such mechanisms use prices for trade matching based on activity in a floor-trading market that usually is in operation during the same time period. For example, RAES operates in tandem with the options trading floor of the Chicago Board of Options Exchange. The crowd in the pit is trading continuously, and the best bid and offer extant on the floor at any given time are transmitted to RAES, which processes retail customer orders of ten or fewer options contracts. Such limited execution mechanisms are relatively rare in futures and options trading, including AUTOEX, AUTOM, and POETS, in addition to RAES. Examples in the case of the trading of stocks and bonds are BEACON, MORRE, PACE, SCOREX, and SOES. Most execution systems of this type are quite old, dating back as far as 1969. The newer generation of automated mechanisms is composed of systems that stand alone. Global reach pertains to whether or not terminals are located outside the home country. In most instances, a "no" in that column implies that there are regulatory restrictions against such an operation, but that is not always the case. The IBIS stock-trading system is under no such legal restriction, for example, but all terminals are located in Germany with no immediate plans for expansion into cross-border trading.

Computerized exchanges easily lend themselves to the idea of cross-border trading. There are no real technological barriers. Despite the frequency with which one reads about "globalization of trading," however, it is clear from these tables that electronic markets are not spearheading the move into international trading activity at present. Only two futures/options systems are oriented this way, with the DTB system planning such operations. The FAST system of the London Futures and Options Exchange specifically advertises its international trading operations as a direct way to increase the number of market participants and attract liquidity. The GLOBEX system of the Chicago Mercantile Exchange will operate in partnership

[8] Twenty systems have been granted the right to operate as nonexchange facilities in the United States, for example. Several of these operate as such electronic bulletin boards. Others have failed by the time of the writing of this paper, including Econ Investment Software, Adler & Co., Security Pacific, Troster Singer, Exchange Services, Transaction Services, and B&K Securities. See Becker, Adkins, Fuller, and Angstadt.[2]

with foreign exchanges and offer overseas terminals. In stock trading, only BEACON of the Boston Stock Exchange maintains a foreign connection, and it is limited to a link with Montreal. The best global reach is provided by INSTINET, which has terminals located around the world. Trading in U.S. equities is supplemented by dealing in U.K., French, German, Dutch, Swiss, Norwegian, Finnish, and Swedish stocks. Many of the problems arising with respect to greater cross-border trading activity through automated exchanges concern regulatory issues and international regulatory cooperation, which are the topics of section 5.

3. AUCTIONS AS ALGORITHMS: TRADE EXECUTION PRIORITY RULES

The trade execution function is an algorithm that performs order matching according to a set of rules governing the priority of submitted bids and offers. The priority rules determine the place of a bid or offer in the queue awaiting execution. A match occurs under several circumstances, depending on the design of the system. In some systems, a match occurs the moment an order rises to the top of the queue, at a price possibly determined outside of the automated system. A match may occur in other designs when a bid or offer at the top of the queue is accepted directly by the touch of a button. In limit-order matching systems, transactions occur when the orders cross; i.e., when the price of the best offer to buy is equal to or greater than that of the best offer to sell.

An example of a specific trade execution algorithm may help. The following subset of the GLOBEX limit-order system trading rules is taken from Domowitz.[8] [9]

1. *Order eligibility.* A new order is eligible to be matched with a standing order, and a trade will result, whenever the following conditions occur:
    1. One order is a buy order and the other is a sell order.
    2. The two orders are for the same contract.
    3. The price of the buy order is greater than or equal to the price of the sell order.
2. *Trade price.* If an order match is possible according to the criteria of Rule 1, then the trade will take place at the price of the standing order.
3. *Trade quantity.* If an order match is possible according to Rule 1, then the trade will take place for a quantity equal to the smaller of the
    1. remaining quantity of the new order or
    2. remaining quantity of the standing order.

[9] There also are special rules governing the setting of an opening price in the GLOBEX system, as well as a facility to directly take an existing bid or offer on the limit order book.

4. *Maximization of total trade size.* If there are multiple standing orders eligible for matching against a new order, then matching will be considered in priority sequence until one of the following conditions is attained:
    1. the new order is completely filled or
    2. all eligible standing orders have been considered.
5. *Standing order priority.*
    1. Price: for buy orders, higher price is higher priority; for sell orders, lower price is higher priority.
    2. Quantity: a standing order for "primary quantity" has a higher priority than that for "secondary quantity" if they are both at the same price. A standing order for secondary quantity has priority over a standing order for primary quantity if the supplementary quantity is at a better price. A supplementary quantity order may be executed only in conjunction with its associated primary quantity.
    3. Time: Within the same price and quantity type, older orders have higher priority.

The first three rules are a part of most trade execution algorithms. The term "standing order" refers to a bid or offer entered previously into the system, which has been saved on the electronic order book. The fact that the trade takes place at the price of the standing order replicates floor trading practice. Variations of rule 4 are not independent of priority rules, and are considered below in that context. It is possible, for example, to have some kind of sharing rule among all orders at the same price, regardless of time of order entry, which is not the case here, given priority rule 5.3.

There are three priority rules that govern this execution algorithm. Best price (5.1) is the chief priority. Following price is time: first in, first out. The final priority is one of display. A trader may split a bid or offer at the same price into primary and secondary amounts. The primary quantity is shown to all system participants. The secondary quantity is not displayed. The displayed quantity has precedence over that which is not displayed. If a trader's secondary quantity cannot be executed at the same time as the primary, the system will cancel the secondary bid or offer, as undisplayed orders have zero priority if they stand alone without some displayed quantity.

The purpose of this section is to describe the trade execution priority rules used in practice.[10] In principle, all automated trade execution systems can be characterized by an ordered list of such rules. The list below is not ordered in any particular fashion, however.

[10] Harris[14] also discusses order precedence rules, but more from a normative point of view with an eye towards improvements of rules in efforts to increase liquidity in the market. The list here is more comprehensive, but covers only rules currently used on existing systems.

1. *Price.* Best price is the highest priority on virtually all systems. Trade-matching systems which take transactions prices from a market exogenous to the system do not have this rule built into the algorithm, except in the indirect sense that the price at which the trade match occurs is the best quote available in the exogenous market.
2. *Price with market-maker exposure.* In some systems, market makers constitute a class of system participants with responsibility to execute part of retail customer order flow. Certain mechanism designs incorporate the feature that an order is exposed to the market maker for a few seconds to allow the possibility of bettering the existing best quote in the market. If the market maker declines to do so, the order is rerouted for execution at the existing quote structure, in accordance with any other priority rules. BEACON, SCOREX, MAX, and NSTS all contain this feature.
3. *Time.* Time priority means first in, first out, but almost always refers to time at a particular price, not time in the system. Even time at price is not always a sufficient description. GLOBEX, for example, allows suspension of orders at a given price, which sacrifices time priority when the order is reactivated.
4. *Modified time.* Modified time priority is used with quantity allocation rules, described below. Several traders may have bids outstanding at the same price. Modified time priority would give preferential treatment in terms of quantity allocation to the trader with the highest time priority order, while treating the remaining participants equally, in the case of an incoming offer eligible for matching. The APT and the proposed HKTS use a modified time priority rule.
5. *Order type.* Order types include market orders, limit orders, block orders, and cross orders. Cross orders refer to two orders from the same system participant, a buy order and a sell order for the same quantity at the same price.[11] The HKTS will give higher priority to such orders. SOFFEX has special procedures to deal with crosses that involve cancellation (zero priority) of one side of a cross under some circumstances. Block orders refer to bids or offers above some specified quantity. Block order priorities are treated under quantity priority below. Market orders, i.e., orders to buy or sell at the best current quotes, almost always have priority over priced limit orders. They are not allowed in some systems, GLOBEX, for example.
6. *Hit and take.* Hitting the bid (accepting the currently outstanding bid) or taking the offer (accepting the currently outstanding ask) refers to an action, rather than a priority, but a description of systems in terms of priority rules would be incomplete without it. It is an action which may supersede other priorities. Not all systems have such a feature. Those that do include GLOBEX, APT, SYCOM, SOFFEX, IBIS, and INSTINET.

[11]The idea here is that the system trader has received such orders from two retail customers who do not have direct access to the system.

7. *Quantity.* Size precedence gives priority to bids or offers for large quantity. Such a priority rule would generally displace a time priority rule. The SOFFEX block-trading facility imposes a size priority for trades above a certain quantity, but subject to numerous qualifications.
8. *Quantity allocation.* Size allocation priorities preempt time priority or are used with a modified time rule. The system may, for example, allocate an equal number of shares or contracts of an incoming eligible offer to each system participant bidding at the same price until the incoming order is filled or all bids with prices eligible for matching are exhausted. Alternatively, the system can allocate incoming orders to eligible system bidders on a pro rata basis, i.e., according to the quantity bid. The Stock Exchange of Hong Kong is currently debating the form of allocation mechanism for the proposed HKTS. APT is a pro rata allocation system.
9. *Display.* There are two possible types of display precedence. The first may be classified as a rule which gives priority to bids and offers which display the size of the order and the identity of the trader over orders which hide identity. Systems to date either give trader identification or they maintain anonymity, but do not prioritize this way. The second gives priority to bids and offers whose size is displayed to the market over orders that are submitted, but not displayed to the system participants. Such a feature favors traders who disclose their order information to the market. GLOBEX, CATS, and NORDEX embody this type of priority rule.
10. *Trader class.* It is customary in U.S. equity markets to give public limit orders precedence over dealer quotes at the same price, regardless of time precedence.[12] Although this rule would be easy to implement on any electronic system, only NSTS, HKTS, and ATS/2 on the list presented in section 2 appear to have this feature.[13]
11. *Preferencing.* Preferencing is the practice of routing a customer's orders to a particular system participant by prior arrangement. Preferencing rules on SOES and SAEF require preferenced orders to be executed at the best quote of any dealer in the market, regardless of whether the preferenced dealer is offering the best price.[14]

All continuous automated auction systems currently in place can be characterized in terms of an ordering of these priorities. The GLOBEX algorithm presented at the first part of this section could be given as (price, time, display, hit/take). Systems such as FAST, ATS, and FACTS are simply (price, time) priority mechanisms. MOFEX is written as (price, order type, time). A complete classification of algorithms is attempted in Domowitz.[10]

[12] See Harris[14] for discussion.

[13] There is some ambiguity in this case as well. Trading rules for ATS/2 specify this class distinction explicitly, but it is not clear that it is enforced as part of the trade execution algorithm.

[14] See Stoll and Huang[22] for a discussion of the advantages and disadvantages of preferencing arrangements.

Periodic single-price auctions such as WAS and TGE, as well as market-opening procedures for continuous automated markets, use priority rules in a very limited way. Such systems are automated forms of the clearinghouse auction discussed in Friedman.[12] Bids and offers are submitted over some period and executed together at a single price at a single point in time. The price is calculated by minimizing the total bid/offer size imbalance and/or by maximizing the total volume traded over possible transactions prices. Unlike the ideal Walrasian tâtonnement procedure, some bids or offers eligible for trade at the chosen price cannot be executed because supply will not precisely equal demand, an effect largely due to the imposition of a discrete minimum price variation. In this case, some priority ordering must be used to decide what trades should be executed. Market orders often are executed first, followed by priced orders in order of price and time.

The statement of priority rules is a bit terse, and some distinctions and elaborations in terms of design deserve mention. The price improvement offered by the market-maker exposure in rule 2 can be automated, for example. The Midwest Stock Exchange offers two algorithms for orders of certain sizes that replace market-maker exposure by automated price improvement in markets with bid/ask spreads wider than one eighth of a point. One of these is quite similar to the practice of stopping orders on the NYSE for price improvement that depends on the relation between the last and next sales on the exchange.[15]

Hitting the bid or lifting the offer is not equivalent to a market order. Market orders operate in electronic systems precisely the same way as they do on the floor, in that a market order is assured execution. Hitting the bid, for example, involves touching a button that signals acceptance of some or all of the size of a bid at some price, usually the best bid. If another trader touches this button first, his trade will get executed. However, if he takes only part of the outstanding bid, the remaining quantity is available to other traders, after which the hit order is discarded from the system. It is possible on GLOBEX, for example, to hit the bid for a price lower than the best bid, in which case the order is filled for as much of the specified quantity as possible down to that price. This is not to say that real-time trading via market orders is not allowed on automated systems. Market orders can and are processed against quotes from the limit order book, as well as market-maker quotes, on many systems. The hit/take option is supplementary to market orders in these cases. In systems such as GLOBEX which only accept limit orders, this option is a primary means of market interaction. Real-time revision of orders is permitted on all systems, including cancellation. Cancellation and resubmission at a different price sacrifices time priority. Cancellation of part of the size of an order without changing the price for the remainder generally retains priority in the system. Quantity priorities obviously are affected by such an action. Suspension of orders, as mentioned in rule 3, is a form of temporary cancellation with the option of simultaneous resubmission of all orders, given that price and size are unchanged. Finally, precedence with respect to trader class refers to the origin of an order, and not to rules dictating the types of individuals allowed to interact

[15] See *The New York Stock Exchange Floor Official Manual*, June, 1990, p. 16.

directly in the system. For systems operated by exchanges, direct participation is heavily restricted, in much the same way as entry to the floor. The purchase or lease of a seat often is required. Training programs are offered and examinations must be passed, in the interest of orderly markets. For proprietary systems, capital requirements usually are the only impediment to system trading.

4. SYSTEM INFORMATION: SCREEN DISPLAY

A distinct advantage of automated trade execution systems is that of information availability. In principle, all possible trading information can be made available to everyone at all times, because it is all embodied in the computer system. The magic of window-screen environments allows great depth of information display in convenient formats. A full discussion of system transparency requires consideration of information flow to three broad classes of market participants: system traders, public investors, and regulators. Attention is restricted to system traders in this paper, as they are the direct participants in the automated auction process. The types of information discussed in this section differ in many respects from those commonly analyzed in the theoretical literature on auctions. Systems do not provide private signals containing information leading to the formation of reservation values. All system signals are common, with the exception of an individual trader's position in the security. Under that assumption, traders receive independent unbiased signals concerning an uncertain reservation value. To the extent that the trading process disseminates private information, systems differentiated with respect to their communication technology have varying effects on market efficiency in the sense of information-theoretic models of trading.

Data potentially available to system users in real time for any individual security includes: high price and low price of the trading session; price and size of the latest trade; the best bid and offer prices (BBO); quantities available at the BBO; prices of all bids and offers in the system; the size available at all prices; trader identification for each quote; counterparty identification for each trade; sales record of the session; aggregate volume traded in the session; number of system participants or terminals active; and relevant information from other markets.

This list excludes information potentially available to a system user that ordinarily would be considered private. The position of a trader in any or all securities traded is one example.

Despite the fact that an automated system could offer all such information, since the computerization of the trading process requires or produces it,[16] systems vary widely with respect to information display. The simplest designs literally

[16]The key exception is information from other markets in the case of the trading of derivative securities. Such information is not required or produced by the computerization of the auction, and could be obtained from other sources in real time.

have no display of their own. These systems are limited to trade-matching algorithms which operate according to time priority only, and use prices from outside the system as transaction prices. Price information and quotation display must be obtained from quote vendors servicing the outside market generating prices. Such systems include AUTOM, RAES, AUTOEX, POETS, BEACON, MAX, SCOREX, PACE, and SOES.

General market information concerning last trade and aggregate volume, for example, is available on most systems that produce such information as part of the trade execution process. Information from other markets is part of a few systems including GLOBEX and NSTS.

Screen-based trading systems are anonymous for the most part. Few offer identification of system users posting bids and offers to system participants. Exceptions are CATS, APT, GTB, INSTINET, and SEATS.[17] It is much more common to observe the availability of the counterparts to executed trades, although this sometimes is private information, limited to the parties themselves.

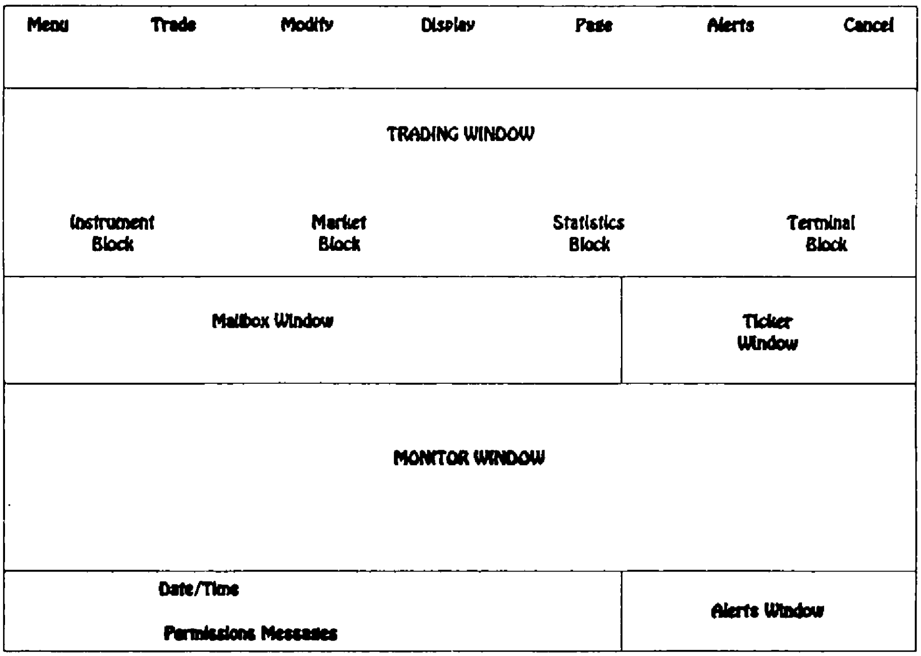


FIGURE 1 Globex Main Display.

[17]Broker identification is required in SEATS for quantities less than $10,000. Disclosure is optional for bids or offers over that amount.

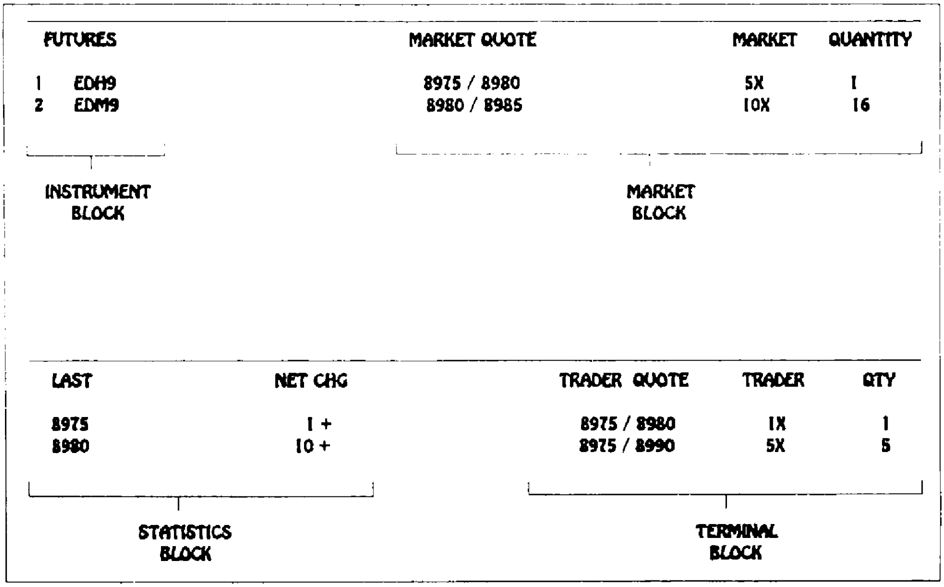


FIGURE 2 Globex Trading Window Detail.

The major difference between systems that has implications for pricing and market efficiency is the availability of current bid and offer information. Many systems do not display the book of bids and offers to all system participants, giving only the best bid and offer in real time. Mechanisms displaying the book include GLOBEX, CORES (stocks, futures and options), SFTS, OTS, CAC, and CATS, among others, but the practice is by no means universal. Further, several systems allow bids and offers into the system that are not shown to market participants. These include GLOBEX, APT, CLOB, CATS, and NORDEX. Not all system participants are treated equally with respect to information in some designs. Such informational differences appear to be limited to trader identification. NSTS offers a "public" limit order book with aggregate quote information, while designated market makers have a screen available which shows all market-maker quotes, identified individually by dealer. CAC has three levels of information, providing quote and trade identification information only to brokers.

The design of the screen display obviously depends on the information to be made available, but also differs depending on the market served by the automated system. Several examples are provided in figures 1 through 7.

Figure 1 illustrates the general layout of the GLOBEX screen. GLOBEX is a futures and options system, and the monitor window displays information on current prices in the underlying spot markets. The alerts window serves as a signaling device to an individual trader, and can be programmed to alert the system user when the price of a contract reaches a certain level, for example. The ticker window gives real-time execution information.

The main trading window is broken up into four blocks of information: financial instruments monitored, market, statistics, and terminal. These blocks are illustrated in Figure 2. The market block gives the BBO with size for the instruments listed on the left of the screen. The statistics information on the normal view of the screen is limited to the price of the last transaction and the net change in price. The terminal block is personalized, containing the system user's own quotations for the instruments listed in the instrument block.

There are three other possible ways to view the main trading screen in GLOBEX, shown in Figure 3. The "statistics view" replaces the terminal block with additional market information on the instruments displayed. The "more view" again is personalized to the trader; it shows the quantities bid or offered by the user that are not shown to other users via the open limit order book. Finally, the "two-up view" replaces the statistics and terminal block with additional market information, including bid-ask spreads trading on the system.

GLOBEX is a limit order system that displays the order book.[18] A typical order window is illustrated in Figure 4, taken from the SYCOM system. Note that

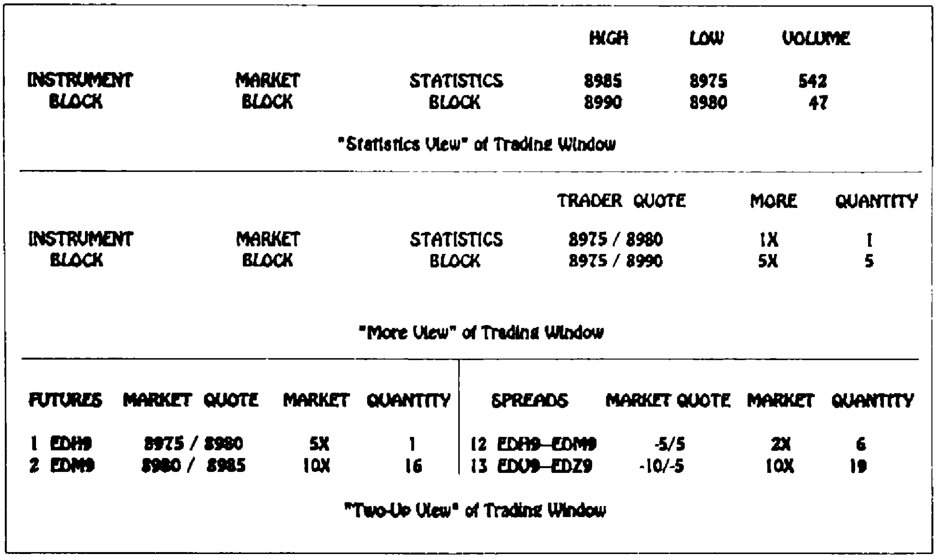


FIGURE 3 Globex Trading Window Detail.

[18] More precisely, the ten best bids and offers with associated size.

CONTRACT SUMMARY

	XBZ9	XBHO	XBZ
OPN	6465	6660	280
HI	6465	6660	280
LO	6464	6659	279
VOL	66	10	12
ASK	6465	6659	280
SIZE	33	31	6
BID	6464	6659	279
SIZE	44	5	2
LST	6464	6659	279
CHG	↓1	↓1	
CLOSE	6465		

CONTRACT—XBZ9
6464

SIZE	ASK	BID	SIZE
33	6465	6464	44
33	6465	6463	44
33	6466	6462	55
33	6467	6461	55

F1		F2		F3		F9	F10
XBZ9		XBHO		XBZ		STRATEGY SCREEN	TRADING SCREEN
B	6464	B	6658	B	279		
A	6465	A	6659	A	280		
T	6464	T	6659	T	279		

FIGURE 4 Sycom Main Screen with Trading Window.

the book contains the aggregate size available at each price, and this is virtually universal among order book displays.[19] The contract summary screen serves the same purpose as the instrument, market, and statistics blocks in GLOBEX. The bottom of the SYCOM main screen contains boxes corresponding to function keys on the terminal keyboard. The BBO and last trade price are illustrated for each contract selected by the user, and a press of the key allows trading screen and strategy screen information for that instrument. There is no analogue to the strategy screen on GLOBEX, because the means of entering bids and offers differ on SYCOM. Bids or offers for a single instrument are entered into the terminal, but not transmitted to the host computer immediately.[20] This allows the trader to create a deck of orders for different securities (and/or the same security at different prices and quantities). The trader can then sequentially release several such orders simultaneously for execution. The strategy window shows the user the set of bids and offers programmed.

[19] An exception is the market maker display in NSTS, which shows all quotes by dealer, with size for each quote.

[20] Contingent orders are not accepted, however, nor can orders for different instruments be spliced together.

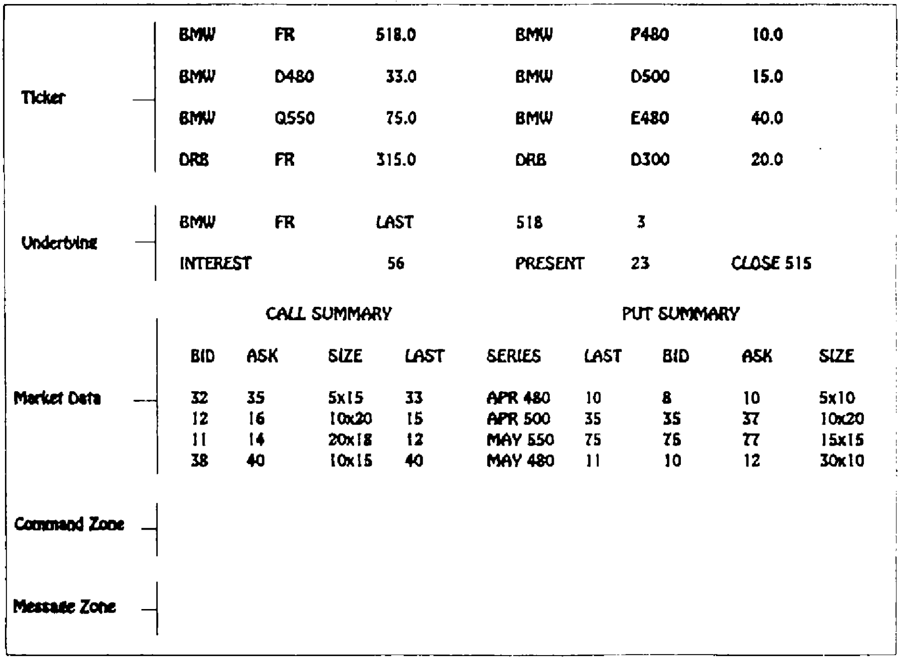


FIGURE 5 DTB Main Trading Screen.

No information on the equities or indexes underlying the futures and options traded on SYCOM is provided by the system itself, but that feature appears on the DTB main screen in Figure 5. Only the BBO for the options listed on the screen is given in the market data window. The ticker window here shows only the futures and options contracts traded and the last execution price. The screen does show the number of current participants, however, a feature not present in the two preceding examples. NSTS of the Cincinnati Stock Exchange links several equity markets through the Intermarket Trading System, and offers a variety of screen displays. Figure 6 contains the screen that differentiates this system from others. The national best bid and offer is displayed, together with the best quotes on seven markets in addition to the automated market. Statistics now include national and system information for the current and previous day's trading. Finally, the box at the right contains the book seen only by designated market makers on the system, including quotes by individual dealer.

Finally, no review of terminal displays would be complete without a pit-trading simulation. Figure 7 shows the screen of the Chicago Board of Trade's AURORA futures system, a market that probably will not be implemented. It is of interest, however, in that AURORA was planned to be a video simulation of the trading pit,

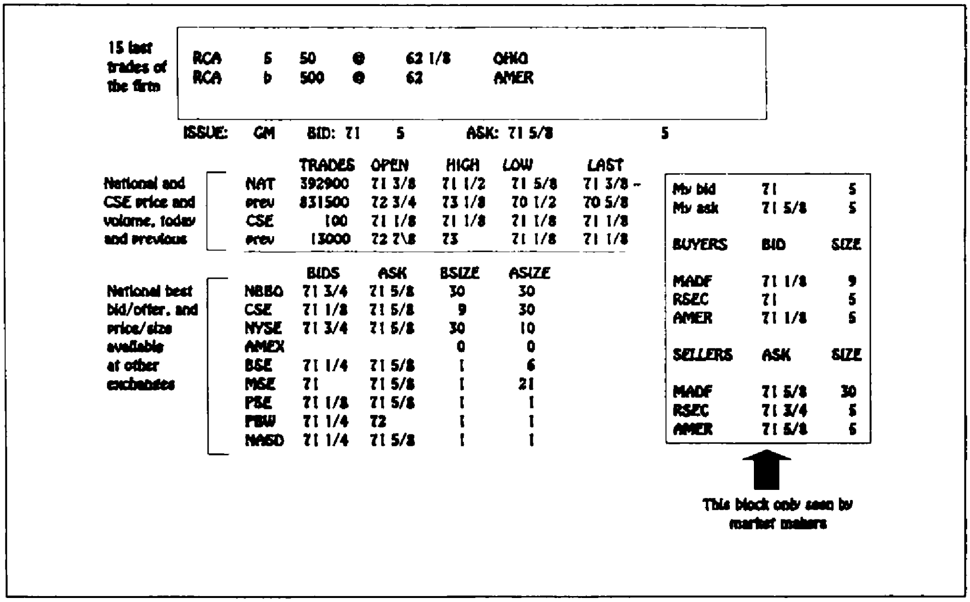


FIGURE 6 Cincinnati Stock Exchange Indepth Market Display.

and the icons in the center represent bids and offers with trader identification and size. The APT system was developed along similar lines. The price at which these bids and offers are made must be the best in the market, replicating pit trading. This price is shown in the upper left. The aggregate size at each price is in parentheses. The boxes on the screen border are reminiscent of screens placed around the usual trading floor. The boxes on the bottom left show contracts for the same instrument at different expiration dates and traded spreads. The ones on the right contain best bid and offer information from other simulated "pits." The position of the system user is given in the bottom right corner, while the box above contains transaction information shown to all participants, including identification of the parties making the trade.

5. REGULATION AND AUTOMATED AUCTION MARKETS

Much of the focus of United States regulatory attention with respect to automated systems is on the soundness of the system as a technical matter. In reviewing automation proposals, the Securities and Exchange Commission (SEC)

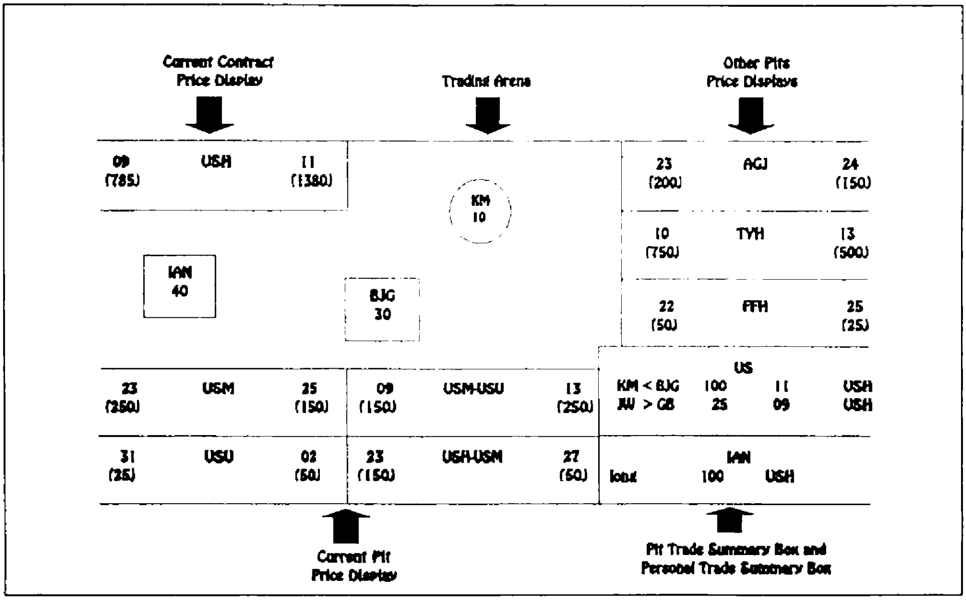


FIGURE 7 Aurora Main Display.

attempts to ensure that the introduction of new systems does not threaten to overwhelm the capacity of existing financial market structures. Surveillance and issues of compliance with existing regulatory standards also are of direct regulatory concern.[21] The United States General Accounting Office (GAO), while lauding the potential of automated systems to control trading abuses, concerns itself with "generic risks associated with automation."[22] Risks delineated by the GAO relate to the soundness of hardware and software with respect to correct information processing and responsive operations, as well as the security of the computer system and the need to ensure continuous service. Although such issues are of obvious importance to both system users and regulators, they involve computer system design, rather than the rules and mechanics of automated auctions, and will not be discussed further.

There is, however, domestic regulatory worry with respect to the pricing efficiency of automated trade execution systems. Section 5(g) of the Commodity Exchange Act states that any market for the trading of futures contracts must show that its activities are not contrary to the public interest. The public interest,

[21] See, for example, Ruder and Adkins[21] for a discussion of the Securities and Exchange Commission's attitude towards regulation of automated systems.

[22] See GAO/IMTEC-89-68,[13] which relates to automation of futures markets, but could as well have been applied to equity markets.

in turn, is defined in Section 3 of the Act in terms of reliable price discovery.[23] Discussion and analysis of this issue and others with respect to floor trading versus automation in futures markets are contained in Domowitz.[7,9]

The global proliferation of systems has attracted international regulatory attention to the fact that screen-based trading embodies features which distinguish it from traditional floor-based trading practices. The capability of computerized systems to link market participants at remote locations raises the issue of international regulatory cooperation. Placing screens in different regulatory jurisdictions can obviously lead to overlapping claims of regulatory authority, for example.

In response, the International Organization of Securities Commissions (IOSCO) formed a working party to consider a variety of issues relating to automated market design and regulation, consisting of representatives from Australia, France, Italy, Japan, Switzerland, the United Kingdom, the United States, and Germany. This list should not be surprising, considering the system locations documented in the previous section. The group suggested the following areas of emphasis: access, operational issues, financial integrity, surveillance, disclosure, security and system vulnerability, order execution algorithm performance, and system transparency. The result was a statement of principles for the oversight of screen-based trading systems, which touched upon each of the issues above.[24] These principles provide a convenient framework for discussion of regulatory issues, as well as give some insight into regulators' concerns. Such concerns in turn translate into broad guidelines for system design. The principles are stated broadly as items to examine in the regulation and design of systems. The discussion below is an attempt to briefly elaborate on some of the issues that bear directly upon automation of auctions, omitting problems with hardware and computer security, in particular.

The pricing algorithm is a primary concern in the consideration of market efficiency. The algorithm performs a similar function to the trading rules in a floor-trading system. Standards should be applied to assess these rules in the same way as they are used to judge floor-trading practices with respect to efficient pricing.[25] Other issues that arise include, for example, the possibility of preferential treatment to certain classes of traders in the interest of increasing market liquidity and the degree to which orders are given an equal opportunity to obtain execution.

From a regulatory perspective, the nature of the algorithm is a factor in determining the potential for trade abuses. Some abuses may be eliminated completely by algorithm design; a trivial example is after-hours trading, which is stopped by

[23] In an attack on the Chicago Mercantile Exchange's GLOBEX system, the Chicago Board of Trade claimed that the proposed mechanism would violate the statutorily recognized goal of reliable pricing in futures markets. See letter from Thomas R. Donovan, President, Chicago Board of Trade, to Jean A. Webb, Office of the Secretariat, Commodity Futures Trading Commission, dated September 8, 1988, concerning the proposal contained in 53 *Federal Register* 25528, July 7, 1988.

[24] See International Organization of Securities Commissions.[15] Although the report specifically addressed systems for the trading of futures and options, the principles enunciated therein are completely general and could easily be applied to equity markets.

[25] This is the approach taken in Domowitz[9]; see also Domowitz and Wang.[11]

simply turning off the computer. Trading abuses which depend on knowledge of traders' identities may be eliminated by making all bids and offers anonymous and removing the possibility of counterparty selection. The addition of the latter to an algorithm may exacerbate problems with disclosure or prearranged trading. In other words, the algorithm governing order execution could affect the manner in which authorities regulate. Periodic single-price auctions would eliminate the very need to worry about prearranged trading or trading ahead of customer orders.[26] All such considerations mandate a thorough examination of the trade-matching algorithm by the regulatory authority.

In reviewing an algorithm, the regulatory authority should consider the flow of information to participants. The system should be able to ensure the equitable availability of timely trade and quotation information to like classes of system participants. In principle, one advantage of automation is the potential of providing all participants, including system traders, outside investors, and regulators with the same comprehensive information set in real time. As pointed out in Domowitz,[10] existing systems do not systematically make the same information available to the public that is given to participants who have system screens. The benefits of full market information must be weighed against potential costs in terms of liquidity, for example. Full reporting of large trades and interdealer quotes to the public may lead to greater system trader risk if such traders are engaged as market makers, thereby resulting in larger bid-ask spreads and a less efficient market. The issue is not clear at this point, and regulatory assessment of such costs on any particular system must be made on a case-by-case basis.

Differences in treatment among classes of system participants must be identified and made clear to all users of the system. The issue of response time, i.e., the time between the transmission of an order from the user's terminal and the receipt of confirmation by the host computer, is a good example peculiar to automated markets. Difficulties may arise in ensuring equal response times when terminals are dispersed over a large geographical area. The need to ensure that response times are equitable over classes of users is more important from the regulatory point of view than the actual response time itself. The problem is most pressing in systems with time priority rules, of course. Performance of a system operating under quantity allocation rules without modified time priority is less likely to be degraded by unequal response times. The issue is not restricted to time priority violations, however. Delays in receipt of a better price or a command to hit the bid due to unequal distance from the host computer will affect market allocations. From the perspective of hardware, optical fiber telephone connections are, therefore, preferable to satellite links. Other solutions include buffers that equalize response time by delaying orders transmitted close to the host relative to those coming from great distances. Complete synchronization of the auction process leads to periodic clearinghouse auctions, and has not been considered a solution to the problem, given the demand for continuous trading.

[26]These themes also are elaborated upon in Corcoran and Lawton.[6]

Associated with this point is the issue of access. Screen-based trading mechanisms have the potential to increase the degree of direct access to a market beyond that which is usually available in floor-trading environments. The types of access permitted in a system and the means by which access controls are enforced can affect the degree and manner of regulatory oversight of a system. Increase in access should not threaten to disrupt the orderly operation of the screen-based market itself, or any related floor or underlying securities markets. Concerns include the qualifications of individuals receiving terminals, the security of the computer system itself, as well as what activities may be conducted by potentially different classes of terminal operators.

The system design itself may be able to provide access control. Consider the issue of credit worthiness, for example. The system can be designed to maintain individual records of all trader positions and their value at any point in time, as well as to receive information concerning current bank balances and letters of credit. The system program could then simply log off a trader who has taken a position beyond credit controls set by the exchange or clearing agency.

This is not an exhaustive list of concerns. It does, however, highlight some issues important to market regulation that arise in automated auctions in particular. The success of cooperative regulatory efforts on an international scale requires that regulators in different jurisdictions think about the same problems in similar ways. The agreement on a set of general principles for regulatory oversight, even if stated extremely broadly, is an important step in the right direction. Variations in domestic law and market custom always will remain to some extent. Flexibility in regulatory policy is important, if only to avoid unnecessarily constraining system design. Such flexibility will be crucial, however, if the potential of automated systems for cross-border trading is to be realized.

6. TAKING OVER THE MARKETS: REFORMING THE FLOOR AND OTHER STORIES

A recent headline in the *Financial Times*, reads, "Farewell to the trading floor as markets plan automation."[27] In fact, the potential for automating financial markets to such an extent has been recognized in policy circles since at least 1963.[28] The first measured discussion of how to think about such a task in the financial literature seems to be that of Black,[3] leading to the Peake, Mendelson, and Williams[19,20] proposal for an automated auction system, culminating in the comprehensive market-cum-trading strategy system of Amihud and Mendelson[1] and the electronic call market design of Cohen and Schwartz.[5]

[27] *Financial Times*, July 22, 1991.

[28] See "Special Study of Securities Markets, Report of the Special Study of the SEC" (1963) in H.R. Doc. No. 95, 88th Congress, 1st Session, pt. 2 at 358 and 678.

Despite such intellectual activity and the progress documented in section 2, a farewell to the institution of floor trading still may take some time. The issue received a public hearing as early as 1977, at a conference on automation in the futures industry sponsored by the Commodity Futures Trading Commission. The discussion revealed a deep split between academics and computer experts, on the one hand, and market practitioners, on the other.[29] Part of the problem, today as well as then, is that both sides of the debate over automation of the markets claim the same set of desirable market characteristics. For example, proponents of computerized trading argue that an improvement in the flow of information is available, resulting in greater transparency of the market for a wider range of market participants than can be had under pit trading. Those that support the floor trading institution in turn believe that information on the exchange floor cannot be replicated by computerized information. Melamed,[17] in particular, emphasizes special aspects of floor trading as stimuli for the development of opinions and ideas that contribute in an essential way to the price discovery process and generation of liquidity in the market. His opposition to automated trading floors at the time was as vehement as his excellent defense of organized floor trading.

Leo Melamed now is considered the founder of the Chicago Mercantile Exchange's (CME) GLOBEX system. The CME views GLOBEX as a response to the major issues of globalization, automation, and off-exchange trading. The CME membership generally appears to support GLOBEX, given the vote approving the system, but accepts the idea in large part because GLOBEX provides trading opportunities after regular floor-trading hours and is not currently viewed as a replacement of the trading pits. Financial incentives also play a part. GLOBEX is structured as a limited partnership, with the membership receiving 70 percent of the profits.[30]

Floor trading still is viewed as preferable by many traders and exchange officials. William Brodsky, president of the CME, believes that floor trading remains the primary focus of liquidity in any trading environment, for example.[31] He cites the past and current liquidity generated by the Chicago floor markets for derivative securities as the evidence.

There is a "chicken before the egg" problem with the argument, as such. Floor trading has been the dominant trading institution in the United States for over a century. The current generation of traders has invested a great deal of human capital in this market. Liquidity, indeed, is focused into the pits at the moment,

[29] See the recorded transcript of proceedings (as opposed to the written collection of papers later published) of the Conference on Automation in the Futures Industry, June 15, 1977, on file at the offices of the Commodity Futures Trading Commission, Washington, D.C. The transcript provides a much better insight into the feelings of the participants on the subject.

[30] The split among members depends on the nature of the seats held on the CME. Full members will receive three units, IMM members two units, and IOM members one unit. This information is garnered from a memo from R. Dufour to Strategic Issues Task Force members, Chicago Board Options Exchange, dated August 30, 1988.

[31] Address to representatives of the Societé Generale Bank of France, Northwestern University, July 3, 1991, and in personal communication.

because floor markets have been the only focus for quite some time. If traders had been given computers and screens before the advent of open outcry trading, they might well view yelling across an open-pit to be an arcane form of doing business. Liquidity would be centered in the computerized trading system.

One "economic experiment" that might shed some light on this issue concerns the growth of trading in futures and options. For many years, such trading took place almost exclusively in Chicago. The number of financial futures and options listed on exchanges worldwide has grown from 16 in 1978 to 205 in 1988, however. Concurrent with this interest in derivative financial securities, the number of futures and options exchanges listing such products has grown from 10 in 1978 to 37 in 1988.[32] The number of new futures and options exchanges has expanded accordingly. Most are overseas, where there is little tradition of open-pit trading for futures and options contracts. Virtually all such new exchanges embody automated trade execution, with over fifteen new foreign automated exchanges planned and constructed since 1988.[33] New stock exchanges abroad often are automated. Movements away from periodic single-price auction in equity markets also have been towards automated continuous auctions rather than floor trading.[34]

There is other evidence that traders may analyze the shortcomings of open-outcry auction and come out in favor of automated systems, given time and information to assess the alternatives. CATS has been in operation on the Toronto Stock Exchange (TSE) since 1977, enabling trading in an increasing number of equity issues over the years. In October 1987, the TSE commissioned McKinsey & Company to conduct a review of floor and computerized exchange operations and to determine the best trading modes for the exchange as a whole.[35] The TSE members suggested four criteria of market quality for analysis: visibility of market information, access, fairness, and cost efficiency.

Surveys of the membership uncovered the fact that the market participants found that the trading floor performs poorly on several dimensions of market quality. Over 90 percent said that the visibility and quality of information from the floor required enhancement. Widespread dissatisfaction (89 percent) existed with respect to trade execution and trade reporting on the floor. The survey indicated that many members preferred the order execution and reporting capabilities of CATS, and suggested using the automated system as the standard against which to measure floor-trading performance. Finally, 86 percent of the membership said that the floor was not completely fair.

These results led to modifications of existing floor-trading practices in ways that are reminiscent of computerized auctions. For example, floor traders must commit orders to a book; undeclared floor orders have no standing. The goal here

[32] See Chapman,[4] tables 6 and 7.

[33] New exchanges not surveyed in section two for lack of definitive information include the Belgian and Austrian futures and options exchanges, among others.

[34] This includes the Paris and Milan stock exchanges, for example.

[35] The final report was published in April 1989, and the material cited below is based on information provided therein; see McKinsey & Company.[16]

is to distribute information about the liquidity of the market to the greater investment community, as well as to traders both on and off the floor. Broad display of market-by-price information to the public is supplemented by information to exchange members with respect to broker identification for each order. This particular asymmetry of information previously was noted as not uncommon in computerized auction environments. Traders away from the floor may execute orders automatically against bids and offers on the book, allowing a short delay (no more than 20 seconds) to allow floor market-maker intervention under some circumstances. Such a rule ensures that traders away from the floor have the same access to the market as floor traders, and moves the floor closer to the instantaneous execution capability of an automated system.

A modified time priority rule, such as that discussed in section 3, is one of the key features of the modified floor. Modified time priority is to be supplemented with a proportional allocation of trades among members desiring to trade at the same price. This procedure may encourage traders to show more size. It also removes market-makers' flexibility in determining whether shares are allocated either proportionately or equally, potentially aiding in the issue of fairness.

McKinsey visits to other exchanges appear to reinforce the idea that the extent to which exchange members believe that the floor offers benefits relative to automated trade execution depends on the "trading heritage" of the trader.[36] Opinions surveyed depended on the particular biases of floor traders' experiences. This makes the McKinsey TSE study of particular interest, in that traders there had been exposed to an automated system for an extended period before the time of the survey. The poll of TSE members reveals that there was no general agreement with respect to the possibility that the benefits of face-to-face trading outweigh the instantaneous order entry, execution, and information dissemination capabilities of the existing automated auction mechanism.

In this context, it is interesting to note that Toronto now proposes to shut down its floor operations completely in favor of the automated auction. A vote is scheduled for the middle of February, 1992, on whether or not to do so by January, 1993. The only strong opposition appears to be from the 105 TSE market makers, together with a subset of traders who actively continue to use the floor.

7. CONCLUDING REMARKS

Markets no longer need embody physical locations to enable trading. As reported by the current chairman of the SEC to the U.S. Congress, market forces now are linked electronically, and computerized trading is an integral part of both domestic

[36]See appendix B of Mckinsey.[16]

and foreign financial markets.[37] The survey information presented here serves to document and reinforce this statement.

The advent of computerized trade execution is described as a revolution in this paper. A revolution, by definition, is a radical change of circumstances in an existing system. Automated execution certainly is viewed as such by current market practitioners. Several reasons have been cited for this movement, including cost, better information, market efficiency and fairness, wider market access, growth in off-exchange trading activity, and competition between exchanges for order flow. Occasionally, the rationale is a bit more idiosyncratic. The SOFFEX system exists in part because at the time a futures exchange was planned, each major financial center in Switzerland already had an exchange of some sort. The political solution was to computerize the futures market, and give all centers equal access.

Wider market access includes the extension of trading hours and a growing interest in trading directly with other partners at the institutional level. The latter is manifested by the growth of crossing networks such as POSIT and INSTINET, for example. Competitive pressures to extend trading hours have motivated computerization of markets in Chicago, Sydney, London, and (to a much lesser extent) New York. Although after-hours trading is an obvious rationale for automation, the majority of systems surveyed in this paper operate during regular trading hours. New developments such as Reuters' Dealer 2000 for the interbank foreign exchange market will support 24-hour trading, but that system will operate during regular hours as well.

Such innovations suggest the importance of globalization of trading activity as a driving force behind automation. In terms of long-range planning, this is undoubtedly the case. The data presented in section 2 indicate that few automated systems operate on a cross-border basis, however, and globalization cannot account directly for the growth in systems. Regulatory problems constitute the main hindrance to international development.

Data are not available to adequately quantify reductions in cost. Execution fees on the fully computerized CSE are substantially smaller than those charged for the same stock issues on the NYSE, however, suggesting cost savings. Pearce Bunting, president of the TSE, has stated that the main reason for ending floor trading is "dramatically reduced costs" for the exchange and its members. Although he declined to estimated such savings, a TSE spokesperson claimed that elimination of the floor will cut space needs alone by about 40 percent. Some tangential evidence might be inferred from related types of systems. For example, the LSE TAURUS share settlement system is expected to cost 65 million pounds over four years, but yield a savings in staff alone of 50 million pounds per year.[38] Revealed preference arguments suggest that if computerized markets are established as the primary means of off-hours activity, they are at least cheaper, including some social costs,

[37]See letter from Richard C. Breeden, chairman, SEC, to Edward J. Markey, chairman, subcommittee on Telecommunications and Finance, U.S. House of Representatives, dated 11 July 1991.

[38]See *The Economist*, 26 October, 1991, p. 98.

than maintaining a full complement of floor-based market-making functions. In that case, failure to replace day trading by completely automated markets again is traceable to vested interests and political pressures.

It is tempting to ascribe advances in automated execution to competition for order flow. Discrimination between competition among exchanges for order flow and competition over the form of the trading institution itself is difficult, however. Competition between exchanges is focused on accessibility, the breadth of instruments traded, reduced fees, and best execution, a catch-all which covers speed of execution, reporting, and pricing.[39] The same factors influence the form of the trading institution. Existing floor-based exchanges which have introduced some form of automated execution obviously find such automation to be a potentially effective competitive device. On the other hand, competition over the trading institution exists in the absence of competition between existing exchanges for order flow. Proprietary systems such as INSTINET and fully automated exchanges such as the CSE are examples of automated execution systems in the form of new entrants competing with floor-trading operations. The growth of derivatives markets in Europe and Asia has spawned automated exchanges which represent "behind the scenes" competition over the form of the trading institution, in that the new exchanges were faced with a choice of institution upon inception and chose automation. Consideration of competition over the form of the trading institution raises an important issue for the student of industrial organization and market structure. This is the topic of market fragmentation, which currently is of great policy interest. Concern has been expressed, in particular, that automated trade execution systems further enable the growth of off-exchange trading, and may result in a "balkanization" of the national securities market inconsistent with the National Market System concept of an integrated nation-wide system of competitively traded securities. As a result, Congress fears that the goal of maintaining an efficient price-setting mechanism for securities will be undermined.[40] The SEC has focused debate on this issue on the costs and benefits relating to systems that use prices from other markets to enable trade execution as opposed to systems that provide a price discovery service.[41] This focus is too narrow, given the existing level of competition involving automated systems that do provide such a service. A broader theoretical and empirical examination of the problem may serve to provide useful and different perspectives.

[39] Legally, best execution is defined simply as execution at the best bid or ask extant in the market. The term is interpreted more broadly here.

[40] See, for example, letter from Edward J. Markey, chairman, Subcommittee on Telecommunications and Finance, U.S. House of Representatives, to Richard C. Breeden, chairman, SEC, dated 16 May 1991. The National Market System concept is formally expressed in the 1975 Securities and Exchange Act.

[41] See Memorandum, from William H. Heyman, director, Division of Market Regulation, to Richard C. Breeden, chairman, SEC, dated 3 July 1991.

ACKNOWLEDGMENTS

Support from the National Science Foundation under grant SES-8921952, Household International, and the Center for Urban Affairs and Policy Analysis, Northwestern University, is gratefully acknowledged. I thank Dan Friedman, John Rust, and Jianxin Wang for helpful comments and suggestions.

APPENDIX

SYSTEM ACRONYMS

ABS	Automated Bond System
APT	Automated Pit Trading
ATS	Automated Trading System
ATS/2	Automated Trading System, updated
AUTO-EX	Automated Exchange
AUTOM	Automated Options Market System
BEACON BSE	Automated Communications and Order Routing Network
BEST	name, no acronym
CAC	Cotation Assistée en Continu
CATS	Computerized Automated Trading System
CLOB	Consolidated Limit Order Book
CORES	Computerized Order Routing and Execution System
CORES-F CORES	for futures
CORES-O CORES	for options
DELTA	name, no acronym
DTB	Deutsche Terminborse
FACTS	Fully Automated Computerized Trading System
FAST	Fully Automated Securities Trading System
GLOBEX	Global Exchange
GTB	Generale Telematico di Borsa
HKTS	Hong Kong Trading System
IBIS	Integrated Trading and Information System
INSTINET	Institutional Trading Network
MAX	Midwest Automated Execution System
MOFEX	Mercado de Opciones y Futuros Financieros
MORRE	Montreal Registered Representative System
NORDEX	name, no acronym
NSTS	National Securities Trading System
OHT	Off-Hours Trading
OTS	Options Trading System

PACE PHLX	Automated Communications and Execution
POETS	Pacific Options Exchange Trading System
POSIT	Portfolio System for Institutional Trading
RAES	Retail Automated Execution System
SAEF SEAQ	Automated Execution Facility
SCOREX	Securities Communication Order-Routing and Execution System
SEATS	Stock Exchange Automated Trading System
SFTS	Stock Futures Trading System
SIB	Sistema de Interconexion Bursatil
S-MART	Securities Market
SOES	Small-Order Execution System
SOFFEX	Swiss Options and Futures Exchange
STS	Securities Trading System
SYCOM	Sydney Computerized Overnight Market
TGE	Tokyo Grain Exchange
TRADE	name, no acronym
WAS	Wunsch Auction System

EXCHANGE ABBREVIATIONS

AMEX	American Stock Exchange
ASX	Australian Stock Exchange
BSE	Boston Stock Exchange
CBOE	Chicago Board Options Exchange
CME	Chicago Mercantile Exchange
CSE	Cincinnati Stock Exchange
FSE	Frankfurt Stock Exchange
GFOE	German Futures and Options Exchange
IFOX	Irish Futures and Options Exchange
LFOX	London Futures and Options Exchange
LIFFE	London Intern'l. Financial Futures Exchange
LSE	London Stock Exchange
ME	Montreal Exchange
MEFF	Mercado Espanol De Futuros Financieros
MOFF	Mercado de Opciones y Futuros Financieros
MSE	Midwest Stock Exchange
MSE	Milan Stock Exchange
NASD	National Association of Securities Dealers
NYSE	New York Stock Exchange
NZFOE	New Zealand Futures and Options Exchange
OSE	Osaka Securities Exchange
PHLX	Philadelphia Exchange
PSE	Pacific Stock Exchange
PSE	Paris Stock Exchange
SEHR	Stock Exchange of Hong Kong
SFE	Sydney Futures Exchange
SOFFE	Swiss Options and Financial Futures
SSE	Singapore Stock Exchange
SSE	Spanish Stock Exchanges
TGE	Tokyo Grain Exchange
TIFFE	Tokyo Intern'l. Financial Futures Exchange
TSE	Tokyo Stock Exchange
TSE	Toronto Stock Exchange

REFERENCES

1. Amihud, Y., and H. Mendelson. "An Integrated Computerized Trading System." In *Market Making and the Changing Structure of the Securities Industry,* edited by Y. Amihud, T. Ho, and R. A. Schwartz. Lexington, MA: Lexington Books, 1985.
2. Becker, B., A. Adkins, G. Fuller, and J. Angstadt. "The SEC's Oversight of Proprietary Trading Systems." Working Paper, Office of Self-Regulatory Oversight and Market Structure, Division of Market Regulation, United States Securities and Exchange Commission, 1991.
3. Black, F. "Toward a Fully Automated Exchange." *Financial Analysts Journal* **27** (1971): July-August, 29–35, November-December, 25–28, 86–87.
4. Chapman, A. B. "International Securities Regulation in a Global Electronic Environment." In *Innovation and Technology in the Markets: A Reordering of the World's Capital Market Systems*, edited by D. R. Siegel. Chicago, IL: Probus, 1990.
5. Cohen, K. J., and R. A. Schwartz. "An Electronic Call Market: Its Design and Desirability." In *The Challenge of Information Technology for the Securities Markets*, edited by H. C. Lucas, Jr. and R. A. Schwartz. Homewood, IL: Dow Jones-Irwin, 1989.
6. Corcoran, A., and J. Lawton. "The Effect of Variations Among Automated Trading Systems on Regulatory Oversight." Paper prepared for the Futures and Options Regulators Meeting, Burgenstock, Switzerland, United States Commodity Futures Trading Commission, 1991.
7. Domowitz, I. "When Is a Marketplace a Market: Automated Trade Execution in the Futures Market." In *Innovation and Technology in the Markets: A Reordering of the World's Capital Market Systems*, edited by D. R. Siegel. Chicago, IL: Probus, 1990.
8. Domowitz, I. "The Mechanics of Automated Trade Execution Systems." *J. Fin. Intermediation* **1** (1990): 167–194.
9. Domowitz, I. "Equally Open and Competitive: Regulatory Approval of Automated Trade Execution in the Futures Market." *J. Futures Markets* (1992): forthcoming.
10. Domowitz, I. "A Taxonomy of Automated Trade Execution Systems." Unpublished manuscript, Northwestern University, 1992.
11. Domowitz, I., and J. Wang. "The Electronic Order Book and Automated Trade Execution in Futures Markets." Unpublished manuscript, Northwestern University, 1991.
12. Friedman, D. "The Continuous Double Auction Market Institution: A Survey." This volume.
13. General Accounting Office. "Automation Can Enhance Detection of Trade Abuses but Introduces New Risks." GAO/IMTEC-89-68, United States General Accounting Office, 1989.

14. Harris, L. E. *Liquidity, Trading Rules and Electronic Trading Systems.* Monograph 1990-4, New York University Salomon Center, Leonard N. Stern School of Business, 1990.
15. International Organization of Securities Commissions. "Screen-Based Trading Systems for Derivative Products." Report of the Technical Committee, International Organization of Securities Commissions, 1990.
16. McKinsey & Company. "Charting the Future Growth of the Toronto Stock Exchange." Final report to the Toronto Stock Exchange, McKinsey & Company, 1989.
17. Melamed, L. "The Mechanics of a Commodity Futures Exchange: A Critique of Automation of the Transaction Process. *Hofstra Law Review* **6** (1977): 149–172.
18. Office of Technology Assessment. "Trading Around the Clock: Global Securities Markets and Information Technology." OTA-BP-CIT-66, United States Government Printing Office, Washington, D.C., 1990.
19. Peake, J. W., M. Mendelson, and R. T. Williams, Jr. "The National Book System." Paper submitted to the National Advisory Board of the United States Securities and Exchange Commission, 1976.
20. Peake, J. W., M. Mendelson, and R. T. Williams, Jr. "Toward a Modern Exchange: The Peake-Mendelson-Williams Proposal for an Electronically Assisted Auction Market." In *Impending Changes for Securities Markets: What Role for the Exchanges?*, edited by E. Bloch and R. A. Schwartz. Greenwich, CT: JAI Press, 1979.
21. Ruder, D., and A. Adkins. "Regulation and the Automation of Information Dissemination and Trading in the United States." In *Innovation and Technology in the Markets: A Reordering of the World's Capital Market Systems*, edited by D. R. Siegel. Chicago, IL: Probus, 1990.
22. Stoll, H. R., and R. Huang. "Major World Equity Markets: Current Structure and Prospects for Change." Working Paper No. 90-32, Owen Graduate School of Management, Vanderbilt University, 1990.

II. Theories

David Easley† and John O. Ledyard‡
†Cornell University, Ithaca, NY 14853, and ‡California Institute of Technology, Pasadena, CA 91125

THREE
Theories of Price Formation and Exchange in Double Oral Auctions

We provide a theory to explain the data generated by experiments with double oral auctions. Our theory predicts convergence to the equilibrium implied by the law of demand and supply and provides an explanation of disequilibrium behavior. The predictions of our theory seem to fit the data better than do the predictions of Walrasian, Marshallian, or game theoretic models. Our theory also suggests that, in demand-supply environments, the double oral auction is remarkably robust in the sense that aggregate performance is similar for a very wide range of individual behaviors.

1. INTRODUCTION

One of the main justifications for the use of equilibrium models in economics is the argument that there are forces which tend to drive agents and their decisions towards an equilibrium if they are not at one already. Market equilibrium models have proven to be extremely powerful in the analysis of many situations; however, attempts to model and explain the forces that do drive an economy to equilibrium

The Double Auction Market, Eds. D. Friedman and J. Rust,
SFI Studies in the Sciences of Complexity,
Proc. Vol. XIV, Addison-Wesley, 1993

have met with little success. Most of the literature on the stability of equilibrium uses the fiction of a disinterested auctioneer who adjusts a single known price for each good in response to stated excess demand resulting from agents' equilibrium plans. The limitations and defects of this approach are well known: for a survey of the literature, see Arrow and Hahn.[1] In addition, as far as we know, the *only* institutional arrangement that even approximates this idealized model of price formation is the London gold market (see Jarecki[9]).

Now, however, a body of data has been generated which provides detailed information on the disequilibrium behavior of traders in auction markets similar to those of organized commodity or stock exchanges. These data are difficult to ignore since they are generated experimentally under controlled conditions, and cannot be explained away by reference to measurement error, unobserved variables, or other fudge factors. In the experiment, a small number of traders, each with limited imperfect information, determine prices and quantities transacted through interactive bargains. There is neither a single price nor a single price quoter. Nonetheless, the quantities exchanged and the prices at which transactions take place typically converge to, or near to, the values predicted by the law of demand and supply. But, in spite of the fact that the traditional demand-supply model appears to yield reasonably accurate predictions of the long-run average prices and quantities in these markets, it fails to yield any insights into the process by which these prices and quantities are obtained.

In this paper we consider several positive theories of the price formation and exchange process for the class of experimental exchange markets called Double Oral Auctions.[1] We examine three of these theories in detail and argue that one of them seems to be the most consistent with the data. The ability of this theory also to explain price formation and exchange in other markets such as the New York Stock Exchange depends, of course, on the degree of parallelism that exists between the two (see Smith[20]). An astronomer's maintained hypothesis is that the physics of the lab is the same as that of the sun; our working hypothesis is that behavior in experimental markets is similar to that in other markets, and that insights discovered in the evidence generated in the lab are potentially transferable to nonexperimental markets with similar institutional structures. Thus, we view the theory in this paper as a first step towards constructing a positive theory of the process of exchange and price formation in many other markets.

2. THE EXPERIMENTAL MARKET

In a double oral auction (DOA) experiment, a pool of subjects (usually eight to twelve) is divided at random into a group of buyers and a group of sellers. The

[1] In fact, some of the auctions are computerized rather than oral. All that matters is that participants can make bids or offers and acceptances, and are informed of others' bids or offers and acceptances.

buyers are given value schedules telling them the amount in cents that they will receive from the experimenter for each unit of the good they purchase. Buyers keep the difference between their value and the price they pay for that unit. The sellers are given cost schedules telling them their cost in cents for each unit of the good they sell. Sellers keep the difference between their selling price and their cost on each unit they sell. Each subject knows his own payoff schedule but is given no information about the others' payoffs. Smith[20] shows how these payoff schedules induce demand and supply schedules. An example of payoff schedules and the induced supply and demand schedules for one experiment is provided in Appendix A.

After they receive their payoff schedules, subjects are allowed to trade during a market period of some fixed length, usually called a market day. Buyers can make bids to buy a unit of the good and sellers can make offers to sell a unit. If a bid or offer is accepted, a binding trade occurs and all traders are informed of the contract price. Once a trade is completed, bids and offers can be made for another unit of the good. No information other than bids, offers, acceptances, and contract prices is transmitted or known by the participants.

When a market period ends, the subjects are given new payoff schedules, identical to their schedules for the previous period, and the experiment is repeated.[2] Market demand and supply conditions are typically held constant across periods so that any equilibrating process that exists has a chance to establish an equilibrium. For a more detailed explanation of auction experiments and the usual results, see Williams[22] and Smith and Williams.[21]

These experiments provide a unique opportunity to examine price formation for two reasons. The first is that, unlike nonexperimental markets, the actual prices and quantities predicted by the law of demand and supply are known to the experimenter. Secondly, complete data on bids, offers, contracts, and their timing is available. An example of a typical design and the data generated is provided in Appendix A. Demand and supply functions can be calculated from the subjects' valuations, and equilibrium prices and quantities can then be computed. The first obvious fact from these experiments is that *actual exchange prices are not equal to those predicted by the law of demand and supply.* In a strict sense, demand-supply theory is rejected by these data. The second obvious fact, however, is that after a very few replications, *transaction prices and quantities converge to near those predicted by the law of demand and supply.* These observations have been replicated many times. The only conclusion one can draw is that the traditional theory needs refining before one has a compelling explanation of the observed behavior in these markets. Not only must "equilibrium" be explained, but we must also explain the "disequilibrium" transactions, the sequence in which they occur, and the process by which participants are "learning."

In our search for a better theory of price formation, we have used several criteria. Firstly, we wanted the theory to predict convergence to the predictions of the law of demand and supply for those experiments in which convergence occurs and to predict nonconvergence in those experiments in which convergence does not occur.

[2]Other designs are also used; see Smith[19] and Smith and Williams[21] for some of these.

Secondly, we wanted the theory to be useful in understanding the dynamics of the adjustment process by making falsifiable predictions about the entire process. A theory which predicts eventual convergence at $T = \infty$, and nothing else, is consistent with the data but not very illuminating. Third, we wanted the predictions of disequilibrium behavior not to be at odds with the data. How one weighs these criteria against one's prior belief in any particular theory is a matter of judgement. Our choice will be evident from the theories we reject and the candidate we offer.

3. THREE POSSIBLE THEORIES

Our goal is to understand how the actual dynamics of these markets work, not how they should work. We recognize that there are a variety of models which purport to explain price adjustments, but we view the existence of experimental data as an opportunity to reject a subset of those theories which seem obviously inappropriate. The set of reasonable theories for these markets can now be constrained by the data in a way that has been unusual for economics. To see what this means, let us consider three candidates for a theory of market dynamics.

Since both the institutional description and the data from the experimental DOA markets reject the Walrasian tâtonnement auctioneer as the appropriate model of price formation, a natural alternative might be a Marshallian theory. In a naive version of this theory, the trading sequence depends on the differences in buyers' values (willingness to pay) and sellers' values. In particular, this theory predicts that trade will occur in the efficient order, i.e., the first trade will occur between the buyer with the highest induced value (Buyer 1 in the example in Appendix A) and the seller with the lowest induced cost (Seller 1 in the example in Appendix A). The second trade is predicted to occur between the buyer and seller with the second values and costs, and so on. This theory does not predict which prices will occur, but it does predict that the total quantity transacted will be the competitive equilibrium quantity. Unfortunately, this theory has little to do with reality. When we look closely at the microdata, we see that the theory is soundly rejected. A cursory glance at the summary data of Appendix A should convince even the most skeptical reader that the predictions of the naive Marshallian theory are not at all consistent with the data. (In IPDA14, the rank correlation coefficient between the order of the true values and the order of the transactions is .369 in week 1 and .273 in week 2.) This is an excellent example of a case in which the experimental setup allows us to test more hypotheses than would be possible if we only had access to nonexperimental market data. Testing the prediction concerning the order in which participants are involved in transactions would be impossible without explicit knowledge of the individual valuations.

A second candidate for a theory is found in Friedman[3] who takes an alternative approach to the problem by redefining the experiment. He studies one day of a DOA with traders who are allowed to resell or repurchase the good being traded. Under a no-congestion condition which requires that at the day's end no trader wants to reset the closing bid or ask prices or accept the outstanding bid or ask, he shows that the final allocation will be at most one transaction away from being Pareto-optimal. No congestion implies that the final ask be no more than the second-lowest cost of selling a unit, the final bid be no less than the second-highest value of buying a unit, and that no one wants to accept the final prices. With resale and repurchase allowed, this insures that all but perhaps one infra-marginal unit has traded and that no more than one extra-marginal unit has traded. Beyond the question of the appropriateness of the no-congestion assumption, the difficulty in applying this theory to the DOA experiments is that the theoretical conclusion relies heavily on the agents' ability to retrade, while retrading is not allowed in many of the experiments. The theory also finesses the issues of learning and dynamics. How no congestion occurs is left unexplained.[3]

A third candidate for a theory would be a model based on game-theoretic considerations. For most of the DOA experiments, there is a complete-information Nash equilibrium (with price-quantity offers or bids as strategies) in which all trades take place at the competitive equilibrium price. However, the use of a Nash equilibrium concept to describe the experimental market has two difficulties. Firstly, the data are not consistent with this equilibrium (not all trades occur at the competitive price). Secondly, the participants in the experiments do not have enough information to calculate the strategies required to support this equilibrium. (They would have to be able to calculate the competitive equilibrium price.) Thus, one must turn to models with asymmetric information.

In the experiments which have been run, details on others' payoffs (and thus on the competitive equilibrium) may only be inferred by the subjects from the public data on bids, offers, and contracts. Thus, the structure in which subjects find themselves is a dynamic game with incomplete information. If an equilibrium were calculated for this game, its predictions could then be compared with the data. We feel that there are at least three difficulties with using this approach to construct a positive theory of double oral auctions. Firstly, as common knowledge about the distribution of valuations and the strategies selected are not controlled in the experiments, it is not clear how to apply game theory, as it currently exists, to the experiments. These are games of incomplete information: they are not games of imperfect information.[4] One could try to ignore this problem and assume that there is, at some level, common knowledge. However, this leads to the second difficulty.

[3]In an important recent paper, Friedman[4] has filled in this gap with a model based on search-theoretic principles. We discuss this interesting model in more detail below in Section 4.D.

[4]In his seminal articles, Harsanyi[7] was very careful to differentiate between incomplete and imperfect information. Beginning with a game of incomplete information, he converted it to a game of imperfect information *from player i's point of view.* There was no guarantee, absent an assumption of objective common knowledge, that the game from i's point of view would be the same as the game from j's point of view. Therefore, without the common knowledge hypothesis, players can be surprised *on the equilibrium path*: they discover that they are in an entirely different game tree than they thought they were. At this point Bayes' rule provides no guidance and players can do anything. With this freedom, one can make any outcome of the experiment a Bayes-Nash (cont'd.)

With an assumption of common knowledge, the natural model is the Bayes-Nash equilibrium. If the subjects are risk neutral, we know from Gresik and Satterthwaite[5] and the revelation principle that any Bayes-Nash equilibrium has the property that no extra-marginal units are traded when subjects only own one unit of the commodity. Yet in the experiments, extra-marginal units are often traded (see, e.g., the data from IPDA14 in Appendix A). If the subjects are risk averse, then we know from Ledyard[10] that virtually anything can be an equilibrium. If risk attitudes are not controlled for (see Roth and Malouf[17]), then the game-theoretic model explains everything.

Our third difficulty with the game-theoretic approach is that, as far as we know, no one has solved for an equilibrium of the appropriate game. Wilson[23] has found strategies for a one-shot version of the DOA which satisfy the necessary conditions for a Bayes-Nash equilibrium. However, Wilson's model predicts that the rank correlation coefficient between the order of true values and trades is one which, as we noted above, is strongly at odds with the data. But the experimental DOAs are repeated, common knowledge is not controlled for and subjects may not be risk neutral as Wilson assumes. Under these circumstances, it is not fair to compare Wilson's predictions with the data. It is also unfair to expect much from the general approach.

Since neither the Marshallian model nor the Friedman model, nor any currently available game-theoretic model appears to be appropriate as a positive theory, and since the goal of building an appropriate game-theoretic model has eluded us and others, development of an alternative model seems warranted. We turn to that next.

4. A POSITIVE THEORY

A. PRELIMINARIES

A participant in a double oral auction experiment has a complex decision problem. He must decide when to bid, how much to bid, and whether or not to accept the trades offered by other subjects. Further, all of these decisions must be made with very imperfect information. The subject does not know the payoffs or expectations of other agents, he does not know the terms of trade that will be available to him in the future, and he does not know the effect of his actions on the actions of others. This is a very complex interactive decision problem with incomplete information in which individuals must choose bidding and acceptance strategies. To place some structure on this problem, we first introduce some notation and definitions concerning the data known to both the experimenter and us.

The *true payoffs or values given to buyers* are integers and are ranked as $V^1 \geq V^2 \geq \ldots \geq V^n \geq 0$, where V^i is the ith highest value and there are n units. A buyer

equilibrium of the *incomplete* information game without objective common knowledge. We show how to do this in footnote 12.

b will be assigned a subset of these units $V^{b1} \geq V^{b2} \geq \ldots \geq V^{bB}$ and will trade them one at a time in the sequence $b1, b2, \ldots, bB$. No recontracting is allowed. The *true costs given to sellers* are integers and are ranked as $0 \leq M^1 \leq M^2 \leq \ldots \leq M^m$, where M^j is the cost of the jth unit and there are m units. A seller will be assigned a subset of these units and will trade them one at a time. No recontracting is allowed. It should be noted that typically the values, V^i, and costs, M^i, are assigned once and remain fixed. Each buyer (seller) knows only his own values (costs) and no participant is given any information as to how these values and costs were chosen. There is no basis for common knowledge assumptions about independence of values or their distributions. Consequently, we neither make such assumptions nor use these concepts in our theory.

Market period or days for an experiment are indexed by $d = 1, 2, \ldots$. The *time remaining in any given* day is indexed by $t = 0, 1, \ldots, T$. Contract prices, bids, and offers are in integer units in the interval $[0, H]$, where $H < \infty$ is some arbitrarily selected upper bound above V^1 and M^m, and during any particular day, d, each participant observes all contract prices, bids, and offers.[5]

To summarize, each buyer knows the rules of the auction, the value of his own units, and the sequence, timing, amount, and identity of all past bids, offers, and contracts. It is these data alone on which the buyer can base his decisions to bid and to accept. A symmetric remark applies to each seller.

B. AN INTUITIVE LOOK

We adopt the spirit of both revealed preference theory and demand-supply analysis by placing assumptions on individual behavior which, we believe, are consistent not only with optimal behavior but also with a vast range of "boundedly rational" rules-of-thumb. We do not model how agents should make their decisions. Instead, we provide criteria which, we believe, sensible individuals in these markets act as if they satisfy. This allows us to construct a theory which is robust to a wide variety of individual behaviors and yet which is reasonably sharp in its predictions about the data. We model "reduced form" behavior by decomposing the decision problem into three main elements: expectations, reservation prices, and bidding strategies. These are most easily explained in reverse order.

Assume that at each instant of time, there is for each buyer (seller) *a reservation price*, possibly different from his true value, which summarizes his willingness to bid up (offer down) to that price or to accept any offer up (bid down) to that

[5]Each participant also observes the timing of each contract, bid, and offer. It is highly probable that the timing of these events is an important piece of information which affects the actions of the buyers and sellers. However, the level of complexity required to incorporate timing into the model seems to outweigh the gains to be achieved. Thus, we ignore it throughout the paper.

price.[6] If each participant has such a reservation price as a function of time, the buyers' side of the double auction can then be thought of as proceeding like an ascending bid (English) auction, *with these reservation prices substituting for the true values.* After some period of time, the outstanding bid will always be held by the buyer with the highest reservation price (not necessarily the highest untraded value), and that bid will be at least as high as the second-highest reservation price. Otherwise, the holder of the second-highest reservation price will bid, causing the holder of the highest reservation price to rebid, and so on. We find it unnecessary to explicitly model this process, and we assume that it occurs instantaneously. Thus, all observed bids will be the reduced form results of the above English auction. Since we have also assumed that the buyer is willing to accept any price lower than the reservation price, an acceptance of an offer will occur whenever that offer is lower than the highest reservation price of a buyer. To make the theory as simple as possible, we assume the buyer with the highest reservation price moves instantaneously faster than any other buyer. (This is only a restriction if offers jump down in large discrete increments and several buyers have similar reservation prices—a situation likely to occur only in the opening minutes of any trading day.) This intuitive view of the bidding is formalized in Assumption 1 below. Sellers' offers are viewed symmetrically in Assumption 1$'$.

Since bids and offers depend on reservation prices and not directly on the induced values, bids and offers ultimately depend on the relationship between reservation prices and the data observed by each agent. This relationship is assumed to depend on two principles of learning. Firstly, it is true under Assumption 1 that whenever the bids and acceptance prices of a buyer are higher than were necessary to complete a transaction, the buyer completes a trade but overpays. We assume that a buyer will realize that he overpaid and will, during the next auction, lower his reservation price. If it is not lowered too much, the buyer should still be able to complete a transaction but at a better price. Secondly, it is true that if a buyer waits too long to bid or, what is the same thing, maintains too low a reservation price during the day, then that buyer may not complete a transaction even though profitable ones are available. We assume that if a buyer could have purchased a unit at less than its value to him, V^i, but did not, then that buyer will realize he underbid and will, either that day or during the next auction, raise his reservation price at each time of day. It is the delicate balance between "paying too much" and "not offering to pay enough" which the buyers must learn in order to be successful in the auction. We do not explicitly model this learning process; instead, we provide assumptions about reservation price behavior which, if satisfied, reflect these learning principles. We summarize this rather simple intuition in Assumption 2 below.

[6] A more sophisticated theory might distinguish between the amount a buyer is willing to bid and the lowest offer he would accept. In particular, buyers may not be willing to bid up to their reservation price (see Wilson[23]). This distinction could be easily incorporated into our model, but it is not apparent that it would add to the explanatory power of the model.

C. BIDDING BEHAVIOR

We start our description of the formal theory with the introduction of a hypothesis concerning the existence of the key unobservable of our model. It is important to realize that we treat reservation prices in this paper in the way that preferences are generally treated in economics. We cannot observe whether subjects really compute reservation prices; we can only assume they act as if they do. For a coherent theory, the reservation prices may need to be related in a systematic way to the true values but, *a priori*, do not need to be.

ASSUMPTION 0: RESERVATION PRICES. For each buyer unit and seller unit, there is an (unobservable) reservation price at each day d and time t, denoted $r_d^i(t) \in R^1$ for buyers and $s_d^j(t) \in R^1$ for sellers.

Assumption 0 only contains notation. To link the unobservables to the data, we need to tie the bids and acceptances to the reservation prices, and then to tie the reservation prices to the true values and costs. As we indicated in the previous section, this is done by assuming that, given reservation prices, bids and acceptances are the reduced form of English auction behavior.

ASSUMPTION 1: BUYERS' BIDS AND ACCEPTANCES.

i. $b_d(t)$, the current outstanding bid in day d, with time t left, is held by buyer i^* where $r_d^{i^*}(t) \geq r_d^i(t)$, for all $i = 1, \ldots, n$.

ii. $b_d(t) \leq r_d^{i^*}(t)$.

iii. $b_d(t) \geq r_d^i(t)$, for all $i \neq i^*$.

iv. Buyer i^* accepts the current outstanding offer, $o_d(t)$, if and only if $o_d(t) \leq r_d^{i^*}(t)$. No other i accepts $o_d(t)$.

Simply stated, at each point in time, the current bid is held by the buyer with the highest reservation price—*not necessarily the buyer with the highest true value.* This bid lies below that reservation price and above the second-highest reservation price. Under Assumption 1, and 1′ below, trades always occur between the buyer with the highest reservation price and the seller with the lowest reservation price. We emphasize that *these need not be the buyer with the highest value and the seller with the lowest cost* since the English auction is based on reservation prices and not on the "true values," V^i and M^i.

For completeness, we make an assumption on the offers and acceptances of sellers that is symmetric with that made for buyers. The only difference is that we have arbitrarily assumed that if seller j^* is willing to accept $b_d(t)$ and buyer i^* willing to accept $o_d(t)$, then the buyer accepts first. We could reverse this without affecting the conclusions to come.

ASSUMPTION 1′: SELLERS' OFFERS AND ACCEPTANCES.

i. $o_d(t)$, the current outstanding offer in day d, with time t left, is held by seller j^* where $s_d^{j^*}(t) \leq s_d^j(t)$, for all $j = 1, \ldots, m$.

ii. $o_d(t) \geq s_d^{j^*}(t)$.

iii. $o_d(t) \leq s_d^j(t)$, for all $j \neq j^*$.

iv. Seller j^* accepts the outstanding bid, $b_d(t)$, if and only if $b_d(t) \geq s_d^{j^*}(t)$ and buyer i^* does not accept $o_d(t)$. No other j accepts $b_d(t)$.

We do not yet have a testable theory since, given any sequence of bids and contracts, it is possible to construct a sequence of reservation prices which, under Assumption 1, would imply the given data precisely. Unless we place some restrictions on the reservation prices, we can explain anything, and therefore nothing.

D. RESERVATION PRICE FORMATION

We now tie the theory down by restricting reservation price behavior in a way which relates it to observable data. This is the way in which we connect bids, contract prices, and the sequence of trades to the initial data known by the experimenter and, thus, provide testable propositions about these auctions.

Reservation prices are assumed to be formed in accordance with the intuitive principles outlined in Section 4.B. We begin by assuming that a buyer's expectations in any period are based on the prices of the previous period. In particular, we assume that the support of the buyer's expectations is the set of prices bounded by the maximum of last period's highest contract price or highest bid, and the minimum of last period's lowest contract price or lowest offer. Based on these expectations, reservation prices are formed over time as follows: (a) for most of a trading day, one's reservation price lies below the true value, V^i, and within the support of the expectations (when this is feasible), (b) if possible, the reservation price is actually below the maximum price in the support since the buyer does not want to "overpay," (c) eventually, if no contract is agreed to, buyers will cave in and let the reservation price rise and approach the maximum price in the support, and (d) if still no contract is completed, the reservation price will rise higher than even the maximum in the support of the expectations.

The sequence of actions (a), (b), and (c) are consistent with optimal behavior in finite-time, nonstrategic search models.[7] If a buyer believes that offers are identically and independently distributed on $[\underline{P}, \overline{P}]$, no matter what he does, and that he will receive a finite number of draws of offers with replacement, it is really

[7] For examples of this literature see Gronau,[6] Lippman and McCall,[11,12] Mortensen,[13] and Cox and Oaxaca.[2]

easy to show that his reservation price satisfies (a), (b), and (c).[8] A buyer who is certain that he can complete a trade at $\overline{P}$ will only move his reservation price to $\overline{P}$ at $t = 0$, the end of the day. If he could not complete a trade at this point, he would presumably be willing to pay more than $\overline{P}$ as he now knows that his beliefs are incorrect. In (d) we assume that he reaches $\overline{P}$ before $t = 0$, perhaps because he is not certain that a trade can be completed at $\overline{P}$.

Before formalizing our assumption on reservation prices, we need to introduce some notation. If a trade occurs at time t of day d, we let $c_d(t)$ be the contract price. Then for each day $d > 1$, let $\underline{P}_d = \text{Min}\{o_{d-1}(t), c_{d-1}(t) : t = 0, \ldots, T\}$ and $\overline{P}_d = \text{Max}\{b_{d-1}(t), c_{d-1}(t) : t = 0, \ldots, T\}$. We assign $[\underline{P}_1, \overline{P}_1] = [0, H]$. The interval $[\underline{P}_d, \overline{P}_d]$ is interpreted as the support of traders' price expectations in day d. Let $\Delta P_d = \overline{P}_d - \underline{P}_d$.

ASSUMPTION 2: BUYER'S RESERVATION PRICE FORMATION. For all buyers $i = 1, \ldots, n$:

i. If i has traded (accepted an offer or had a bid accepted) in day d before time t, then $r^i_d(t) = 0$.

ii. For each day d there is time $\hat{t}^i_d > 1$ such that, if i has not traded in d before t, then:

   a. For all $t > \hat{t}^i_d$;
      $\text{Min}\{V^i, \overline{P}_d\} > r^i_d(t) \geq \underline{P}_d$ if $\Delta P_d > 1$ and $V^i > \underline{P}_d$;
      $\text{Min}\{V^i, \overline{P}_d\} \geq r^i_d(t) \geq \text{Min}\{V^i, \underline{P}_d\}$ otherwise.

   b. $r^i_d(\hat{t}^i_d) = \text{Min}\{V^i, \overline{P}_d - 1\}$.

   c. For all $t < \hat{t}^i_d$;
      $r^i_d(t) = \text{Min}\{V^i, b_d(t+1) + 1\}$ if
      $b_d(t+1) \in \{\overline{P}_d, \overline{P}_d - 1\}$ and $b_d(t+1)$ unaccepted;
      $r^i_d(t) \in \{r^i_d(t+1), \text{Min}\{V^i, b_d(t+1) + 1\}\}$ if
      $b_d(t+1) > \overline{P}_d$ and $b_d(t+1)$ unaccepted;
      $r^i_d(t) = r^i_d(t+1)$ otherwise.

Assumption 2(i) sets the reservation price for traded units to zero to indicate that they have left the market. The conditions in Assumption 2(ii)(a) embody the intuition that, as a result of learning, reservation prices will not be "too high" early in the trading day. The conditions in Assumption 2(ii)(b,c) embody the intuition that, towards the end of the day, if the buyer has not completed a transaction, then

[8]Recently, Friedman[4] has provided a model of the double oral auction in which agents are Bayesian and expected utility maximizers who ignore the strategic feedback effects of their own actions. He derives reservation strategies which provides a choice-theoretic underpinning to our model. (See Friedman,[4] p. 57–58.) The main difference between his model and ours lies in our use of bounds $\overline{P}$ and $\underline{P}$ on the support of possible prices (he assumes priors are positive over all prices $[0, H]$) and property (d) which describes what happens when, as a Bayesian, the buyer is surprised that no offer is in $[\underline{P}, \overline{P}]$. (Friedman's agents are never surprised.)

that buyer will learn to raise his reservation price slowly. Towards the end of the day, reservation prices will not be "too low."[9]

To complete the model we make a symmetric assumption about sellers' reservation prices which we call Assumption 2'.

ASSUMPTION 2': SELLERS RESERVATION PRICE FORMATION. For all sellers $j = 1, \ldots, m$:

i. If j has traded (accepted a bid or had an offer accepted) in day d before time t, then $s_d^j(t) = \overline{P}$.

ii. For each day d there is a time $\overline{t}_d^j > 1$ such that, if j has not traded in day d before time t, then:

   a. For all $t > \overline{t}_d^j$;
   $\text{Max}\{M^j, \overline{P}_d\} \geq s_d^j(t) > \text{Max}\{M^j, \underline{P}_d\}$ if $\Delta P_d > 1$ and $M^j < \overline{P}_d$;
   $\text{Max}\{M^j, \overline{P}_d\} \geq s_d^j(t) \geq \text{Max}\{M^j, \underline{P}_d\}$ otherwise.

[9] Our assumptions on reservation prices can also be stated with a simple, Markov information structure. The only information used from past days is $(\underline{P}_d, \overline{P}_d)$. The only information used from the current day is the previous bid, offer, and contract price, and an indicator of whether the individual has traded. Let the trade indicator for individual i at time t of day d be $h_d^i(t)$,

$$\text{where } h_d^i(T) = 0 \text{ and } h_d^i(t-1) = h_d^i(t) + \begin{cases} 1 & \text{if } i \text{ trades at } t; \\ 0 & \text{otherwise.} \end{cases}$$

The information from the current day is then $I_d^i(t) = \{b_d(t), o_d(t), c_d(t), h_d^i(t)\}$ for any $t < T$ and $I_d^i(T) = \phi$. Any trader's reservation price evolves according to a transition probability which is time and information dependent and parameterized by the trader's value and $(\underline{P}_d, \overline{P}_d)$. To simplify the notation we drop the indices i and d, and we consider the problem from a typical buyer's point of view. The distribution of the buyer's reservation price at time t, $r(t)$, is given by $R(I(t+1),\ t,\ r(t+1))$. Let $A = \{r \in [0, \overline{P}]\ :\ \text{Min}(V, \overline{P}_d - 1) \geq r \geq \text{Min}(V, \underline{P}_d)$ if $\Delta P_d > 1$ and $\text{Min}(V, \overline{P}_d) \geq r \geq \text{Min}(V, \underline{P}_d)$ otherwise$\}$. Then we can write Assumption 2 equivalently as

Assumption 2*: Buyers' Reservation Price Formation.

(i) If $h(t+1) = 1$, then $R(I(t+1), t, r(t+1))(0) = 1$.
(ii) If $r(t+1) \in A$, $h(t+1) = 0$, and $b(t+1) < \overline{P}_d - 1$, or $b(t) = c(t)$, then $\int_A dR(I(t+1), t, r(t+1)) = 1$.
(iii) For any $r \geq \overline{P}_d - 1$, $R(I(t+1), t, r(t+1))(r) = 0$ if $r > r(t+1) + 1$.
(iv) If $r(t+1) < V$, $b(t+1) \in [\overline{P}_d - 1, \overline{P}_d]$ and $b(t+1) \neq c(t+1)$, then $R(I(t+1), t, r(t+1))(r(t+1)+1) = 1$.
(v) There is a $t^* > 1$ such that if $h(t^*+1) = 0$, then

$$\sum_{r \geq \text{Min}(\overline{P}_d - 1, V)} R(I(t^*+1), t^*, r(t^*+1))(r) = 1.$$

Either 2 or 2* can be used in the rest of the paper; we use 2.

b. $s_d^j(\bar{t}_d^j) = \text{Max}\{M^j, \underline{P}_d + 1\}$.

c. For all $t < \bar{t}_d^j$;
$s_d^j(t) = \text{Max}\{M^j, o_d(t+1) - 1\}$ if
$o_d(t+1) \in \{\underline{P}_d, \underline{P}_d + 1\}$ and $o_d(t+1)$ unaccepted;
$s_d^j(t) \in \{s_d^j(t+1), \text{Max}\{M^j, o_d(t+1) - 1\}\}$ if
$o_d(t+1) < \underline{P}_d$ and $o_d(t+1)$ unaccepted;
$s_d^j(t) = s_d^j(t+1)$ otherwise.

Assumptions 1, 1′, 2, and 2′ constitute the full set of premises for our theory. One obvious and intended omission is any (direct) tie between reservation prices and valuations—other than the obvious constraint that buyer i's reservation price be less than i's value. In particular, we make no assumptions about the relative rankings of values and reservation prices.[10] While such an assumption might tighten the predictions of the model, most of the theorems we are interested in and most of the implications consistent with the data do not require it. We think this is an attractive feature of the model in that decision making is decentralized. Agents need know nothing about each other but only need to look at observable data to decide what to bid and whether to trade. Convergence to competitive equilibrium under these conditions simply highlights the robustness of the double oral auction as a market institution. Any monotonic link between values and (even randomized) reservation prices would require some coordination between the agents. This in turn would seem to require some prior beliefs on values and a common strategy. This is unnecessary and inconsistent with the spirit of our analysis.

We have made two implicit assumptions which should be recognized. First of all, we assume that each buyer's and seller's behavior is independent of the total number of participants in the market. That is, a buyer's choices of bids and acceptances are the same whether he is a monopolist or one of 100 buyers. Although this runs counter to conventional economics, experimental evidence suggests that if the number of buyers and the number of sellers are both greater than two, then this assumption is satisfied. Further, even if there is a single seller, what little evidence there is suggests that the model we propose may still be appropriate. We leave as an open empirical question just how few participants, if any, are needed before our theory is not applicable.

The second implicit assumption is that buyers and sellers with multiple units to purchase or sell will decide on strategies for each unit separately. That is, the bids and acceptances a buyer makes for his, say, highest valued (first) unit are assumed to be independent of the total number of units he may want to buy. This is not "rational behavior," but the interaction effects are difficult to model (we know of

[10] An example of such an assumption is $[V^i > V^j] \Rightarrow [r_d^i(t) \geq r_d^j(t)]$ $\forall$ d, t, i, j. This particular hypothesis, which is closely related to the hypotheses of the Marshallian model and game-theory models, yields predictions seriously at odds with the data. See Section 7 for other possible connections between valuations and reservation prices.

no literature which does this[11]). The simplicity this assumption gives the theory is, we feel, well worth the price.[12]

A question that naturally arises is whether optimal behavior for a game-theoretic formulation of the DOA is consistent with our behavioral rules. Since this would be a dynamic incomplete-information game, and since the common knowledge that is an integral part of recent game theory is not controlled for in the experiments, current theory provides little guidance. However, as we indicated in section 3, footnote 4, if one recognizes the important difference between a game of *imperfect* information with *objective common knowledge* priors and a game of *incomplete* information, then one can understand how Bayesian game theory is consistent with our rules. In particular, even assuming risk neutrality, we can construct a vector of strategies, one for each agent, such that i's component of a Bayes equilibrium for the DOA game *from i's subjective point of view*, such that revisions in beliefs as the game is played using these strategies satisfy the Bayes' rule *on nonzero probability events*, and such that the trades, bids, and offers are consistent with Assumptions 1, 1′, 2, and 2′. Because of the absence of objective common knowledge, agents may be surprised, even on the equilibrium path, but neither the Bayes rule nor the Bayes-Nash equilibrium prevents this possibility.[13] Nevertheless, we do not believe that this game-theoretic behavior is what subjects in DOA experiments are really doing. Thus we prefer to analyze our more general model

[11]Noussair[15] has recently solved this problem for a uniform-price sealed-bid mechanism, but it remains unsolved in general.

[12]Holt, Langan, and Villamil[8] report a series of experiments in which traders had multiple units with payoffs structured to give some traders market power on some units. Their data are nonetheless reasonably consistent with the predictions of our theory. So our implicit assumption that traders decide on strategies for each unit separately seems not to be at odds with the facts.

[13]We describe the equilibrium from one agent's point of view. Suppose that every trader believes that values are drawn such that two or more buyers and sellers have p^* as their value and that $S(p^*) = D(p^*)$. (This is the way a typical experiment is set up. The only new feature here is having all traders believe the same p^* and believe it with certainty.) Further, suppose that this subjective belief is common knowledge. Consider a buyer's strategy which is to bid p^* if $V \geq p^*$, to bid V if $p^* > V$, to accept any offer $o \leq p^*$ if $o \leq V$, and to reject any offer $o > p^*$. The buyer is assumed to follow this strategy forever and sellers are assumed to follow symmetric strategies. To complete the description of an equilibrium, we have to describe the updating of beliefs. With the proposed strategies any bid above p^* or offer below p^* is a zero probability event so the Bayes rule has no implications in this case. Let p^*_d be the common believed price for day d with $p^*_o = p^*$. If no zero-probability events are observed in d, then $p^*_{d+1} = p^*_d$. If any zero-probability event is observed, let p^*_{d+1} be one of the zero-probability prices in $(\underline{P}_d, \overline{P}_d)$. The claim is that these strategies and updating rules define a Bayes-Nash equilibrium. Consider any buyer. He knows that his bids can affect beliefs only if the bids are above p^*_d. But this is undesirable, at least until the end of the day, as a high bid has him paying more than he believes is necessary and can only raise future prices. At the end of the day, a buyer who does not hold the outstanding bid of p^*_d knows that he cannot trade at p^*_d as he believed. (This will happen if and only if there are two or more buyers who have not yet traded at $T - 1$. Further, the buyer who will not be able to trade as expected knows who he is.) This is a zero-probability event (according to his belief) so he can plan to do anything in this contingency. To complete the description of the equilibrium strategy, we assume that he bids $p^*_d + 1$ if $p^*_d + 1 \leq V$ and V otherwise. (cont'd.)

recognizing that one set of behavioral rules satisfying Assumptions 1, 1′, 2, and 2′ are consistent with a game-theoretic treatment.

Now a final observation. We believe that there are traders in the experiments whose behavior is, at least for a few iterations, vastly different from behavior which would be consistent with our assumptions. In particular there are traders, who hold out for a highly profitable trade to the end of the day even though they never complete one. These traders usually modify their behavior after a few days. Those who do not lose a considerable amount of opportunity income. We do not attempt to explain their irrationality.

We turn now to the derivation of a number of testable implications of the theory. We then confront these with the data from a small number of representative experiments. At that point, the reader should be able to decide whether or not our model offers a realistic description of actual behavior in double auctions.

5. THEOREMS

In this section we trace through some of the implications of our theory. As will become apparent, most of the action will occur when there is an "excess demand or supply" of two or more units remaining in the auction, as there are then competitive pressures on bids and offers. Thus we are interested in the following concepts.

DEFINITION Let $D^c(P) = \#\{V^i \geq P\}$, $D^o(P) = \#\{V^i > P\}$, $S^c(P) = \#\{M^j \leq P\}$, and $S^o(P) = \#\{M^j < P\}$. Let $P_* = \min\{P : D^c(P) \leq S^o(P) - 2\}$ and $P^* = \max\{P : S^c(P) \leq D^o(P) - 2\}$.

P_* is the minimum price at which there is an excess supply of two units and P^* is the maximum price at which there is an excess demand of two units. For IPDA14 in Appendix A, P_*=4.31 and P^*=3.99. In Figure 2, Appendix B, P_*=101 and P^*=99. An excess of two is important to provide the competitive forces that will drive prices. To see this, consider the following propositions. All results are stated under Assumptions 0, 1, 1′, 2, and 2′. Remember $\underline{P}_d$ is the lowest contract price or offer observed during day $d-1$ and is a "lower bound" on the agent's support in day d. $\underline{P}_{d+1}$ will be the lowest contract price or offer observed during day d.

Now consider a buyer who contemplates a defection from the proposed equilibrium by refusing to trade. This can only be valuable if it changes beliefs. But any one buyer knows he can cause only one seller to remain untraded so this strategy will result in no lower offers. This structure describes a Bayes-Nash equilibrium which produces behavior consistent with Assumptions 1, 1′, 2, and 2′. We have not explored the possibilities for refinements.

LEMMA 1:

a. If $\underline{P}_d \geq P_*$, then $\underline{P}_{d+1} < \underline{P}_d$. If $\underline{P}_d < P_*$, then $\underline{P}_{d+1} < P_*$.

b. If $\overline{P}_d \leq P^*$, then $\overline{P}_{d+1} > \overline{P}_d$. If $\overline{P}_d > P^*$, then $\overline{P}_{d+1} > P^*$.

That is, there are competitive forces driving minimum contract prices below P_* and keeping them there. These same forces drive maximum contract prices above P^* and keep them above P^*.

PROOF: We prove (a); the proof of (b) is symmetric.

Suppose $\underline{P}_d \geq P_*$ and $\underline{P}_{d+1} \geq \underline{P}_d$. As $\underline{P}_d \geq P_*$, we have $D^c(\underline{P}_d) \leq S^o(\underline{P}_d)-2$. The number of trades in day d is no more than $D^c(\underline{P}_d)$ as by hypothesis all trades have been at price $\underline{P}_d$ or above. Thus at t=2 there are at least two sellers j and j' with M^j, $M^{j'} < \underline{P}_d$ who have not yet traded. Then by applying Assumptions $2'$(ii)(c) and $1'$ repeatedly, we have $o_d(0) \leq \underline{P}_d - 1$. But then $\underline{P}_{d+1} < \underline{P}_d$ which contradicts $\underline{P}_{d+1} \geq \underline{P}_d$.
Suppose $\underline{P}_d < P_*$ and $\underline{P}_{d+1} \geq P_*$. Then all trades have been at prices at or above P_*. A minor modification of the argument above then yields a contradiction. QED

LEMMA 2: Suppose $\Delta P_d > 1$; then

a. If $D^o(\underline{P}_d) \geq S^o(\overline{P}_d)$, then $\underline{P}_{d+1} > \underline{P}_d$.

b. If $S^o(\overline{P}_d) \geq D^o(\underline{P}_d)$, then $\overline{P}_{d+1} < \overline{P}_d$.

That is, there are competitive pressures driving (the lowest) contract prices up if they are too low relative to the highest prices and driving maximum contract prices down if they are too high relative to the lowest prices.

PROOF: We prove (a); the proof of (b) is symmetric.

Suppose that $D^o(\underline{P}_d) \geq S^o(\overline{P}_d)$ and $\underline{P}_{d+1} \leq \underline{P}_d$. Then there exists a time t' such that either $o_d(t') = \underline{P}_{d+1}$ or $c_d(t') = \underline{P}_{d+1}$. Since $\underline{P}_{d+1} \leq \underline{P}_d$ and $\Delta P_d > 1$, it follows from Assumption $2'$(ii)(a),(c) that there exists a time $\hat{t} > t'$ such that $o_d(\hat{t}) = \underline{P}_{d+1}$ was not accepted. Therefore, as $\Delta P_d > 1$, Assumption 1(ii)(a) implies that all units $V^i > \underline{P}_d$ have been traded before time $\hat{t}$. So the number of trades before time $\hat{t}$ is at least $D^o(\underline{P}_d)$ which is $\geq s^o(\overline{P}_d)$. Then as Assumption $2'$(ii)(a) implies that if $\Delta P_d > 1$, all $M^j < \overline{P}_d$ trade before any $M^j \geq \overline{P}_d$, we know that all $M^j < \overline{P}_d$ have been traded before time $\hat{t}$. So all $M^j \leq \underline{P}_d$ have been traded before time $\hat{t}$. Then by Assumption $2'$, $s^j_d(t)$ and

$o_d(t) > \underline{P}_d$ for all $t \leq \hat{t}$ and all j. This contradicts either $o_d(t') = \underline{P}_{d+1} \leq \underline{P}_d$ or $c_d(t') = \underline{P}_{d+1} \leq \underline{P}_d$. QED

Buyers' reluctance to pay too much and sellers' reluctance to accept too little eventually force minimum and maximum contract prices closer together. Of course, the difference between maximum and minimum contract prices does not necessarily decrease every day. In a day where there is excess demand at the upper bound $\overline{P}$, prices may rise, but they will not go above the cost of unit number $D^o(\underline{P})$. This occurs because units up to $D_o(\underline{P})$ trade first (if $\Delta P > 1$) and these can all be traded at prices no more than $M^{D^o(\underline{P})}$. Thus the statistic that falls, or at least does not rise, in every period is the maximum of $\overline{P}$ and $M^{D^o(\underline{P})}$.

DEFINITION Let $u_d = \max\{\overline{P}_d, M^{D^o(\underline{P}_d)}\}$ and $\ell_d = \min\{\underline{P}_d, V^{S^o(\overline{P}_d)}\}$.

In IPDA14, Appendix A, suppose $\underline{P}_d$=4.00 and $\overline{P}_d$=4.10. Then u_d=4.20 since $D^o(\underline{P}_d)$=6 and $\ell_d = \underline{P}_d$=4.00 since $S^o(\overline{P}_d)$=4. The next lemma is useful in the proof of the results of main interest further on.

LEMMA 3: If $\Delta P_d > 1$, then $u_{d+1} \leq u_d, \ell_{d+1} \geq \ell_d$, and $|u_{d+1} - \ell_{d+1}| < |u_d - \ell_d|$.

PROOF: There are two cases to consider: (1) $S^o(\overline{P}_d) \geq D^o(\underline{P}_d)$ and (2) $D^o(\underline{P}_d) \geq S^o(\overline{P}_d)$. We prove the lemma under case 1; the proof under case 2 is symmetric. We first need to establish:

CLAIM 1: If $\Delta P_d > 1$ and $S^o(\overline{P}_d) \geq D^o(\underline{P}_d)$, then $\underline{P}_{d+1} \geq \ell_d$.

PROOF: Suppose that $\Delta P_d > 1$, $S^o(\overline{P}_d) \geq D^o(\underline{P}_d)$, and $\underline{P}_{d+1} < \ell_d$.

From the definition of $\underline{P}_{d+1}$ we know that there is a time t' in day d such that $o_d(t') = \underline{P}_{d+1}$ or $c_d(t') = \underline{P}_{d+1}$. Then as $\underline{P}_{d+1} < \ell_d \leq \underline{P}_d$, there must be a time $\hat{t}$ in day d such that $o_d(\hat{t}) = \ell_d$ was not accepted. This implies that all $V^i \geq V^{S^o(\overline{P}_d)}$ have traded before time $\hat{t}$. So the number of units traded before time $\hat{t}$ is at least $S^o(\overline{P}_d)$. By Lemma 2(b) we have $\overline{P}_{d+1} < \overline{P}_d$. So the number of units traded in day d is no more than $S^o(\overline{P}_d)$. Thus the number of units traded in day d, before time $\hat{t}$, is $S^o(\overline{P}_d)$. So all $M^j < \overline{P}_d$ have traded before time $\hat{t}$. Then there does not exist a seller unit $M^j \leq \underline{P}_{d+1} < \ell_d$ to offer $o_d(t') = \underline{P}_{d+1}$ or accept a contract at $c_d(t') = \underline{P}_{d+1}$. This contradicts $\underline{P}_{d+1} < \ell_d$. The proof of Lemma 3 now follows directly from Claims 2 and 3.

CLAIM 2: If $\Delta P_d > 1$ and $S^o(\overline{P}_d) \geq D^o(\underline{P}_d)$, then $\ell_{d+1} \geq \ell_d$.

PROOF: By Lemma 2, $\overline{P}_{d+1} < \overline{P}_d$. So $S^o(\overline{P}_{d+1}) \leq S^o(\overline{P}_d)$.

This implies that $V^{S^o(\overline{P}_{d+1})} \geq V^{S^o(\overline{P}_d)}$. By claim 1, $\underline{P}_{d+1} \geq \ell_d$. Now $\ell_{d+1} = \text{Min}\{\underline{P}_{d+1}, V^{S^o(\overline{P}_{d+1})}\} \geq \text{Min}\{\underline{P}_{d+1}, V^{S^o(\overline{P}_d)}\} \geq \ell_d$.

CLAIM 3: If $\Delta P_d > 1$ and $S^o(\overline{P}_d) \geq D^o(\underline{P}_d)$, then $u_{d+1} < u_d$.

PROOF: By Lemma 2(b), $\overline{P}_{d+1} < \overline{P}_d$ and by Claim 1, $\underline{P}_{d+1} \geq \ell_d = \text{Min}\{\underline{P}_d, V^{S^o(\overline{P}_d)}\}$. Thus $D^o(\underline{P}_{d+1}) \leq \text{Max}\{D^o(\underline{P}_d), S^o(\overline{P}_d)\} = S^o(\overline{P}_d)$. So $M^{D^o(\underline{P}_{d+1})} \leq M^{S^o(\overline{P}_d)} < \overline{P}_d$. Then $u_{d+1} = \text{Max}\{\overline{P}_{d+1}, M^{D^o(\underline{P}_{d+1})}\} < \overline{P}_d \leq u_d$. So $u_{d+1} < u_d$. QED

The forces embodied in Lemmas 1, 2, and 3 serve to drive contract prices together and into the interval $[P^*, P_*]$. If supply and demand balance at this point, prices will stay in this interval. Before proceeding to Theorem 1, we need to show that the interval is well defined.

CLAIM 4: $P_* \geq P^*$.

PROOF: Suppose $P^* \geq P_*$. Then $S^c(P_*) \leq D^o(P_*) - 2$ and $D^c(P_*) \leq S^o(P_*) - 2$. So $D^c(P_*) + 2 \leq S^o(P_*) \leq S^c(P_*) \leq D^o(P_*) - 2$. This implies $D^c(P_*) < D^o(P_*)$ which is false.

THEOREM 1: If $D^c(P^*) = S^c(P_*)$, then there exists a day $d^* < \infty$ such that $P^* \leq \underline{P}_d < P_*$ and $P^* < \overline{P}_d \leq P_*$ for all $d \geq d^*$.

PROOF: As the price set—the integers in $[0, \overline{P}]$—is finite, Lemma 1 implies that there is finite day $\overline{d}$ such that $\underline{P}_d < P_*$ and $\overline{P}_d > P^*$ for all $d \geq \overline{d}$. Then by Lemma 3 there is a finite day $d^* \geq \overline{d}$ such that $[\underline{P}_{d*}, \overline{P}_{d*}] \subseteq [P^*, P_*]$ and $\Delta P_{d*} \leq 1$.

We now prove the theorem by an induction argument. Suppose $[\underline{P}_d, \overline{P}_d] \subseteq [P^*, P_*]$ for some day $d \geq d^*$. We need to show that this implies $[\underline{P}_{d+1}, \overline{P}_{d+1}] \subseteq [P^*, P_*]$. Suppose not, say $\underline{P}_{d+1} < P^*$. Then there is a time t' in day d such that $o_d(t') < P^*$ or $c_d(t') < P^*$. As $P^* \leq \underline{P}_d$, there is a time $\hat{t} \geq t'$ such that $o_d(\hat{t}) = P^*$ was not accepted. As $P^* \leq \underline{P}_d$, this implies that all units $V^i \geq P^*$ have traded before time $\hat{t}$. So the number of units traded is at least $D^c(P^*) = S^c(P_*)$. As $P_* \geq \overline{P}_d$, this implies that all units $M^j \leq P_*$ have traded before time $\hat{t}$. Then there is no seller with a unit $M^j < P^* < P_*$ to offer $o_d(t') < P^*$ or to accept $c_d(t') < P^*$. The proof that $\overline{P}_{d+1} \leq P_*$ is symmetric. By the induction argument above and Lemma 1, we have a day $d^* < \infty$ such that $P^* \leq \underline{P}_d < P_*$ and $P^* < \overline{P}_d \leq P_*$ for all $d \geq d^*$. QED

Theorem 1 applies to experiments which have a Walrasian equilibrium price P^e and quantity Q^e. These experiments fall into three groups. Firstly, if there are multiple units at the Walrasian equilibrium price (and if $D^c(P^e - 1) = S^c(P^e + 1)$), then Theorem 1 predicts that prices will eventually remain within one cent of P^e (as $P_* = P^e + 1$, $P^* = P^e - 1$) and that quantity traded will be at least $Q^e - 1$ and no more than the maximum of $S^c(P_* + 1)$ and $D^c(P^* - 1)$. Secondly, if there is only one unit at the Walrasian equilibrium price and $D^c(P^*) = S^c(P_*)$, the situation in IPDA14 in Appendix A, then Theorem 1 predicts that eventually the maximum price will be no more than one cent above the minimum of the value of the first infra-marginal buyer (V^{Q^e-1}) and the cost of the first extra-marginal seller (M^{Q^e+1}). For IPDA14, this is 4.30. The prediction for the minimum price is symmetric. In the limit, prices tend to keep out extra-marginal units and to keep in infra-marginal units. For this class of experiments, the prediction is again that the quantity traded will eventually remain in the interval $[Q^e - 1, \max\{S^c(P_* + 1), D^c(P^* - 1)\}]$. Finally, if the experiment presents an interval of prices, any of which can be a Walrasian equilibrium, with no units at any of these prices and if $D^c(P^*) = S^c(P_*)$, the predictions of Theorem 1 are again that prices eventually remain in $[P^*, P_*]$. However, it is possible to design payoff schedules with one unit at the Walrasian equilibrium or with no units at any Walrasian equilibrium so that $D^c(P^*) \neq S^c(P_*)$. This case and cases where there is no Walrasian equilibrium are addressed by the following theorem.

THEOREM 2:

a. If $D^c(P^*) > S^c(P_*)$, then there exists a day $d^* < \infty$ such that $P^* \leq \underline{P}_d < P_*$ and $P^* < \overline{P}_d \leq M^{D^c(P^*)}$ for all $d \geq d^*$.

b. If $S^c(P_*) > D^c(P^*)$, then there exists a day $d^* < \infty$ such that $V^{S^c(P_*)} \leq \underline{P}_d < P_*$ and $P^* < \overline{P}_d \leq P_*$ for all $d \geq d^*$.

PROOF: We prove part (a); the proof for (b) is symmetric. We first need to establish the relationship between P^*, P_*, $M^{D^c(P^*)}$, and $V^{S^c(P_*)}$.

CLAIM 5: If $D^c(P^*) > S^c(P_*)$, then $M^{D^c(P^*)} > P_* > V^{S^c(P_*)} \geq P^*$.

PROOF:

i. Suppose $P_* \geq M^{D^c(P^*)}$. Then $S^c(P_*) \geq D^c(P^*)$. A contradiction.

ii. Suppose $V^{S^c(P_*)} \geq P_*$. Then $D^c(P_*) \geq S^c(P_*) \geq S^o(P_*)$. But by definition, $D^c(P_*) + 2 \leq S^o(P_*)$.

iii. Suppose $P^* > V^{S^c(P_*)}$. Then $D^c(P^*) < S^c(P_*)$. A contradiction.

By the argument in Theorem 1 we know that there is a day $d^* < \infty$ such that $[\underline{P}^*_d, \overline{P}^*_d] \subseteq [P^*, P_*]$. By Claim 5, $M^{D^c(P^*)} > P_*$. So $[\underline{P}^*_d, \overline{P}^*_d] \subset [P^*, M^{D^c(P^*)}]$. The proof now proceeds by induction. We need to show that if $[\underline{P}_d, \overline{P}_d] \subseteq [P^*, M^{D^c(P^*)}]$, then $[\underline{P}_{d+1}, \overline{P}_{d+1}] \subseteq [P^*, M^{D^c(P^*)}]$. There are two cases to consider: (1) $\overline{P}_d \leq P_*$ and (2) $\overline{P}_d > P_*$.

Case 1: $\overline{P}_d \leq P_*$. As $M^{D^c(P^*)} > P_* \geq \overline{P}_d$ and $V^{S^c(P_*)} \geq P^*$ by Claim 5, an argument similar to the proof of Theorem 1 shows that $\overline{P}_{d+1} \leq M^{D^c(P^*)}$ and $\underline{P}_{d+1} \geq P^*$.

Case 2: $\overline{P}_d > P_*$. We know that $\underline{P}_d < P_*$ for all $d \geq d^*$, so $\overline{P}_d > P_*$ implies that $\Delta P_d > 1$. So by Lemma 3, $u_{d+1} \leq u_d$. By definition $u_d = \text{Max}\{\overline{P}_d, M^{D^o(\underline{P}_d)}\}$ and by hypothesis $\underline{P}_d \geq P^*$. So $D^o(\underline{P}_d) \leq D^o(P^*)$. Thus, $M^{D^o(\underline{P}_d)} \leq M^{D^o(P^*)} \leq M^{D^c(P^*)}$. By hypothesis $\overline{P}_d \leq M^{D^c(P^*)}$. So $u_d \leq M^{D^c(P^*)}$. By definition $u_{d+1} = \text{Max}\{\overline{P}_{d+1}, M^{D^o(\underline{P}_{d+1})}\}$. Now $u_{d+1} \leq u_d \leq M^{D^c(P^*)}$. So $\overline{P}_{d+1} \leq M^{D^c(P^*)}$.

We also need to show that $\underline{P}_{d+1} \geq P^*$. Suppose not. Then $\underline{P}_{d+1} < P_*$. This requires $S^o(\underline{P}_d) > D^c(\underline{P}_d)$. Thus, $S^c(P_*) \geq S^o(P_*) \geq S^o(\underline{P}_d) > D^c(\underline{P}_d) \geq D^c(P^*)$. This contradicts $D^c(P^*) > S^c(P_*)$. So $\underline{P}_{d+1} \geq P^*$.

Theorem 2 now follows from the induction argument above and Lemma 1. QED

Although Lemmas 1, 2, and 3 imply that prices are eventually contained in the interval $[P^*, P_*]$, they need not stay in this interval if supply and demand are not equal there. For example, if $D^c(P^*) > S^c(P_*)$ and low-value buyers (those with $P^* - 1 \leq V^i < P_*$) trade first, the remaining high-value buyers may bid prices up. However, they need not, and so will not, bid more than $M^{D^c(P^*)}$ in order to complete a trade. So the range of prices could expand to be $[P^*, M^{D^c(P^*)}]$.

In subsequent days it will shrink until it is again contained in $[P^*, P_*]$. It seems unlikely that this process would continue, and our theory does not predict that it will, only that it might. In fact, Lemma 3 implies that for all supply and demand configurations if all extra-marginal units are excluded by $[\underline{P}_d, \overline{P}_d]$, then the interval will shrink to at most one cent and then remain fixed.

6. COMPARISONS OF THE PREDICTIONS WITH THE DATA

Our prediction of convergence seems consistent with the experimental data, but it is not directly testable with these data as the number of repetitions necessary for convergence is not specified. In any case, obtaining the competitive equilibrium in the limit is only a first test of a theory of price formation in double oral auctions. We have rejected the models considered in Section 3, at least in part, on the basis of their incorrect predictions about dynamics. In this section we compare the predictions of our model with experimental data. There are three categories of data for which our theory has implications: The sequence of minimum and maximum prices, the sequence of trading partners, and the number of units traded.

The three lemmas in Section 5 directly yield predictions about the dynamics of minimum and maximum prices. Lemma 1 implies that these prices move to bracket the competitive equilibrium price and that once this is accomplished, the theoretical equilibrium price remains in the interval $[\underline{P}, \overline{P}]$. Lemmas 2 and 3 imply that minimum and maximum prices respond to the forces of demand and supply. The prediction is that the minimum price will rise if demand at $\underline{P}$ exceeds supply at $\overline{P}$ and that the maximum price will fall if supply at $\overline{P}$ exceeds demand at $\underline{P}$. In the excess demand case ($D^o(\underline{P}) \geq S^o(\overline{P})$), the maximum price may rise, but the prediction is that it will go no higher than the level necessary to allow the $D^o(\underline{P})$th unit to trade. For the excess supply case, the prediction is that although the minimum price may fall it will not go below $V^{S^o(\overline{P})}$.

Remember that $\underline{P}_d$ is the lowest contract price or offer that occurred during the day *before d*. $\overline{P}_d$ is the highest contract price or bid observed during the day *before d*.

PREDICTION 1. Prices (See Lemmas 1, 2, and 3):

i. If $\underline{P}_d \geq P_*$, then $\underline{P}_{d+1} < \underline{P}_d$. If $\underline{P}_d < P_*$, then $\underline{P}_{d+1} < P_*$.

ii. If $\overline{P}_d \leq P^*$, then $\overline{P}_{d+1} > \overline{P}_d$. If $\overline{P}_d > P^*$, then $\overline{P}_{d+1} > P^*$.

iii. If $\Delta P_d > 1$ and $D^o(\underline{P}_d) \geq S^o(\overline{P}_d)$, then $\underline{P}_{d+1} > \underline{P}_d$ and $\overline{P}_{d+1} \leq u_d$.

iv. If $\Delta P_d > 1$ and $S^o(\overline{P}_d) \geq D^o(\underline{P}_d)$, then $\overline{P}_{d+1} < \overline{P}_d$ and $\underline{P}_{d+1} \geq \ell_d$.

TABLE 1 Violation Percentages[1]

	Descriptors							Analysis		
Experiment	Exp	Marg	Com	Units	Q^e	NYSE	Que	Price	Seq	Quant
IPDA8	No	1	5	8	6	Yes	No	13.9	0	11.1
IPDA9	No	1	5	8	6	Yes	No	9.4	1.6	0
IPDA10	Yes	1	5	10	8,6	Yes	No	6.3	0	0
IPDA11	No	1	5	10	8	Yes	No	18.8	1.3	0
IPDA14	Yes	1	5	8	6	Yes	No	9.4	0	0
IIPDA14	No	3	10	21	15	Yes	No	12.5	0	0
IIPDA22	No	0	10	16,11	11	No	No	3.1	NA	12.5
IIPDA25	Yes	2	10	12	7	No	Yes	21.9	0.8	12.5
IIPDA57	No	3	10	21	15	Yes	Yes	16.7	1.0	8.3
Average								12.7	0.7	5.3

1

PDA =Plato Double Auction

Exp =experienced subjects (have participated in another PDA)

Marg =number of extra marginal units

Com =commission in cents per unit traded

Units =number of units on each side of the market

Q^e =competitive equilibrium quantity

NYSE =New York Stock Exchange rules (new bids and offers must improve on outstanding bids and offers)

QUE =electronic queuing of bids and offers (see Smith and Williams[21])

Price =% violation of price predictions

Seq =% violation of trading sequence predictions

Quant =% violation of quantity traded predictions

Our model yields no predictions for day 1 of any experiment. In any experiment in which supply or demand was shifted, we treat the first day after the shift as day 1 of a new experiment. In each non-initial day of an experiment, we have four possible violations of price predictions: violations of 1(i) and 1(ii), and the two predictions of either 1(iii) or 1(iv). So, in an experiment running for ten days with no shifts, there are 36 possible price violations. The entry for price violations is the number of violations divided by the number of possible violations. In each non-initial day the number of possible trading sequence violations is the number of buyer plus seller units $(n + m)$. The number of actual violations is the number of units traded out of order. In each non-initial day there is one quantity prediction, and thus one possible quantity violation.

TABLE 2 Price Violations

Price violations of x\$ or less not counted	Percentage of price violations over nine DOAs
$x = 0.01$	6.3
$x = 0.05$	3.3
$x = 0.10$	1.3

Our prediction about the sequence of trading partners follows from the proofs of the Lemmas. It is essentially that sellers with costs below $\overline{P}$ trade before those with costs above $\overline{P}$ and buyers with values above $\underline{P}$ trade before those with values below $\underline{P}$.

PREDICTION 2. Trading Sequence:

i. If $\Delta P_d > 1$:
All sellers with cost below $\overline{P}_d$ trade before any sellers with cost at or above $\overline{P}_d$. All buyers with value above $\underline{P}_d$ trade before any buyers with value at or below $\underline{P}_d$.

ii. If $\Delta P_d \leq 1$:
All sellers with cost at or below $\overline{P}_d$ trade before any sellers with cost above $\overline{P}_d$. All buyers with value at or above $\underline{P}_d$ trade before any buyers with value below $\underline{P}_d$.

Our prediction about the number of units traded is that it will be at least the competitive equilibrium for demand and supply curves truncated at $\overline{P}$ and $\underline{P}$, respectively, less one unit.

PREDICTION 3. Quantity Traded:
The quantity traded in day d will be $Q_d \geq \text{Max}\{K : \hat{V}_d^K \geq \hat{M}_d^K\} - 1$ where $\hat{V}_d^K = \text{Min}\{\underline{P}_d, V^K\}$ for $K = 1, \ldots, n$ and $\hat{M}_d^K = \text{Max}\{\underline{P}_d, M^K\}$ for $K = 1, \ldots, m$.

Table 1 summarizes, for nine DOA experiments, violations of our predictions about prices, sequence of trades, and number of units traded as a percentage of total possible violations. This table is based on data from Williams,[22] and on unpublished data which were made available by Vernon Smith. We realize this table is neither easy to understand nor conclusive empirical verification of our theory. We offer it only as supporting evidence of plausibility. We have yet to see a DOA experiment which is much different in its violations of Predictions 1 to 3 than those in Table 1.

To see the total number of price violations in perspective, Table 2 illustrates the margins of error. This table reports the total number of violations of our price predictions (over all nine experiments) which were more than x cents, as a percentage of the total number of possible violations of our price predictions.

To put sequence and quantity violations in context, it is useful to compare them with the violations of the sequence and quantity predictions of the Marshallian theory and the sequence predictions of the game theory approach. The Marshallian theory predicts that units will trade in the order of value and that all profitable trades will occur. The violations of this prediction as a percentage of possible violations in IPDA8 is 42.5%. The game theory approach[23] predicts that units will trade in the order of value but yields no further prediction on the number of trades. The violations of this prediction as a percentage of possible violations in IPDA8 is 29.4%.

7. FURTHER EXPERIMENTS

There is now a role for further interaction between theory and experiments. The class of experiments described in Section 6 motivated our theory, and it in turn suggests several experiments which could lead to refinements or rejection of the theory. There are several aspects of our theory which could be tested. Firstly, we do not assume that traders' reservation prices and bids or offers converge to their true values at the end of each day. The data that we have seems to reject such an assumption. However, without this assumption we can establish convergence only to an interval determined by P_* and P^*. Our theory admits as an equilibrium a situation in which one extra-marginal unit is included or in which one infra-marginal unit is excluded. For example, Theorem 1 applies to the demand and supply configuration in Figure 1 of Appendix B to predict equilibrium in the interval [114,148]. The placement of the first extra-marginal units in that figure has no effect on our equilibrium prediction. Charles Plott and Chris Worrell have run a DOA experiment using the configuration of Figure 1. Their data suggest that prices converge into the interval [133,139] determined by the first extra-marginal units. This conclusion does not reject our theory, but it does suggest that the theory might be refined to produce sharper convergence results.

Secondly, we have refrained from placing any direct assumption on the relative (between agent) rankings of true values and reservation prices. Possible ranking hypotheses on buyers' reservation prices include (1) if $V^i > V^j$, then $r_d^i(t) > r_d^j(t)$ and (2) if $V^i > \overline{P}_d$ and $V_j \leq \underline{P}_d$, then $r_d^i(t) > r_d^j(t)$. Hypothesis 1 is clearly rejected by the data, but whether hypothesis 2 is rejected depends on one's standard of acceptance. The absence of a ranking hypothesis is responsible for the relatively weak prediction of Theorem 2. In DOAs where $D^c(P^*) \neq S^c(P_*)$, our theory predicts convergence into the interval $[P^*, P_*]$, but it then admits the possibility of cycles between prices in this interval and prices as low as $V^{S^c(P_*)}$ if $S^c(P_*) > D^c(P^*)$ or prices as high as $M^{D^c(P^*)}$ if $D^c(P^*) > S^c(P_*)$. In the presence of either ranking hypothesis (and a symmetric hypothesis on sellers' reservation prices) cycles would not occur and prices would remain in $[P^*, P_*]$. We have some data about experiments where our theory admits the possibility of cycles. In both IPDA8 and IPDA9 (reported in Section 6), $S^c(P_*) > D^c(P^*)$. In neither of these experiments do we see cycles; prices seem to remain approximately in $[P^*, P_*]$. However, the extra-marginal seller unit at $P_* - 1$ is occasionally traded, so it is possible that cycles would have arisen had the experiments continued beyond ten days.[14] This suggests two possible further experiments. Firstly, IPDA8 could be run for more days to decide whether cycles will appear. Secondly, an experiment with a design more likely to produce cycles could be run. The supply and demand configuration of Figure 2 in Appendix B is such a design. The prediction of Theorem 2 for this configuration is that prices will remain in the interval $[V^{S^c(P_*)}, P_*] = [70, 101]$. Our conjecture is that in the experiments any cycles would eventually disappear, with prices remaining in $[P^*, P_*]$ and perhaps following a time path during each day starting at P^*, and then rising during the day. By offering a small discount (to P^*) early in the trading day, infra-marginal sellers could insure that they complete a trade. Prices would then rise by one or two cents as marginal traders complete their trades. If this occurs, it suggests that the theory might be further refined.

There are several other experiments which could lead to refinements or rejection of our theory. Firstly, our theory is silent about the fine details of organizing a DOA. All that counts is that traders can make bids or offers and acceptances, and that they are informed of others' bids, offers, and acceptances. Thus the predictions of the theory are unchanged by the use of New York Stock Exchange rules, electronic queues, or other details. However, the data are not unchanged by these details (see Smith and Williams[21]). It may be that sharper predictions would result if these details were taken into account. Secondly, the theory does not apply to experiments in which one side of the market is not allowed to bid or offer. See Plott and Smith[16] for some of these experiments. Third, the theory does not yield predictions about the effect of shifts in supply and demand curves. There is now data from experiments in which supply and demand curves are shifted systematically. The theory would need to be refined to yield useful predictions about the effect of such changes in market conditions.

[14] In IPDA8 the extra-marginal unit at $P_* - 1$ is seller 2's second unit. This unit is traded in both days 9 and 10 of the experiment.

The methodology of using experiments to test the predictions of theory can also be applied to the alternative theories that we have described. For instance, Wilson's game theoretic model of DOAs does not directly apply to the existing DOA experiments, but a DOA experiment could be designed to test the theory. Trader's values and costs could be drawn independently across days from distributions which the traders know, risk attitudes could be controlled for, and the experiment could be repeated for a number of days to allow for learning about the game and about strategies. Wilson's predictions could then be compared to data from the final day of the experiment.

8. CONCLUSION

The theory presented here is deterministic and, although it does not completely describe precise paths of bids, offers, and contracts, it does place fairly tight bounds on these data. One observation not in accord with these bounds is grounds for rejection of the theory, and in fact there are a number of such observations. However, the low percentage of observations which violate the obvious implications of the theory seems acceptable for a theory which yields fairly precise predictions about prices and trades.

The potential importance of this theory is not just that it seems to describe what happens in DOA experiments, but also that it is the beginning of a positive theory of how market prices are formed and of how they adjust to changes in demand and supply conditions. The question of price formation has a long history of ad hoc and unsuccessful attempts at an answer. Our theory is also ad hoc in the sense that we make assumptions on individual behavior which are not derived from an optimizing model. However, our assumptions seem sensible, are consistent with the behavior implied by at least one optimizing, game-theoretic model and, more importantly, they seem to do a reasonable job of describing actual bids, offers, and contracts. There is now a target for experimentalists to reject with data or for theorists to improve on by obtaining a better fit to the data.

9. SOME ACKNOWLEDGMENTS AND A HISTORICAL NOTE

This paper benefited from discussions in seminars at Cornell, Northwestern, and Stony Brook, and an NSF Conference on Experimental Economics at the University of Arizona. It has also been affected by several referees, some of whom have been helpful in forcing us to more clearly present our ideas. This version is significantly different from earlier versions. We would like to thank Vernon Smith and Arlington Williams for making data on their Plata DOA experiment data available to us.

We would like to thank the editors for inviting us to participate in this volume. We thank John Rust for encouraging us to expose our theory in the testbed of the computerized Double Auction Market at the Santa Fe Institute. We feel reasonably happy with our fifth place finish and like to think that even though our theory was designed for continuous-time open-outcry systems rather than the sychronized double auction, it (the theory) held up pretty well. We thank Dan Friedman for not forgetting our paper and for forcing us, once again, to drag it out of retirement. Finally we thank Bob Wilson for his insights and Vernon Smith and Arlington Williams for providing data and support in the early days of 1980-81.

Because the first version of this paper appeared in 1981 and the last significant revision was made in 1988, some of the references and most of the data may seem somewhat outdated. The task of constructing a new type of theory to explain a new type of data derived from nascent experimental markets was actually begun by us in 1978. Since 1981, game theory arose as a new paradigm in economics and many editors were loath to publish a non-optimizing base theory such as that we had proposed. Our last 1988 revision was written to respond to these concerns. Since then, game theorists analyzing dynamics have turned from models with common knowledge and Bayes equilibrium to models of Bayesian learning, fictitious play, and other non-optimizing models of behavior which do not require, for example, common knowledge of rationality. We are glad others are pursuing the path we took in 1980. We believe that many of these "learning" models are consistent with our assumptions, but that is another paper.

Not only has theory changed since 1981 but experimental technology has significantly advanced. As one can see, from articles in this volume, the simple DOA experimental market we have described in this paper has become more sophisticated and interesting. In our sporadic efforts at casual empiricism, we have yet to see data in the more recent experimental markets that significantly differs from the predictions of our theory. For example, DOA's where commissions are not paid look like the DOA's we analyze. Of course what is needed is a serious empirical study to see whether our impressions are valid. As far as we know, our challenges to theorists and experimentalists in Sections 7 and 8 remain unaccepted. We hope someone responds.

APPENDIX A

TABLE 3 Values and Costs for DOA #IPDA14

	Week 1		Week 2	
	Unit 1	Unit 2	Unit 1	Unit 2
BYR1	5.20	3.80	3.70	3.60
BYR2	5.00	4.00	3.80	3.50
BYR3	4.80	4.20	3.90	3.40
BYR4	4.60	4.40	4.00	3.30
SLR1	3.70	4.40	3.10	3.30
SLR2	3.80	4.30	2.90	3.50
SLR3	3.90	4.20	2.70	3.70
SLR4	4.00	4.10	2.50	3.90

TABLE 4 Period 9

	MKR	TM	BIDS	OFFERS	TKR	TM
1	B2	297	3.30			
2	B1	295	3.34			
3	B2	292	3.35			
4	S4	289		3.45	B3	285
5	B2	281	3.30			
6	B4	278	3.35			
7	B2	275	3.36			
8	S3	274		3.50		
9	B1	268	3.39			
10	B4	258	3.40		S2	252
11	S1	254		3.45	*	
12	B2	247	3.30			
13	B1	245	3.40			
14	S3	244		3.50		
15	S1	236		3.45		
16	B2	231	3.41			
17	S3	191		3.44		
18	S1	181		3.43	B2	154
19	B4	151	3.30			
20	B2	147	3.31			
21	S2	146		4.50		
22	B1	146	3.40			
23	B2	140	3.41			
24	S1	136		3.45	B1	68
25	B2	122	3.42		*	
26	B1	112	3.43		*	
27	B2	106	3.44		*	
28	S2	63		4.50		
29	B4	62	3.30			
30	B1	61	3.40			
31	S3	57		3.49		
32	B2	55	3.41			
33	B1	45	3.43			
34	B2	41	3.44			
35	B2	23	3.45		S3	17
36	B4	11	3.30			
37	S2	10		3.50	B1	0
38	B1	8	3.40		*	

TABLE 5 Summary Data from IPDA14

	Contract Price	Buyer	Seller
Day 1	4.25	2	3
	4.20	4	1
	4.50	3	2
	4.40	4	4
	4.30	1	2
Day 2	4.35	3	1
	4.30	4	2
	4.30	4	2
	4.30	1	4
	4.25	2	4
Day 3	4.25	4	1
	4.35	3	3
	4.30	4	2
	4.27	2	4
	4.25	1	3
Day 4	4.30	4	1
	4.39	3	4
	4.30	4	4
	4.26	2	2
	4.25	1	3
	4.20	3	2
Day 5	4.30	4	1
	4.26	2	2
	4.35	3	4
	4.26	1	3
	4.25	4	4
Day 6	3.35	4	1
	3.30	3	2
	3.35	3	3
	3.32	2	1
	3.35	1	4
Day 7	3.31	2	2
	3.35	3	1
	3.40	3	4
	3.38	2	1
	3.45	1	3

TABLE 5 Summary Data from IPDA14 (cont'd.)

	Contract Price	Buyer	Seller
Day 8	3.40	4	2
	3.40	1	1
	3.40	1	1
	3.40	3	4
	3.40	2	3
Day 9	3.45	3	4
	3.40	4	2
	3.43	2	1
	3.45	1	1
	3.45	2	3
	3.50	1	2
Day 10	3.41	2	2
	3.44	1	4
	3.45	4	1
	3.45	1	3
	3.50	3	1

APPENDIX B

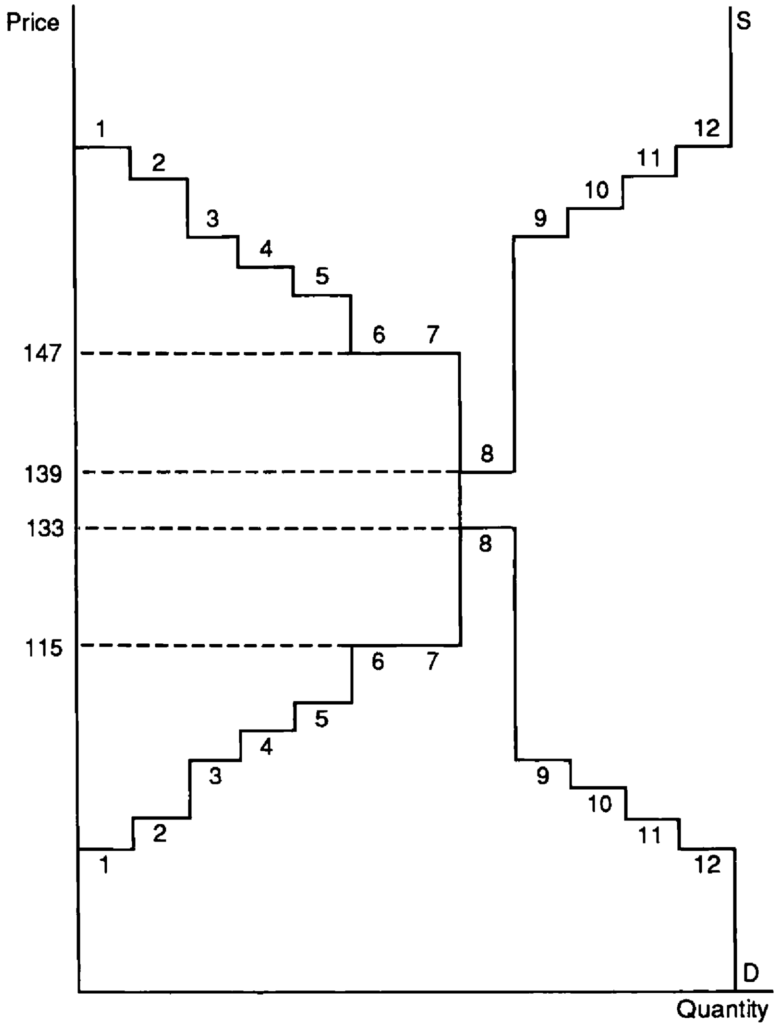


FIGURE 1 An induced demand-supply schedule.

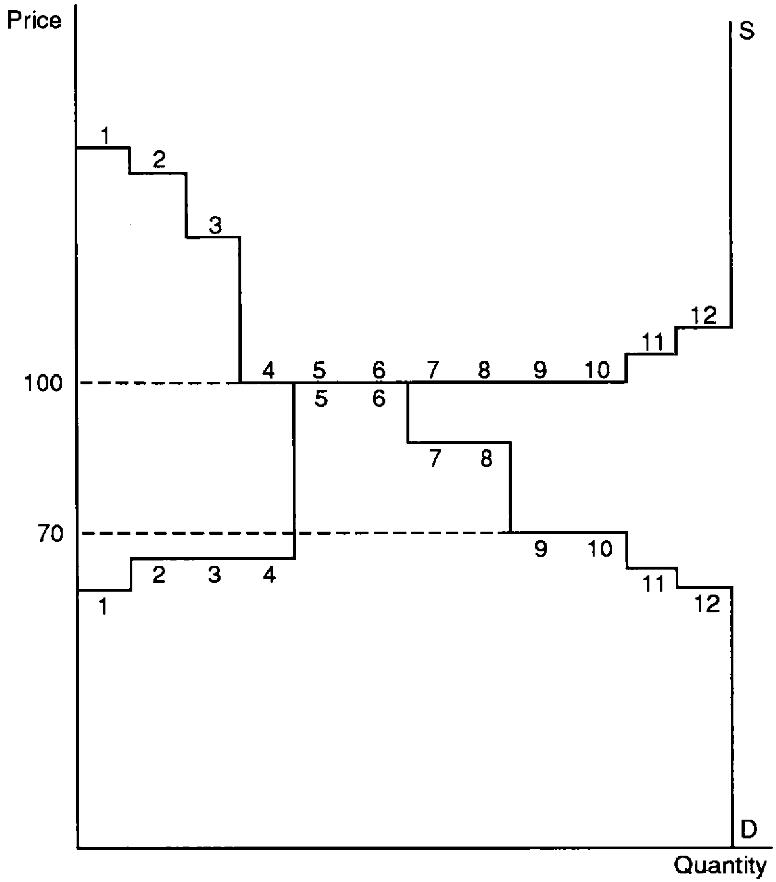


FIGURE 2 Supply and demand schedules for further experiments. Data is from an unpublished experiment by Charles Plott and Chris Worrell; reprinted by permission of the authors.

REFERENCES

1. Arrow, K. J., and F. H. Hahn. *General Competitive Analysis*. San Francisco: Holden-Day, 1971.
2. Cox, J. C., and R. L. Oaxaca. "Laboratory Experiments with a Finite Horizon: Job Search Model." *J. Risk & Uncertainty* **2** (1989): 301–329.
3. Friedman, D. "On the Efficiency of Double Auction Markets." *Am. Econ. Rev.* **74** (1984): 60–72.
4. Friedman, D. "A Simple Testable Model of Double Auction Markets." *J. Econ. Behav. & Organization* **15** (1991): 47–70.
5. Gresik, T. A., and M. Satterthwaite. "The Rate at Which a Simple Market Converges to Efficiency as the Number of Traders Increases: An Asymptotic Result for Optimal Trading Mechanisms." *J. Econ. Theory* **48** (1989): 304–332.
6. Gronau, R. "Information and Frictional Unemployment." *Am. Econ. Rev.* **61** (1971): 290–301.
7. Harsanyi, J. "Games with Incomplete Information Played by Bayesian Players." Parts I, II, and III. *Mgmt. Sci.* **14** (1967-68): 159–182, 320–334, 486–502.
8. Holt, C. A, L. W. Langan, and A. P. Villamil. "Market Power in Oral Double Auctions: Convergence to Competitive Equilibrium Prices Reconsidered." *Econ. Inquiry* **24** (1986): 107–123.
9. Jarecki, H. G. "Bullion Dealing, Commodity Exchange Trading and the London Gold Fixing: Three Forms of Commodity Auctions." In *Bidding and Auctioning for Procurement and Allocation*, edited by Y. Amihud, 146–154. New York: New York University Press, 1976.
10. Ledyard, J. "The Scope of the Hypothesis of Bayesian Equilibrium." *J. Econ. Theory* **39** (1986): 59–82.
11. Lippman, S. A., and J. J. McCall. "The Economics of Job Search: A Survey." *Econ. Inquiry* **14** (1976): 155–189.
12. Lippman, S. A., and J. J. McCall. "The Economics of Job Search: A Survey." *Econ. Inquiry* **14** (1976): 347–368.
13. Mortensen, D. T. "Job Search, the Duration of Unemployment, and the Phillips Curve." *Am. Econ. Rev.* **60** (1970): 847–862.
14. Myerson, R., and M. Satterthwaite. "Efficient Mechanisms for Bilateral Trade." *J. Econ. Theory* **29**: 265–281.
15. Noussair, C. "Auctions with Multi-Unit Demands." Mimeo, California Institute of Technology, 1991.
16. Plott, C. R., and V. L. Smith. "An Experimental Examination of Two Exchange Institutions." *Rev. Econ. Studies* **45** (1978): 133–153.

17. Roth, A., and M. Malouf. "Game Theoretic Models and the Role of Information in Bargaining." *Psych. Rev.* **86** (1979): 574–594.
18. Smith, V. L. "Experimental Economics: Induced Value Theory." *Am. Econ. Rev.* **66** (1976): 274–279.
19. Smith, V. L. "Reflections on Some Experimental Market Mechanisms for Classical Environments,." In *Research in Marketing, Choice Models for Buyer Behavior*, edited by Leigh McAlister, 13–47. Greenwich: JAI Press, 1982.
20. Smith, V. L. "Microeconomic Systems as an Experimental Science." *Am. Econ. Rev.* **72** (1982): 923–955.
21. Smith, V. L., and A. W. Williams. "An Experimental Comparison of Alternative Rules for Competitive Market Exchange." In *Auctions, Bidding and Contracting: Uses and Theory*, edited by M. Shubik. New York: New York University Press, 1983.
22. Williams, A. "Computerized Double-Auction Markets: Some Initial Experimental Results." *J. Business* **53** (1980): 235–258.
23. Wilson, R. "Equilibrium in Bid-Ask Markets." In *Arrow and the Ascent of Economic Theory: Essays in Honor of Kenneth J. Arrow*, edited by G. Feiwel. London: McMillan, 1987.

Mark A. Satterthwaite† and Steven R. Williams‡
†Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208, and ‡ Department of Economics, University of Illinois, Urbana, IL 61820

FOUR
The Bayesian Theory of the *k*-Double Auction

INTRODUCTION

Supply, demand, and trade at a market-clearing price are the most fundamental concepts of microeconomics. Marshall developed this analysis to describe properties of the outcome of trading that he believed to be essentially correct despite the fact that traders do not have "thorough knowledge of the circumstances of the market" (Marshall,[11] ch. II, sec. 2) and, hence, try to manipulate the terms of trade in their favor. The supply-demand analysis is regarded in a different way in this paper. The "Marshallian cross" is interpreted here as an institution for determining the market price-quantity pair, and the behavior of self-interested, imperfectly informed traders when confronted with this institution is analyzed. We study a call market—i.e., a market in which bids determine a demand curve, asks determine a supply curve, and all trades clear simultaneously at a market-clearing price.[1] This paper describes the

[1]Such an institution is commonly used to arrange trade in thin markets and for price discovery in more liquid markets at the opening of the trading day. See Schwartz[20] for a discussion of call markets. In his survey of institutions, Friedman[3] classifies call markets as one-shot clearinghouse markets.

institution, summarizes a theory that supports Marshall's intuition, and discusses how those theoretical predictions may be confronted with experimental evidence.[2]

For simplicity, the theory is restricted to a trading environment consisting of m sellers and m buyers.[3] Each seller has for sale a single, indivisible unit of a homogeneous good, and each buyer is interested in purchasing one unit of the good. We assume that every trader has a reservation value for the good—cost c_i for a seller and value v_i for a buyer—that represents the value in money he places on a unit. Each trader privately knows his own reservation value. The assumption of private information is critical, for as Hayek[6] emphasized, assuming instead that some "single mind" possesses all information relevant to trading trivializes the problem markets exist to solve.

The institution works as follows. At a set time each trader submits a sealed bid V_i if he is a buyer or offer C_i if he is a seller.[4] The offers (or asks) and bids are arrayed to form "reported" supply and demand curves, a market-clearing price p is selected, and units are exchanged among those sellers who offered less than p and those buyers who bid at least p. The market then disbands with no opportunity for recontracting.[5] We call this institution a *double auction* (or DA), because both sides of the markets jointly determine the price through their offers/bids. In particular, for $k \in [0, 1]$, the *k-double auction* (or *k-DA*) is the particular institution that selects $kb+(1-k)a$ as the price when $[a, b]$ is the interval in which a market-clearing price can be selected.

Figure 1 illustrates the institution for the case of $m = 3$, $k = 1$, and a specific realization of costs and values. What Marshall considered as the "true" supply and demand curves of the underlying economic environment are depicted by the step functions SS and DD: the three sellers had costs c_i equal to 0.30, 0.57, and 0.90, while the three buyers had values v_i equal to 0.73, 0.64, and 0.14. These curves depict willingness to trade as given by the traders' privately known reservation values. Sellers submitted offers C_i equal to their costs c_i. Buyers, however, attempted to

[2] While this paper mostly concerns our work and our joint work with Aldo Rustichini, we also wish to cite the early work of Roberts and Postlewaite[16] and the more recent work of Chatterjee and Samuelson,[2] Wilson,[22] and Gresik and Satterthwaite,[4] upon which we built.

[3] The case in which there are different numbers of traders on each side of the market is considered in Williams[21] and Rustichini, Satterthwaite, and Williams.[17] Henceforth, the latter paper will be referred to as RSW.

[4] Because of the assumption of unitary supply/demand, there is no point in allowing traders in this restricted environment to submit multiple offers/bids or offers/bids together with a quantity. The institution—if not the theoretical results—described here is easily generalized to this richer environment.

[5] This is different from most real-world call markets in which some form of aftermarket typically exists. It is also different from the versions of the uniform-price double auctions that McCabe, Rassenti, and Smith[12] study. There, bids/offers may be submitted sequentially over time, and each trader has the opportunity to revise his bid/offer in response to others before the market clears. The possibility of trade in an aftermarket or revision of a submitted bid/offer would influence behavior in the call market.

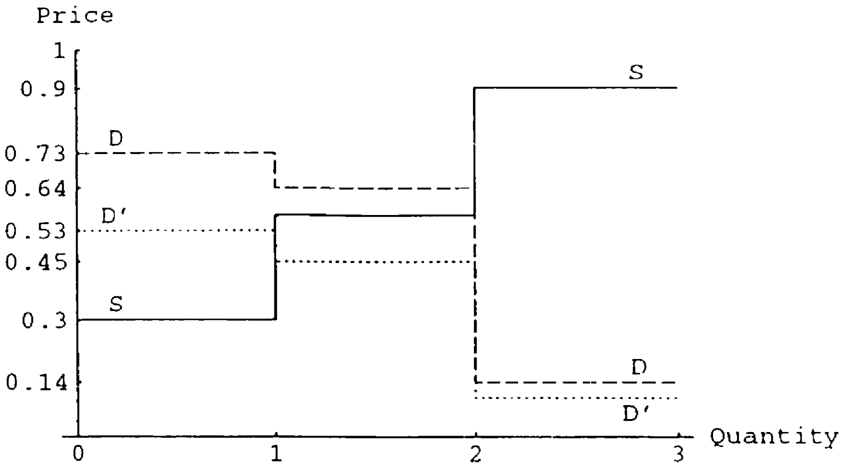


FIGURE 1 An example of a DA with $m = 3$ and $k = 1$. SS and DD are the true supply and demand curves and $D'D'$ is the demand curve strategically reported by buyers. Sellers report their true supply curve because $k = 1$. The price relative to reported supply and demand is 0.53; one unit is traded at that price.

manipulate the price by submitting bids less than their values; these bids of 0.53, 0.45, and 0.10 are represented on the figure by the dotted step function $D'D'$. The interval of market-clearing prices is $[0.45, 0.53]$. In the 1-DA the price is set at 0.53, and the buyer who bid 0.53 trades with the seller who offered 0.30.

Three points should be noted about the example of Figure 1. First, the buyer who bid 0.53 regrets *ex post* that he did not bid lower, for if he had bid 0.46, he would have still received a unit but at a price of 0.46. This possibility of being able to influence price is what leads each buyer to bid less than his reservation value and causes the reported demand $D'D'$ to lie below true demand DD.[6] Second, the DA assigns the available $m = 3$ units to the three traders who reported the highest values. In this example, the three traders are the one buyer who buys and the two sellers who retain their units. Third, the tendency of buyers to choose their bids strategically may lead to inefficient outcomes. Efficiency requires that the $m = 3$ available units be assigned to those traders who truly value them most highly. Two units should be traded: the buyers with values 0.73 and 0.64 should trade with the sellers with costs 0.30 and 0.57. Buyer two fails to trade because he attempted to manipulate price in his favor by bidding 0.45. As illustrated by the crossing of the true supply and demand curves SS and DD at quantity two, the outcome of

[6]Sellers do not ask for more than their true costs in this example because $k = 1$ and, consequently, they cannot influence the price at which they trade. This is the special case of the 1-DA that is discussed in the next section. If $k < 1$, then sellers, too, have an incentive to misrepresent their costs, which shifts the reported supply curve above the true supply curve.

trading would have been efficient if buyers had not acted strategically by reporting the bids that generated the demand curve $D'D'$.

Figure 1 reveals the difficulty of the problem that a trading institution must solve, even in such a simple setting. Because reservation values are private, the institution must elicit information about traders' values from the traders themselves and then use those elicited values to allocate the units available for trade. But because the institution must use traders' reported values, traders may have an incentive to make offers/bids that are self-serving rather than honest. Thus the institution must strike a balance between providing incentives to reveal private information accurately and fully using the information that is revealed to make the allocation.[7]

Theoretical research suggests that the k-DA makes this tradeoff remarkably well. This research, which we review in the next three sections, has focused on "equilibrium" misrepresentation. It shows that when the number of traders is small (m equal to one or two), both somewhat efficient and very inefficient equilibria exist. No convincing theory has been developed that suggests which of these equilibria are most likely to occur in practice. As the number of traders grows, however, the set of equilibria shrinks, reported supply and demand quickly converge to true supply and demand, and the outcome of trading converges to the competitive equilibrium price and quantity. Our research thus supports Marshall's argument that true supply and demand essentially determine the outcome of trade even though traders have incomplete information and act strategically.

The theoretical analysis of the k-DA can be radically complicated if even minor modifications of its rules are made so that it no longer uses a market-clearing price to mediate trade. The fifth section describes such a modification, which Kagel and Vogt[9] used during the pilot phase of an experimental investigation of the k-DA, and discusses the analytical problems it causes. Finally, in the sixth section we review the testable propositions that the theory implies and discuss the possibility of using experimental methods to test these implications.

[7]The k-DA takes the information traders reveal at face value and fully uses it in making its allocation of the available supply. In the mechanism design approach, the institution is carefully created for the specific trading environment so that traders have an incentive to honestly reveal their private information. This is achieved, however, by not fully using the information that is revealed; e.g., two traders may reveal that they should trade, but the mechanism will not have them trade because to do so would destroy the incentives for truthful revelation. This characteristic of the mechanism design approach is illustrated by the optimal trading mechanisms that are constructed in Myerson and Satterthwaite,[13] Gresik and Satterthwaite,[4] and Wilson.[23]

ENVIRONMENT, TRADER'S DECISION PROBLEM, AND THE BAYESIAN MODEL OF EQUILIBRIUM

ENVIRONMENT

There are m sellers and m buyers where $m \geq 2$, all of whom are risk neutral.[8] If seller i succeeds in selling his unit, then his utility is $p - c_i$ where p is the price he receives and c_i is the cost that he privately places on the unit. Otherwise, his utility is zero. Analogously, each buyer's utility is zero when he fails to trade and $v_i - p$ when he does, where v_i is the value he privately places on obtaining a unit.

INSTITUTION

Each trader submits a sealed offer (C_i) or bid (V_i) that is conditioned on his private knowledge of his cost c_i or value v_i. The offers and bids are arrayed in ascending order $s_{(1)} \leq s_{(2)} \leq \ldots \leq s_{(2m)}$; $s_{(j)}$ is thus the jth order statistic of the $2m$ bids and offers. Price is set within the interval $[s_{(m)}, s_{(m+1)}]$ of possible market-clearing prices at $p = (1-k)s_{(m)} + ks_{(m+1)}$ where k is a parameter selected from $[0, 1]$ that is fixed prior to the market being opened. Buyers who bid at least $s_{(m+1)}$ and sellers who offer at most $s_{(m)}$ trade. The only exception is if $s_{(m)} = s_{(m+1)}$, and the interval of market-clearing prices is degenerate. In this event, the quantity supplied may fail to equal the quantity demanded and, if necessary, a fair lottery is run to determine who will trade among those traders who bid or offered p and who are on the long side of the market.[9]

The cases of $k = 1$ (as in Figure 1) and $k = 0$ are special theoretically. Because price equals $s_{(m+1)}$ in the 1-DA, a seller cannot influence in his favor the price at which he actually trades. It is therefore in his best interest to submit his true reservation value as his offer, and reported supply is the same as true supply.[19] Similarly, a buyer in the 0-DA has the incentive to submit his true value as his bid. For $k \in (0, 1)$, a trader on either side of the market can influence price in his favor. He therefore has an incentive to shade his offer/bid away from his true reservation value.

[8] RSW treats the case in which all traders on the same side of the market have the same, possibly risk-averse, utility function. Risk aversion of this form does not substantially complicate the analysis. Chatterjee and Samuelson,[2] Satterthwaite and Williams,[18] and Leininger, Linhart, and Radner[10] consider the k-DA when $m = 1$.

[9] It is common to explain the k-DA by writing asks in increasing order $C_{(1)} \leq C_{(2)} \leq \ldots \leq C_{(m)}$ and bids in decreasing order $V_{(1)} \geq V_{(2)} \geq \cdots \geq V_{(m)}$. Asks and bids are then paired $C_{(1)} \leq V_{(1)}$, $C_{(2)} \leq V_{(2)}, \ldots$, until $C_{(j+1)} > V_{(j+1)}$, so that j is the total quantity traded. Our alternative explanation highlights that the m traders who report the highest values receive the units, and identifies strategic misreporting as the source of inefficiency.

THE TRADER'S DECISION PROBLEM

Buyers and sellers' decision problems are symmetric. Therefore, consider a specific buyer, buyer one; let $v \equiv v_1$ be the value he places on the good, and let λ be the bid he is considering. Buyer one's decision is risky, for the λ he submits affects both the probability that he will succeed in trading and, if he does trade, the expected price he will pay. Suppose he tests a value of λ by calculating the change in his expected utility if he raises his bid by a small amount $\Delta\lambda$. This calculation requires that he have a notion in the form of two probabilities concerning how the other m sellers and $m-1$ buyers are likely to bid.

Let $\zeta_{(1)} \leq \zeta_{(2)} \leq \dots \leq \zeta_{(2m-1)}$ be the random array of bids and offers made by the traders other than buyer one, noting in particular that $\zeta_{(j)}$ is the jth-order statistic from the restricted sample that excludes buyer one's bid. This contrasts with $s_{(j)}$, which is the jth-order statistic of the full sample of size $2m$ that includes his bid. To calculate the change in his expected utility, buyer one needs to know, first, the probability that increasing his bid by $\Delta\lambda$ will cause his bid to jump over $\zeta_{(m)}$, for if he does jump over $\zeta_{(m)}$, he goes from not trading to trading and picks up a trade of value approximately $v-\lambda$. Second, he needs to know the probability that his bid $\Delta\lambda$ is bracketed by $\zeta_{(m)}$ and $\zeta_{(m+1)}$, for if so, then increasing his bid by $\Delta\lambda$ will cause the price at which he trades to increase by $k\Delta\lambda$ to $(1-k)\zeta_{(m)}+k(\lambda+\Delta\lambda)$. His change in expected utility is therefore

$$(\nu - \lambda)\ Pr\{\zeta_{(m)} \in (\lambda, \lambda + \Delta\lambda)\} - k\Delta\lambda\ Pr\{\lambda \in (\zeta_{(m)}, \zeta_{(m+1)})\}\,, \qquad (2.1)$$

where the first term is buyer one's expected gain from switching from being an unsuccessful bidder to being a successful bidder, and the second term is his expected loss from causing the price to rise by $k\Delta\lambda$.

Buyer one may, in practice, estimate these probabilities from the empirical distribution of other traders' bids and offers, and thence compute his optimal bid using Eq. (2.1). We return to this possibility below in Section 6 when we discuss experiments concerning the k-DA. A second possibility, which is more fruitful theoretically, is that buyer one deduces these probabilities from his knowledge of the strategies of the other traders and his beliefs concerning the distributions of sellers' costs and other buyers' values.

THE BAYESIAN GAME MODEL

Some definitions are needed for this second approach. Since traders' reservation values are private, buyer one is uncertain about other traders' costs and values. Let distribution F with density f represent buyer one's subjective beliefs concerning any seller's cost c_i and let distribution G with density g represent his subjective beliefs concerning any other buyer's value v_i. Assume buyer one regards the cost or value of any other trader as distributed independently of his own value v and the values and costs of every other trader. A *strategy* for a trader is a function that specifies an offer/bid for each of his possible reservation values. Let $S(\cdot)$ and $B(\cdot)$

be the strategies that buyer one believes are being used by all sellers and all other buyers, respectively. Thus, if seller i has cost c_i, he offers $C_i = S(c_i)$.

Given these definitions and assumptions, the following three probabilities are well defined:

- $K(\lambda)$ is the probability that if $m-1$ buyers bid using strategy B and $m-1$ sellers offer using strategy S, then exactly $m-1$ bids/offers are less than λ;
- $L(\lambda)$ is the probability that if $m-2$ buyers bid using strategy B and m sellers offer using strategy S, then exactly $m-1$ bids/offers are less than λ;
- $M(\lambda)$ is the probability that if $m-1$ buyers bid using strategy B and m sellers offer using strategy S, then exactly m bids/offers are less than λ. Note that $M(\lambda) \equiv Pr\{\lambda \in (\zeta_{(m)}, \zeta_{(m+1)})\}$, as used in Eq. (2.1).

Formulas for these probabilities can be found in RSW.[17]

Given S, B, F, and G, expression (2.1) implies that the formula for the marginal expected utility of buyer one with value v and bid λ is:

$$\frac{\partial U_B(v,\lambda)}{\partial \lambda} = (v-\lambda)h_B(\lambda) - kM(\lambda) \tag{2.2}$$

where $h_B(\lambda)$ is the density of $\zeta_{(m)}$. If $c = c(\lambda) \equiv S^{-1}(\lambda)$, $\dot{c} = c'(\lambda) = 1/S'[c(\lambda)]$, $v = v(\lambda) \equiv B^{-1}(\lambda)$, and $\dot{v} = v'(\lambda) = 1/B'[v(\lambda)]$ are all well defined, then

$$h_B(\lambda) = mK(\lambda)f(c)\dot{c} + (m-1)L(\lambda)g(v)\dot{v}. \tag{2.3}$$

The first term on the right-hand side, when multiplied by $\Delta\lambda$, is the probability that buyer one's bid, if he increases it by $\Delta\lambda$, will jump over the offer of one of the m sellers, whose bid happens to be $\zeta_{(m)}$. Similarily, the second term is the probability of passing one of the other $m-1$ buyers whose bid happens to be $\zeta_{(m)}$.[10]

Buyer one's *best response strategy* to (S, B) specifies, for each of his possible reservation values v, a bid λ that maximizes his expected utility. With suitable regularity assumptions, if sellers are using strategy S and other buyers are using strategy B, then buyer one can select his optimal bid by setting Eq. (2.2) equal to zero and solving for λ.

[10] Pick a particular seller i. The probability that seller i's offer C_i is in $(\lambda, \lambda+\Delta\lambda)$ is $f(c_i)\dot{c} \times \Delta\lambda$ because c_i is a random variable with density f, $C_i = S(c_i)$, and C_i has density $f(c_i)\dot{c} \equiv f[c(C_i)]c'(c_i)$. Given $C_i \in (\lambda, \lambda+\Delta\lambda)$, $K(\lambda)$ is the probability it is $\zeta_{(m)}$. Therefore, $K(\lambda)f(c)\dot{c}\Delta\lambda$ is the probability that (i) buyer one's bid jumps over C_i and (ii) C_i is $\zeta_{(m)}$.

BAYESIAN NASH EQUILIBRIUM

Following Harsanyi,[5] the equilibrium concept that has been used to study the k-DA theoretically is symmetric Bayesian Nash equilibrium. Suppose, exactly as described above for buyer one, the subjective beliefs of every trader concerning the reservation values of other sellers and buyers are described by the distributions F and G, and that this is common knowledge among all the traders. Suppose further that traders' reservation values are elements of $[0, 1]$, that F and G are C^1 functions on $[0, 1]$, and that the densities f and g are strictly positive on $[0, 1]$. Consider a pair of strategies (S, B). Together they are a symmetric *Bayesian Nash equilibrium* for the k-DA if (i) for each seller i strategy S is the best response strategy to the other $m-1$ sellers playing S and the m sellers playing B and (ii) for each buyer i strategy B is the best response strategy to m sellers playing S and the other $m-1$ buyers playing B. Asymmetric equilibria in which each trader plays a distinct strategy, B_i or S_i, may exist, but as of yet have proven intractable to analysis. "Equilibrium" in this paper thus means "symmetric equilibrium."

The reason for using the Bayesian Nash equilibrium concept in studying the k-DA institution is that it models rational equilibrium behavior in a setting with incomplete information. Information is incomplete in that each trader's reservation value is private and other traders only have beliefs about his value. Behavior is rational in that expected utility is maximized conditional on one's information.

There are, however, at least two drawbacks of the Bayesian Nash equilibrium concept. First, no well-developed theory exists that explains how traders starting *de novo* jointly learn a set of equilibrium strategies (especially if more than one equilibrium exists). In other words, the theory does not include a process that results in the Bayesian Nash equilibrium. This poses a problem for the experimentalist, who receives little guidance from the theory concerning how equilibrium behavior is to be elicited from subject traders. Second, though the Bayesian Nash solution concept was created to model rational behavior in a setting with incomplete information, it makes the strong informational assumption that F and G are common knowledge among the traders in order to support the rationality of their behavior.

ELEMENTARY GEOMETRY OF EQUILIBRIUM STRATEGIES

Let $\underline{C} = \lim_{c\downarrow 0} S(c)$ and $\overline{V} = \lim_{v\uparrow 1} B(v)$. RSW show that for a given equilibrium (S, B), there exists numbers $\overline{c}$ and $\underline{v}$ such that: (i) a seller with value above $\overline{c}$ or a buyer with value below $\underline{v}$ trades with probability zero, while a seller with value below $\overline{c}$ or a buyer with value above $\underline{v}$ trades with positive probability; (ii) S is increasing over $[0, \overline{c}]$ and B is increasing over $[\underline{v}, 1]$; and (iii) $\lim_{v\downarrow \underline{v}} B(v) = \underline{v} = \underline{C}$ and $\lim_{c\uparrow \overline{c}} S(c) = \overline{c} = \overline{V}$. This geometric relationship is depicted in Figure 2 for a pair of continuous equilibrium strategies. In words, (i) implies that a buyer with too low a value or a seller with too high a cost will never trade. A trader with such a value feels no pressure to bid reasonably, e.g., a seller with cost c above $\overline{c}$ may

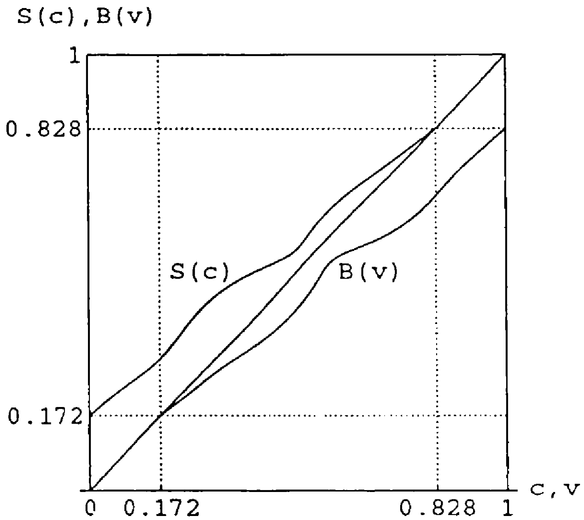


FIGURE 2 Equilibrium strategy pair (S, B) for $m = 2$, $k = 0.5$, and (F, G) uniform on $[0, 1]$. Note that $\underline{C} = \underline{v} = 0.172$ and $\overline{c} = \overline{V} = 0.828$.

choose an arbitrarily large number as his offer. The intervals $[0, \overline{c}]$ and $[\underline{v}, 1]$ are thus called the intervals over which serious offers/bids are made.[11] The diagonal is included in the figure so that the misrepresentation $S(c) - c$, $v - B(v)$ is easily seen.

CONVERGENCE OF EQUILIBRIA TO EX POST CLASSICAL EFFICIENCY AS THE NUMBER OF TRADERS INCREASES

The introduction discussed the incentives of traders to try to manipulate the price in their favor by overbidding in the case of sellers and underbidding in the case of buyers. This strategic misrepresentation causes inefficiency because trades that should occur do not necessarily occur. The economist's intuition is that as the market's size (measured here by m) becomes large, traders become essentially pricetakers (truthtellers) and the problems stemming from strategic misrepresentation vanish. Theorem 1 shows that this intuition is correct within the k-DA model. In particular, for every Bayesian Nash equilibrium, misrepresentation as measured by $S(c) - c$ and $v - B(v)$ is $O(1/m)$.

[11]On a more technical note, B must be differentiable a.e. in $[\underline{v}, 1]$ and S must be differentiable a.e. in $[0, \overline{c}]$ because they are increasing over these intervals. This substantiates the first-order approach outlined above for the case of arbitrary equilibria (S, B). The convergence result in the next section rests upon these first-order conditions.

THEOREM 1 (RSW, Th. 5.1). Consider any equilibrium (S, B) in which trade occurs with positive probability, every seller always offers at least as much as his cost, and every buyer bids at most his value. A number κ exists, whose value is a function of F and G but not of m, or (S, B), such that, for all $v \in (\underline{v}, 1]$ and $c \in [0, \bar{c})$,

$$S(c) - c \leq \frac{\kappa}{m} \tag{3.1}$$

and

$$v - B(v) \leq \frac{\kappa}{m}. \tag{3.2}$$

Additionally, $\underline{v} \leq \kappa/m$ and $\bar{c} \geq 1 - \kappa/m$.

Inequalities (3.1)–(3.2) bound the misrepresentation for a serious offer or bid; the last sentence bounds the intervals over which misrepresentation cannot be bounded. Together these bounds describe the rate at which misrepresentation vanishes as the number of traders on each side of the market increases. It is worth reiterating that this theorem applies to all Bayesian Nash equilibria of the k-DA, including ones in which S or B jump upward discontinuously at particular values of c or v.

Some intuition for this rate of convergence is obtained by outlining a proof of Eq. (3.2) for differentiable strategies S and B. Recall the buyer's first-order condition (2.2). The first term in formula (2.3) for $h_B(\lambda)$ is the probability that the buyer, by increasing his bid, will jump over $\zeta_{(m)}$ and that $\zeta_{(m)}$ is a seller's offer. Omitting this nonnegative term from Eq. (2.2) produces the inequality

$$(v - \lambda) \leq \frac{kM(\lambda)}{(m-1)L(\lambda)\dot{v}} \tag{3.3}$$

where $B(v) = \lambda$. Now imagine the graph of the increasing function $\lambda = B(v)$ in the v, λ plane over its domain $[0, 1]$. It lies below the $\lambda = v$ diagonal, reflecting underbidding by a buyer. The amount of misrepresentation, $v - B(v)$, is the vertical distance between the graph and the diagonal. Misrepresentation increases as v increases if and only if $B'(v) < 1$, or equivalently, if $\dot{v} > 1$. Suppose $v - B(v)$ is maximized at $v' < 1$. Then at v' necessarily $\dot{v} \geq 1$, for otherwise a $v'' > v'$ would exist at which misrepresentation is larger. Thus

$$v' - B(v') \leq \frac{kM(\lambda')}{(m-1)g(v')L(\lambda')\dot{v}} \leq \frac{k}{(m-1)g(v')}\frac{M(\lambda')}{L(\lambda')} \tag{3.4}$$

where $\lambda' = B(v')$ and $\dot{v}$ is also evaluated at v'. Recall the definitions:

- $L(\lambda)$ is the probability that if $m-2$ buyers bid using strategy B and m sellers offer using strategy S, then exactly $m-1$ bids/offers are less than λ;
- $M(\lambda)$ is the probability that if $m-1$ buyers bid using strategy B and m sellers offer using strategy S, then exactly m bids/offers are less than λ.

These two probabilities are obviously similar; not surprisingly the ratio $M(\lambda)/L(\lambda)$ is bounded.[12] This, together with the assumption that $g(v)$ is continuous and positive on $[0,1]$, implies that $v' - B'(v') \leq \kappa/m$ for some $\kappa > 0$, as Theorem 1 states.

If each trader in a k-DA offered/bid his true reservation value (contrary to his self-interest), then the resulting allocation would be *ex post* classical efficient: no gains from trade could remain because a buyer would fail to trade if and only if his benefit v_i were less than the cost c_i of every seller who fails to trade.[13] Theorem 1's result that as m increases, the amount of misrepresentation decreases as κ/m, therefore, implies that in expectation the k-DA rapidly approaches—but does not reach—*ex post* classical efficiency as the number of traders increases. Theorem 2 makes this implication precise.

For given m, k, F, G, and equilibrium (S, B), and for a given sample of the traders' reservation values, the *gains from trade* realized by a k-DA is

$$\sum_{i \in T_B} v_i - \sum_{i \in T_S} c_i \tag{3.5}$$

where T_B and T_S are the sets of buyers and sellers, respectively, who successfully trade. The *expected gains from trade* is the expected value of the realized gains from trade when traders' reservation values are distributed according to F and G. The *potential expected gains from trade* is the expected gains from trade if each trader were to honestly report his reservation value rather than following his equilibrium strategy. The *relative efficiency* of an equilibrium (S, B) is its expected gains from trade divided by the potential expected gains from trade.

THEOREM 2 (RSW, Th. 6.1). Consider any equilibrium (S, B) in which trade occurs with positive probability, every seller always offers at least as much as his cost, and every buyer bids at most his value. A constant ξ exists, whose value is a function of F and G but not of m or (S, B), such that the relative efficiency of (S, B) is at least

$$1 - \frac{\xi}{m^2}.$$

Theorem 2 is complementary to an important theorem of Wilson[22]: for sufficiently large m, equilibria of the k-DA are incentive efficient, provided that the equilibria satisfy some regularity conditions. This means that after traders have learned their values but before trade occurs, it cannot be common knowledge among the traders that switching to another equilibrium or changing the rules of the auction would be Pareto improving. Wilson's theorem thus establishes that for large markets the k-DA's allocations are optimal within the constraints that the traders'

[12] Proving this is the main work of a formal proof of the theorem.

[13] See Holmstrom and Myerson[7] for a taxonomy of standards of efficiency under different informational assumptions.

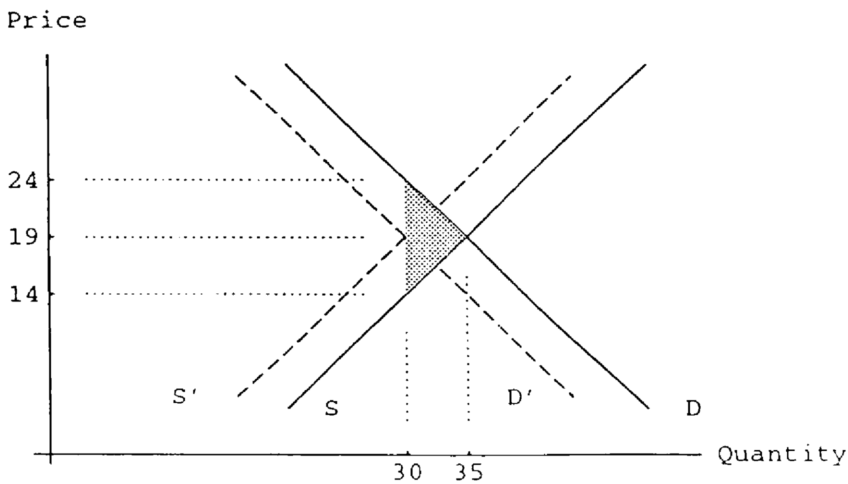


FIGURE 3 True supply and demand are the solid lines S and D. Reported supply and demand are the dashed lines S' and D'. The shaded triangle represents the gains from trade lost because S' and D' are reported.

private information and strategic incentives impose. Our Theorem 2 establishes that these constrained optimal allocations are, in fact, close to the classical, full information, optimal allocations.

Figure 3 provides a simple, supply-demand intuition for this result. In the figure the supply and demand curves drawn in solid lines represent buyers' and sellers' true market supply and demand with market-clearing quantity $q' = 35$. The dotted curves represent their reported supply and demand curves: sellers have over-reported the cost of their units by $\kappa/m = 5$ and buyers have under-reported the benefit of their units by an equal amount. If we assume that both supply curves have slope 1 and both demand curves have slope -1, then the market-clearing quantity for the reported supply and demand curves is $q'' = q' - \kappa/m = 30$. The gains from trade that are unrealized as result of trading q'' units instead of q' units is the area of the shaded triangle; consistent with Theorem 2 its area is $(\kappa/m)^2 = 25$. This is only an analogy, of course, for Figure 3 reflects neither the discreteness nor the incomplete information of our model.[14]

[14] It may be possible to deepen this analogy between our convergence to efficiency result and the area of the Harberger triangle by following Bulow and Roberts[1] who showed that results of auction theory concerning a single seller and potential bidders with unknown reservation values have direct parallels in the theory of a discriminating monopolist in standard price theory. Their analogy replaces a bidder whose value is drawn from a distribution G on a continuum with a continuum of buyers whose values are distributed according to G.

COMPUTATION AND MULTIPLICITY OF EQUILIBRIA

The traders' first-order conditions define a system of ordinary differential equations that cannot be solved in closed form but are easy to solve numerically. Computation of equilibria is important for three reasons. First, computation suggests that a continuum of smooth equilibria exist when k is in $(0, 1)$. This is important because we do not have a general proof of existence of equilibria.[15] Second, numerical examples suggest that a k-DA may be almost fully efficient with small values of m (e.g., $m = 6$ in the case of uniform F and G). Examples of such computations are presented in RSW and in Satterthwaite and Williams.[18,19] The bounds in Theorems 1 and 2 only suggest rapid convergence to efficiency; computation of κ and ξ in the theorems provides bounds on misrepresentation and inefficiency that prove to be very coarse in comparison to the actual performance of computed equilibria. Third, numerical computation produces explicit predictions of bidding behavior in the k-DA, which is surely helpful for experimental testing of this institution.

SYSTEM OF EQUATIONS DETERMINING $(\dot{c}, \dot{\lambda}, \dot{v})$

Pick a point (c, λ, v) that satisfies the inequalities $0 < c < \lambda < v < 1$. Suppose (perhaps counterfactually) that equilibrium strategies (S, B) exist such that $S(c) = \lambda$ and $B(v) = \lambda$. Pick a representative buyer i. Given that (S, B) is an equilibrium, then the buyer's first-order condition $\partial U_B(v, \lambda)/\partial\lambda = 0$ from Eqs. (2.2) and (2.3) is satisfied at (c, λ, v):

$$mK(\lambda)f(c)\dot{c} + (m-1)L(\lambda)g(v)\dot{v} = \frac{kM(\lambda)}{(v-\lambda)} \tag{4.1}$$

or, in matrix form,

$$[D_{BS}(c,\lambda,v) \quad D_{BB}(c,\lambda,v) \quad 0] \times \begin{bmatrix} \dot{v} \\ \dot{c} \\ \dot{\lambda} \end{bmatrix} = \frac{kM(\lambda)}{v-\lambda} \tag{4.2}$$

where $\dot{\lambda} = \partial\lambda/\partial\lambda = 1$. The first-order condition of a representative seller is similar. Putting the two together with the tautology $\dot{\lambda} = 1$ gives a system of ordinary differential equations[16]:

$$\begin{bmatrix} D_{BS}(c,\lambda,v) & D_{BB}(c,\lambda,v) & 0 \\ D_{SS}(c,\lambda,v) & D_{SB}(c,\lambda,v) & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \dot{v} \\ \dot{c} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} kM(\lambda)/(v-\lambda) \\ (1-k)N(\lambda)/(\lambda-c) \\ 1 \end{bmatrix}. \tag{4.3}$$

[15]The 1-DA is a special case. As noted earlier, a seller's dominant strategy in the 1-DA is honest reporting, i.e., $S^*(c) = c$. There is at most one smooth strategy B such that (S^*, B) is an equilibrium in the 1-DA, and for generic F, G a piecewise, smooth, equilibrium strategy B exists.[21] If F and G are the uniform distribution, then this equilibrium takes the simple form $S^*(c) = c$ and $B(v) = mv/(m+1)$. Uniqueness holds because, in addition to the first-order conditions, there is an initial condition that B must satisfy—$B(0) = 0$. Parallel results hold for the 0-DA.

[16]Formulas for D_{SS}, D_{SB}, and N can be found in RSW.

A smooth equilibrium (S, B) defines a solution curve to this ordinary differential equation by the formula $(c, \lambda, v) = (S^{-1}(\lambda), \lambda, B^{-1}(\lambda))$.

COMPUTING EQUILIBRIA

The system (4.3) may be numerically integrated to obtain a solution using any of a number of standard techniques.[17] A simple approach to computing solutions that represent equilibria is as follows. Pick an initial point $P_0 = (c_0, \lambda_0, v_0)$. Solve Eq. (4.3) at P_0 for the vector of derivatives $P_1' = (\dot{c}_1, \dot{\lambda}_1, \dot{v}_1)$ and pick a small, positive step $\Delta\lambda$. Compute a new point $P_1 = (c_0 + \dot{c}_1\Delta\lambda,\ \lambda_0\dot{\lambda}_1\Delta\lambda, v_0 + \dot{v}_1\Delta\lambda)$. At P_1 solve Eq. (4.3) to obtain P_2' and then to compute P_2. Continue this iterative process generating points $P_2, P_3, P_4, \ldots$, as long as neither event E_+ nor event e_+ occurs:

- **Event $\mathbf{E_+}$** occurs at point $P_{n'} = (c_{n'}, \lambda_{n'}, v_{n'})$ if the inequalities $0 < c < \lambda < v < 1$ are violated.
- **Event $\mathbf{e_+}$** occurs at point $P_{n'}$ if $P'_{n+1} = (\dot{c}_{n'+1}, \dot{\lambda}_{n'+1}, \dot{v}_{n'+1})$, the vector of derivatives, violates either $\dot{c}_{n'+1} \geq 0$ or $\dot{v}_{n'+1} \geq 0$.

If event e_+ occurs at point $P_{n'}$, then, because equilibrium strategies must be increasing, no smooth equilibrium goes through the initial point P_0. Select a new initial point at which to restart the algorithm. If event E_+ occurs, return to P_0, reverse the sign of $\Delta\lambda$ so as to generate points in the opposite direction, and iteratively generate points $P_{-1}, P_{-2}, P_{-3}, \ldots$ as long as neither event E_- nor event e_- occurs. Events E_- and e_- are defined exactly as E_+ and e_+ except in the definition of e_-, the vector of derivatives P'_{-n-1} is tested against the inequalities.

If the process terminates at point $P_{n''}$ with event E_- occurring, then the points $\{(c_{n''}, \lambda_{n''}), \ldots, (c_0, \lambda_0), \ldots, (c_{n'}, \lambda_{n'})\}$ and $\{(v_{n''}, \lambda_{n''}), \ldots, (v_0, \lambda_0), \ldots, (v_{n'}, \lambda_{n'})\}$ numerically describe equilibrium strategies S and B, respectively, that go through the initial point P_0, provided that at each of the points satisfaction of the first-order conditions are sufficient for maximizing traders' expected utilities. A condition that guarantees that the first-order conditions are sufficient for utility maximization is that F/f and $(G-1)/g$ are increasing functions on $[0, 1]$.[18]

If event E_+ occurs at point $P_{n'} = (c_{n'}, \lambda_{n'}, v_{n'})$, then $c_{n'}$ approximates the upper endpoint $\bar{c}$ of the interval over which a seller makes serious offers. Similarly, if E_- occurs at $P_n'' = (c_{n''}, \lambda_{n''}, v_{n''})$, then $v_{n''}$ is approximately $\underline{v}$, i.e., the lower endpoint of the interval over which a buyer makes serious bids. Consequently E_- and E_+ are the correct tests at which to terminate the construction of (S, B).

[17]See Press et al.,[15] ch. 15, for a sample of effective numerical algorithms.

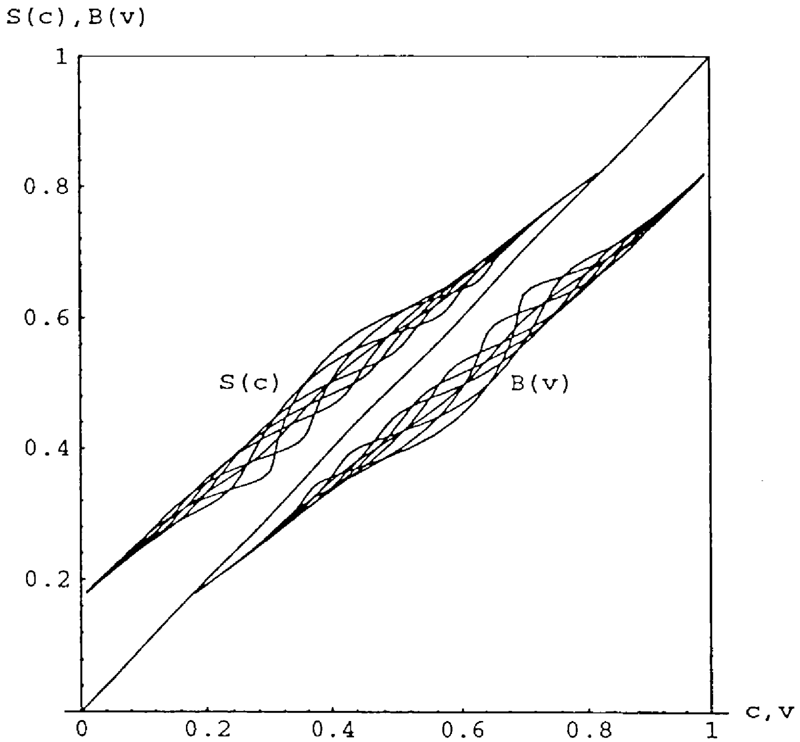


FIGURE 4 Equilibrium strategies (S, B) for $m = 2$, $k = 0.5$, and (F, G) uniform. The equilibrium (S, B) pairs depicted here approximate the full range of equilibrium behavior for $m = 2$.

MULTIPLICITY OF EQUILIBRIA

The freedom in selecting the initial point P_0 suggests that a great variety of equilibria may exist. For a given m, k, F, and G, the entire set of smooth equilibria can be represented by using the following procedure. Fix λ at the competitive price $\overline{\lambda}$ of the limiting continuum market (i.e., the solution to $F(\overline{\lambda}) = 1 - G(\overline{\lambda})$ so that the expected number of sellers below $\overline{\lambda}$ equals the expected number of buyers above). Construct a grid of initial points $(c, \overline{\lambda}, v)$ by selecting the cost c from the interval $(0, \overline{\lambda})$ and the value v from the interval $(\overline{\lambda}, 1)$. Calculate solutions to the system (4.3) for all the starting points and discard those solutions that are not equilibria, i.e., those along which either B' or S' turn negative. If the grid is made fine enough, then the resulting set of equilibria approximates the entire set of smooth equilibria.

Figure 4 shows the bundles of strategies S and B that result when this procedure is carried out for the case of $m = 2$, $k = 0.5$, and uniform F and G. Figure

5 repeats the procedure for $m = 4$. Comparison of the two figures makes obvious what Theorem 1 states must be the case: the maximal amount of misrepresentation of any of the strategies graphed for the $m = 4$ case is approximately half that of the $m = 2$ case. Note also that while an infinite family of equilibria exists for each value of m, the speed with which the equilibrium set shrinks as m increases essentially resolves, for large m, this indeterminacy.

A MODIFIED SEALED-BID DOUBLE AUCTION

Supply-demand analysis is so useful largely because it describes the outcome of trading without reference to the details of the institution through which trade is conducted. The work described to this point substantiates this analysis only for one particular class of institutions, the k-DA. Ideally we would like to establish what features of an institution insure that the outcome of trading is essentially the Marshallian price-quantity pair.

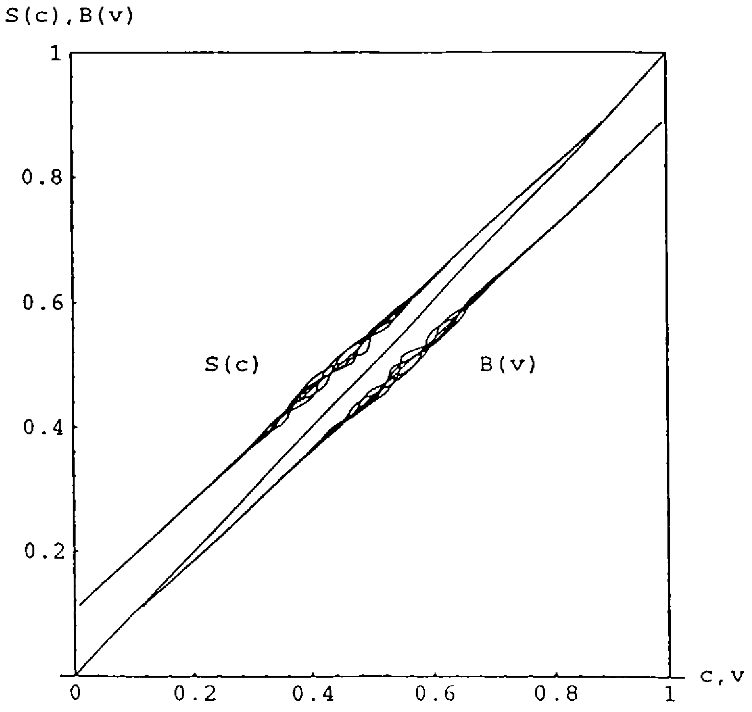


FIGURE 5 Equilibrium strategies (S, B) for $m = 4$, $k = 0.5$, and (F, G) uniform. The equilibrium (S, B) pairs depicted here approximate the full range of equilibrium behavior for $m = 4$.

A perturbation of the 1-DA that Kagel and Vogt[9] introduced for experimental use provides some insight into this question and its complexity. On initial examination the changes seem innocuous theoretically and, perhaps, helpful for the purposes of experimental work. Thorough analysis, however, shows that this modified DA (MDA henceforth) has quite different properties than the standard 1-DA. Most notably, the price in the MDA may not clear the market; as a consequence, sellers' incentives in the MDA are counterintuitive and equilibria in pure strategies apparently do not exist.

RULES OF THE MDA

The rules of the MDA are identical to the rules of the 1-DA except for one change in the algorithm for selecting the price. In the 1-DA, price is set at $s_{(m+1)}$ whether the trader who offered/bid $s_{(m+1)}$ is a seller or a buyer. The MDA sets the price at $s_{(m+1)}$ if a buyer bids $s_{(m+1)}$; otherwise, the price is set equal to the bid that exceeds $s_{(m+1)}$ by the smallest amount. In other words, the price is set equal to the smallest of the bids of those buyers who get to buy. Trade still occurs between buyers who bid at least $s_{(m+1)}$ and sellers whose offers were no more than $s_{(m)}$.

For example, suppose $m = 3$, sellers' offers are 0.43, 0.49, and 0.81, and buyer's bids are 0.64, 0.32, and 0.11. The MDA sets price at $s_{(5)} = 0.64$. This contrasts with the 1-DA, in which price would equal $s_{(4)} = 0.49$. Both mechanisms prescribe that the only trade is between the seller who offered 0.43 and buyer who bid 0.64. The price 0.64 does not clear the market, however, for the seller who offered 0.49 does not trade.

INCENTIVES TO MISREPRESENT

While a buyer in the 1-DA sets the price only if his bid equals $s_{(m+1)}$ in the set of all $2m$ bids and offers, his bid is the price in the MDA if it equals $s_{(m+j)}$ for $j \geq 1$ and no bids lie between $s_{(m)}$ and $s_{(m+j)}$. This increased likelihood of setting price increases the expected reward from underbidding, which means that a buyer in the MDA has an incentive to bid further below his value than would be the case with the 1-DA.

The change in rules also affects sellers' incentives. The MDA eliminates truthful reporting as a dominant strategy and gives a seller an incentive to make an offer below his cost. As in the 1-DA, a seller has no incentive to overbid because he cannot affect the price at which he trades. Sellers have an incentive to underbid, however, because the MDA's price is not market clearing. When a price is chosen in the MDA that exceeds a value that clears the market (i.e., whenever $s_{(m+1)}$ is a seller's offer), there is an excess supply available at the price, with only those sellers whose offers were below $s_{(m+1)}$ being able to trade. This possibility of an excess supply provides a seller with an incentive to underbid so as to include himself among those sellers who get to sell. To see this, consider again the example above in which sellers' offers are 0.43, 0.49, and 0.81, buyers' bids are 0.11, 0.32, and

0.64, price is set at $s_{(5)} = 0.64$, and seller two who offered 0.49 is excluded from trade. Suppose 0.49 is seller two's true cost c_2. After the bids/offers are opened, seller two regrets that he did not underbid by making an offer less than 0.43, for if he had done so, then he would have traded profitably at the unchanged price of 0.64. Underbidding, of course, also includes the possibility of selling at a price below one's cost; in equilibrium, a seller in the MDA weighs these two effects in determining the optimal amount by which he underbids.

CHARACTERIZATION AND COMPUTATION OF EQUILIBRIA

Characterization and computation of equilibria is straightforward for the k-DA; it is quite the opposite for the MDA, which implies that it is hard to obtain testable predictions. The source of the difficulty is that a seller (and similarly, a buyer) who checks that his offer λ is optimal must compute not only the likelihood that λ is less than $\zeta_{(m)}$, but also the likelihood that price will be set equal to $\zeta_{(m+1)}, \zeta_{(m+2)}, \ldots, \zeta_{(2m-1)}$. This implies that each trader's first-order condition requires global knowledge of other traders' strategies, not just their local properties around the offer/bid λ he is testing. This contrasts with the k-DA in which each trader's first-order condition uses only local information about other traders' strategies (as shown by Eq. (4.3)). Consequently, no simple characterization of equilibria through the first-order conditions is possible and the computational technique we used for the k-DA of constructing traders' equilibrium strategies through a series of small steps fails for the MDA.

Our experience in attempting to compute equilibria for the particular case of uniform F and G has caused us to question whether pure strategy, symmetric equilibria even exist in the MDA.[18] Computationally we found that if $S(\cdot)$ has the property that, for all $c > 0$, $S(c)$ is strictly positive, then the best response $S^*(\cdot)$ of a seller to (S, B) is, for c sufficiently close to zero, to offer $S^*(c) = 0$. But if $S(\cdot)$ has the property that, for c sufficiently close to 0, $S(c) = 0$, then the best response $S^*(\cdot)$ to (S, B) is, for all c, to offer $S^*(c) > 0$. Hence, no symmetric pure strategy equilibrium appears to exist.[19] We are uncertain if mixed strategy equilibria exist. This lack of theoretical or computational results cripples the MDA's usefulness as an institution for experimental testing because there are no solid predictions concerning what behavior it should induce from subjects.

[18] We were able to compute ε-equilibria for the MDA by constructing a sequence of buyers' and sellers' strategies through myopic adjustment. Specifically, we began with S_0 and B_0, which were not an equilibrium pair. We computed S_1, the best response of a seller to other traders playing S_0 and B_0. To construct S_1 we computed the seller's optimal offer for a variety of costs and then fitted a Chebyshev approximation to the resulting cost-offer pairs. Given S_1, we computed B_1, a buyer's best response to other traders playing S_1 and B_0. We continued this process until it converged to an ε-equilibrium, which in our very limited experience occurred quickly.

[19] This nonexistence conjecture may have little or no relevance to experimental investigations of the MDA because ε-equilibria do appear to exist that fail to be best responses only for values of c close to zero. A seller with such a value of c is almost certain to sell his unit. Thus, an experimental subject is unlikely to worry much about his offer for such values of c.

COMMENTS

Despite the similarity of the MDA with the 1-DA, its use of a price that is not market clearing makes it difficult to analyze. It is not obvious that results concerning the 1-DA have analogues for the MDA. Consideration of the MDA thus suggests the subtlety of defining the class of institutions for which supply-demand analysis meaningfully predicts the outcome of trade.

EXPERIMENTAL TESTING OF THE BAYESIAN NASH DA THEORY

From an experimental viewpoint the theory of the k-DA has at least three virtues. First, the theory is mainstream in that the Bayesian Nash solution concept is currently the most widely accepted method for modeling strategic behavior when information is incomplete. Second, the model on which the theory is based is easily translated into the lab for it fully specifies the generation of preferences, the information traders receive, the actions they may take, and the algorithm for computing the price and allocation. Third, the theory generates a number of testable hypotheses about buyer and seller behavior in symmetric equilibria and how it changes as m, the number of traders on each side of the market, increases. The three most obvious are (i) sellers offer no less than their reservation values and buyers bid no more, with offers strictly less when $k < 1$ and bids strictly more when $k > 0$; (ii) the maximal amount by which sellers overbid and buyers underbid relative to their reservation values is proportional to β/m where $\beta > 0$; and (iii) the relative efficiency of any equilibrium must be greater than $1 - \gamma/m^2$ where $\gamma > 0$.[20] Moreover, given the distributions F and G, the constants β and γ are calculable as is the entire set of equilibria with smooth strategies (S, B). If the theory is true in the positive sense, then the behavior of experimental subjects should conform to all three implications.

A DIFFICULTY IN DEVISING A CREDIBLE TEST

Despite the simplicity of the model, the directness of its transfer into the laboratory, and the precision of the predictions, it is unclear how to falsify the theory. The problem is this. The Bayesian theory of double auctions posits sophisticated behavior on the part of traders within a game of incomplete information. How to behave in a k-DA is unlikely to be transparent to an inexperienced participant. He must make careful inferences from noisy data about other traders' strategies and

[20] In addition, for any m and any (F, G), an upper bound on the relative efficiency can be computed by constructing an optimal mechanism and calculating its relative efficiency. Gresik and Satterthwaite[4] describe the construction of an optimal mechanism.

then optimize his own bidding against their imperfectly understood behavior. As a consequence, an experimental subject may need to participate in a k-DA a very large number of times before he can accomplish the learning necessary to play in accordance with the theory's predictions.

During the early stages of a subject's experience with the k-DA, he almost certainly does not recognize his optimal behavior. He may therefore fall back on his life experience with economic institutions and adopt a behavior that, by analogy, seems appropriate. This initial rule-of-thumb behavior does not falsify the theory because the possibility remains that the subject will, with sufficient experience, change his behavior in a way that brings it into conformance with the theory. Real traders, as opposed to experimental subjects, have enormous experience and therefore may consistently behave in agreement with the theory even though their learning was, in fact, painfully slow. Consequently, the question any experimental test of the theory must confront, particularly if it obtains negative results, is: How much experience must subjects have for the test to be credible?

As an example of this, Kagel and Vogt[9] had subjects play the MDA that is described above in Section 5. Sellers have an incentive in the MDA to make offers that are less than their reservation values. They, in fact, made offers that were greater than their reservation values and thus acted in accordance with the standard intuition most individuals carry into trading situations in which price is negotiated: a seller should ask for more than what the object is worth to him.[21]

Experimentalists understand that learning is critical, but they have not made much progress in defining its magnitude or speed.[22] Nevertheless, at least three complementary approaches may be taken to improve the likelihood that experimental results contradicting k-DA theory will be regarded as credible. First, theory may be developed that gives a sense of how many repetitions a subject would have to play a particular k-DA situation in order to at least have a reasonable chance of learning equilibrium behavior. Second, the experiment may be designed to make the learning process for the subjects easier. Third, subjects can be matched against computer players in order to allow controlled measurements of how fast learning proceeds in the k-DA environment.

[21] Kagel and Vogt also had subjects participate in the 1-DA. Though sellers have a dominant strategy in the 1-DA of truthful reporting, they also tended to follow the standard intuition of asking for more than their reservation values.

[22] See Kagel's review,[8] section 1.6, of learning within private value auctions. The evidence is weak, and he comes to no firm conclusions concerning either its speed or its effectiveness. More definitely, within a series of public goods experiments, Palfrey and Rosenthal[14] report that subjects demonstrated a very limited ability to learn implicitly critical statistics describing other players' strategies.

THEORY-BASED APPROACHES TO IMPROVED CREDIBILITY

To our knowledge even the most elementary notions of how long learning should take in the k-DA environment have not been developed. For example, suppose a seller has reservation value c and is trying to ascertain if changing his bid to λ'' from λ' would increase his expected utility. How many repetitions of the k-DA would he have to play in order to reject the hypothesis that λ' and λ'' yield identical expected utilities? Answers to questions such as this would give some sense of how quickly a fully rational participant could correct nonequilibrium behavior and would place a lower bound on the number of repetitions an actual, boundedly rational subject would have to play the k-DA before he could be expected to play in accordance with equilibrium predictions.[23]

AN EXPERIMENTAL DESIGN THAT AIDS LEARNING

The simplest way to design a k-DA experiment with m traders on each side of the market is to take $2m$ experimental subjects, assign half to be sellers and half to be buyers, and have them play n repetitions of the k-DA. Each repetition played is an independent event in that each time every trader independently draws a new reservation value from the appropriate underlying distribution F or G. In this setup each seller implicitly defines his entire strategy $S(\cdot)$ as he selects offers C_i in response to different values of his cost c_i. Simultaneously, each buyer is implicitly defining his entire strategy $B(\cdot)$. It is clearly difficult, however, for a trader to find his entire optimal strategy as the other traders are simultaneously adjusting their own strategies.

One way of simplifying the learning is to ask each subject to define only one point of his strategy. Specifically, assemble a large group of $2M$ subjects, with $M \gg m$, assign half to be buyers and half to be sellers, and give each subject a permanent reservation value so that the distribution of the M sellers' costs approximates F and the distribution of M buyers' values approximates G. Conduct n repetitions of the k-DA by drawing independently for each repetition a random market of m buyers and m sellers from the two large pools of M sellers and M buyers. Within each repetition the individual trader faces a market of $2m - 1$ other traders whose reservation values are distributed as they would be if new reservation values were drawn for each repetition. Since each individual seller keeps the same cost c_i for all n repetitions of the k-DA, he is required only to choose a single point $S(c_i)$ on his strategy, a much easier task than computing the entire function $S(\cdot)$. Similarly, each buyer only has to choose the optimal bid for one value.

[23]Such calculations may be particularly illuminating for the k-DA because in it a trader's bid/offer can be *ex post* non-optimal only if (i) it is one of the bids/offers that determine price, or (ii) the price is between his value and his bid/offer. Consequently, a non-optimal bid/offer only occasionally affects a trader's realized payoffs, i.e., the k-DA's feedback to traders is generally weak and sporatic. Additionally, this feedback may be misleading, for an *ex ante* optimal bid is *ex post* nonoptimal a non-negligible fraction of the time.

COMPUTER-SIMULATED PLAYERS AND LEARNING BEST RESPONSES

A more tightly controlled approach can be taken to investigate how rapidly—if at all—a trader learns an optimal strategy. Suppose, for instance, a seller played the k-DA against a set of $2m-1$ computer-simulated players who are programmed to play equilibrium strategies. Each repetition he would submit an offer simultaneously with the computer submitting the other $2m-1$ bids/offers that it calculates by substituting randomly generated reservation values into equilibrium strategies (S, B). The subject could be assigned a new, independently drawn cost c_i each repetition as in the standard design discussed above, or he could keep a single c_i across all repetitions as in the modified design.

The main advantage of having the subject play against a computer is that the subject faces a stationary problem, i.e., the subject does not face a distribution of bids and offers from the other $2m-1$ traders that is changing as they revise their strategies on the basis of their own learning.[24] Positive results from running this tightly controlled design would not establish that traders tend to play Bayesian Nash equilibrium strategies in the k-DA, for this design does not address how a group of traders can simultaneously learn equilibrium behavior. Negative results, however, would be evidence against the theory's positive validity because the ability of a single trader to learn his best response against equilibrium behavior is certainly a necessary condition for a set of traders to learn equilibrium behavior.

CONCLUDING COMMENT

The Bayesian Nash equilibrium theory of the k-DA provides some theoretical justification that the predictions of supply/demand analysis may be valid even in small markets with imperfectly informed, strategic traders. This is consistent with experimental research on a variety of other institutions—particularly the oral double auction—that show high efficiencies even with small numbers of traders. Our proof of convergence is essentially based upon the diminishing likelihood that a trader can affect the price relative to his likelihood of missing a profitable trade. This, it would seem, is the basis of the intuition that traders are compelled to act as pricetakers in any reasonable market institution. It is consequently attractive to conjecture that the incentives elicited by the k-DA may be general and drive the good experimental performance of other institutions in their richer environments.

This may be true, but as of yet the case has not been adequately made. Two points, in particular, should be kept in mind. First, our convergence results apply only to the k-DA. Our difficulty in understanding the MDA with its subtle deviation from a market-clearing price underscores this fact. Ideally we seek a theory that establishes necessary and sufficient conditions for an institution-environment pair

[24] A disadvantage of unknown importance is that a subject may play differently if he knows that he is playing against computer-simulated traders as opposed to real traders.

to exhibit the quadratic rate of convergence to *ex post* efficiency that Theorem 2 establishes for the k-DA.

Second, whether or not the Bayesian theory developed above positively describes the behavior of subjects in the k-DA is not established. If the Bayesian theory is not upheld positively in the k-DA where the theory is properly worked out, then surely it should not be used to explain the efficiency of other institution-environment pairs that have not been properly studied. Testing the theory, however, is difficult because optimal behavior in the k-DA is far from transparent to experimental subjects. Only if experienced subjects fail to behave in accordance with the theory's predictions will it be falsified. But no criterion exists to decide if a group of subjects is sufficiently experienced that data from their actions can falsify the theory. A convincing test of this theory thus seems to require an adequate theory of learning in Bayesian games.

ACKNOWLEDGMENTS

We gratefully acknowledge the useful comments of the editors, an anonymous referee, and Keith Weigelt. This material is based upon work supported by the National Science Foundation under Grant No. SES 9009546.

REFERENCES

1. Bulow, J., and J. Roberts. "The Simple Economics of Optimal Auctions." *J. Pol. Econ.* **97** (1989): 1060–1090.
2. Chatterjee, K., and W. Samuelson. "Bargaining Under Incomplete Information." *Operations Research* **31** (1983): 835–851.
3. Friedman, D. "The Continuous Double Auction Market Institution: A Survey." This volume.
4. Gresik, T., and M. Satterthwaite. "The Rate at Which a Simple Market Converges to Efficiency as the Number of Traders Increases: An Asymptotic Result for Optimal Mechanisms." *J. Econ. Theory* **48** (1989): 304–332.
5. Harsanyi, J. "Games with Incomplete Information Played by Bayesian Players, Parts I, II, and III." *Mgmt. Sci.* **14** (1967-68): 159–182, 320–334, 486–502.
6. Hayek, A. "The Use of Knowledge in Society." *Am. Econ. Rev.* **35** (1945): 519–530.
7. Holmstrom, B., and R. Myerson. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* **51** (1983): 1799–1819.
8. Kagel, J. "Auctions: A Survey of Experimental Research." Xerox, March, 1991. To appear in *The Handbook of Experimental Economics.*
9. Kagel, J., and W. Vogt. "Buyer's Bid Double Auctions: Preliminary Experimental Results." This volume.
10. Leininger, W., P. Linhart, and R. Radner. "Equilibria of the Sealed-Bid Mechanism for Bargaining with Incomplete Information." *J. Econ. Theory* **48** (1989): 63–106.
11. Marshall, A. *Principles of Economics.* London: MacMillan, 1949.
12. McCabe, K., S. Rassenti, and V. Smith. "Designing a Uniform Price Double Auction: An Experimental Evaluation." This volume.
13. Myerson, R., and M. Satterthwaite. "Efficient Mechanisms for Bilateral Trade." *J. Econ. Theory* **29** (1983): 26-281.
14. Palfrey, T., and H. Rosenthal. "Testing Game-Theoretic Models of Free Riding: New Evidence on Probability Bias and Learning." In *Laboratory Research in Political Economy,* edited by T. Palfrey, 239–268. Ann Arbor: University of Michigan Press, 1991.
15. Press, W., B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes: The Art of Scientific Computing.* Cambridge: Cambridge University Press, 1986.
16. Roberts, J., and A. Postlewaite. "The Incentive for Price-Taking Behavior in Large Exchange Economies." *Econometrica* **44** (1976): 115–128.
17. Rustichini, A., M. Satterthwaite, and S. Williams. "Convergence to Price-Taking Behavior in a Simple Market." D.P. 914, Center for Mathematical Studies in Economics and Management Science, December 1990.

18. Satterthwaite, M., and S. Williams. "Bilateral Trade with the Sealed Bid k-Double Auction: Existence and Efficiency." *J. Econ. Theory* **48** (June 1989): 107–33.
19. Satterthwaite, M., and S. Williams. "The Rate of Convergence to Efficiency in the Buyer's Bid Double Auction as the Market Becomes Large." *Rev. Econ. Stud.* **56** (1989): 477–498.
20. Schwartz, R. *Equity Markets: Structure, Trading, and Performance.* New York: Harper & Row, 1988.
21. Williams, S. "Existence and Convergence of Equilibria in the Buyer's Bid Double Auction." *Rev. Econ. Stud.* **58** (1991): 351–374.
22. Wilson, R. "Incentive Efficiency of Double Auctions." *Econometrica* **53** (1985): 1101–1115.
23. Wilson, R. "Design of Efficient Trading Procedures." This volume.

Robert Wilson
Stanford Business School, Stanford, CA 94305-5015 USA; phone: 415-723-8620; fax 725-7979; internet: fwilson@gsb-yen.stanford.edu

FIVE
Design of Efficient Trading Procedures

This chapter describes a formulation of the problem of mechanism design when the participants have private information. Allocations that are efficient within the constraints imposed by incentive compatibility and individual rationality are characterized in terms of necessary conditions. The formulation includes the design of optimal trading procedures, such as auctions, as a special case.

INTRODUCTION

One motive for economic analysis is to improve efficiency. Organizational designs, contract negotiations, and market trading procedures are significant arenas for improvements. In recent years, analyses of these topics have used new theories of mechanism design and implementation. In general, this work aims to construct for each economic environment the rules of a game that yields an efficient outcome when the participants use equilibrium strategies. In this sense the game is efficient, compared to other games that could be used, or the contract or organizational structure that embodies the rules is efficient.

The design of procedural rules invokes a different and weaker criterion for efficiency than the traditional criterion of Pareto optimality. The Pareto criterion requires the strong property that no alternative outcome could improve the welfare of every participant. How the outcome is achieved is moot; for example, a Walrasian allocation is supposedly an efficient outcome, but the issue of how markets are organized to determine prices and trades is not addressed. In contrast, the design approach includes practical matters of implementation. This implies, for instance, that monopoly profits may be unavoidable in a market with few traders, and therefore, the objective of the design is to minimize the distortionary effects of the participants' influence on prices. This modification of the Pareto criterion is sometimes called second-best optimality.

A further aspect of the design approach is that it takes explicit account of each participant's private information, including his preferences as well as other economically relevant data. That is, designs are constrained by practical limitations: some information cannot be observed by others, and some actions cannot be monitored or verified. Some environments admit mechanisms that encourage participants to reveal their private information, but usually the incentives required to promote truthful revelation impose costs in the form of Pareto inefficient outcomes. That is, like monopoly rents, informational rents are often inescapable. The design objective is again to minimize the distortionary effects of these rents. The requirement that sufficient incentives must be provided to induce participants to reveal private information, or to undertake specific actions, is called the incentive-compatibility constraint on the design.

A standard application is contracting between a principal and an agent, in which the agent's information and action are unobservable: in this case the principal's reward to the agent can be based only on observables such as output, and the agent's action (depending on his information) presumably serves his own interests as modified by the schedule of output-contingent rewards offered by the principal. The same features recur in markets. In an auction, for instance, bidders with private information about the value of the item for sale can obtain informational rents: in equilibrium each bids less than his estimate of the value, although this effect is diminished if there are many bidders.

The formulation of efficiency criteria that recognize incentive-compatibility constraints is due to Holmström and Myerson.[15] They define three forms that are increasingly restrictive. To state their criteria it is useful to identify each participant's private information with his type and to apply each criterion to outcomes that are contingent on the list of all their types. Each such list is one of the possible states of the environment, and an allocation is a function that assigns to each list an outcome contingent on that list, subject to the proviso that the allocation is implementable—in the sense that it is achievable by some procedure consistent with incentive compatibility. The criteria differ according to whether they are invoked contingent on the list itself, contingent on the possible types of each participant, or *a priori* before any participant obtains private information. The criterion for *ex post* efficiency says that an allocation is efficient if no other implementable allocation improves each participant's list-contingent welfare for every list that might occur.

The criterion for *interim* efficiency says that an allocation is efficient if no other allocation improves the type-contingent expected welfare of every participant for every type he might be. Lastly, the criterion for *ex ante* efficiency requires that no other allocation improves the expected welfare of every participant, calculated by taking the expectation over the types he might become. In practice, these criteria can be implemented by using welfare weights for the participants that are list-contingent, type-contingent, or constant. An efficient design is, therefore, one that maximizes the expectation of the weighted sum of the participants' expected benefits, subject to the incentive-compatibility constraints and any other feasibility constraints imposed by the environment. Thus, the design problem is a special kind of constrained optimization problem, as we illustrate in Section 1.

Ex post efficiency is rarely invoked, because it is a very weak criterion; for example, it ignores the benefit that one participant could gain from insuring against what others' types might be. *Interim* efficiency applies to environments in which participants' types are fixed data, though each participant's type is unknown to others. For example, this is usually the case after they have arrived at a market or contract negotiation. The strongest criterion, *ex ante* efficiency, is useful to design organizational structures or market procedures in anticipation that they will be used later in many different environments that will differ according to the types of the participants involved.

In many cases there is a specific feasibility constraint that is usefully treated in a special way. If a participant has an outside option enabling him to refuse participation in the procedure, then the procedure must assure him benefits no less than what he could get from the outside option. This is the usual case in organizational design because there is an exogenous labor market, and in exchange markets because one can refuse to trade. Participation constraints are also called individual rationality constraints. Their restrictiveness is inverse to the ordering of efficiency criteria. A participation constraint is *ex ante*, *interim*, or *ex post* according to whether the outside option can be exercised only before knowing one's type, after knowing one's type, or at the conclusion of the procedure. An *ex post* participation constraint is highly restrictive, since it enables one to reject his list-contingent outcome if it is inferior to his outside option; nevertheless, this is the case in some markets because each trader can refuse exchange even after other traders' information has been revealed by their bids. Many auctions, however, require bidders to commit to purchasing at prices no more than their offered bids (e.g., they submit sealed tenders that represent contractual commitments), and in this case an *interim* constraint is more descriptive. In Section 1 we use an *interim* participation constraint, but the methods are adaptable to other formulations.

Much of mechanism design theory relies on a simple proposition known. In practice the design prescribes an implementation in the form of a game whose rules specify the actions available to each participant and the outcome that results from each combination of actions they might choose. This is then implemented as an equilibrium in which each participant chooses an optimal strategy, namely a specification of his action depending on his information. Thus, the outcome is obtained via the translation from types to actions, and then from actions to the outcome.

From an abstract perspective, however, this is merely a single mapping from types to the outcome, namely an allocation as defined above. Thus, it suffices to adopt a formulation in which the designer chooses a rule for assigning an allocation to each list of types. This leaves unsolved the task of designing a practical implementation, but it provides a convenient test of the efficiency of each implementation that might be proposed; in addition, the form of an efficient allocation rule often provides clues about how to construct a game that implements it.

To ensure that an allocation rule is incentive compatible, one uses the requirement that each participant must have an incentive to reveal truthfully his private information. This can be appreciated intuitively by supposing that an implementation is actually used: in this case a participant essentially tells the manager of the procedure to use his reported information and his equilibrium strategy to compute the optimal action used in the game. Because the type-contingent action is optimal, a false report cannot obtain better expected benefits, and therefore, truthful reporting is optimal. A game in which each participant's equilibrium strategy is merely to report truthfully his private information is called a direct revelation game.[1]

These ingredients can be summarized as follows for the case that the efficiency criterion and the participation constraints are formulated in their *interim* forms. The objective is to maximize a weighted sum of the participants' expected utilities by choosing an allocation, represented as a function specifying the outcome for each participant as a function of the entire list of their reported types. Each participant's weight can be contingent on his reported type. The constraints that restrict the choice of allocation require incentive compatibility in the form of incentives for truthful reporting of one's type, and individual rationality in the form of assurance that each participant's type-contingent outside option is not preferred to his type-contingent outcome from the allocation. The solution of this maximization problem defines the outcome function for a direct revelation game that implements an *interim* efficient allocation. As mentioned, this formulation bypasses the practical problem of implementing the allocation as a game other than direct revelation; e.g., in an exchange context the direct revelation game represents the reduced form of the rules of an auction and the equilibrium bid and offer strategies used by the buyers and sellers.

[1] Stronger results can be obtained in some cases by exploiting the dynamic structure of the game and insisting on stronger equilibrium criteria, such as subgame perfection or exclusion of weakly dominated strategies, but we do not address these amendments here.

This formulation is representative of the constructions developed in the work of Myerson[35,36] and subsequent authors who emphasize the design of efficient trading procedures.[2] In addition, it encompasses a large literature on regulation, organizational design, contracting, and negotiation that focuses on principal-agent relationships affected by adverse selection and moral hazard.[3]

Most of this literature imposes restrictive assumptions to get clean characterizations of sufficiency conditions. The basic methodology is widely applicable, however, to obtain necessary conditions for general formulations. The key to this development is an article by Mirrlees[28] that shows how the Divergence Theorem from multivariate calculus simplifies the derivation. In this chapter we employ Mirrlees' method to present a capsule summary of necessary conditions that characterize an efficient allocation. As the setting for a general formulation, we use the context of selection of an investment project by a group of investors, each of whom might be risk averse, have private information, etc.[4] In this setting the problem emphasizes efficient selection of the project and subsequent sharing of the profits. By a simple reinterpretation, however, this formulation applies equally to a principal-agent relationship in which the project is the action undertaken by the agent in response to the incentive represented by the share of profit. The formulation also applies to the design of efficient trading procedures, such as auctions, by interpreting the project as the set of buyers and sellers who trade and the resulting prices and quantities, as we illustrate in Section 4. Thus, a unified analysis applies to investment, contracting, and market trading.

The formulation allows two features often excluded in standard formulations: multidimensional correlated private information and nonlinear utilities representing risk aversion. The formulation is static so we cannot address implementations that exploit equilibrium refinements of extensive-form games.[5] No account is taken of moral hazard associated with private actions until a partial formulation is provided in Section 3.

1. FORMULATION

Consider a group of individuals indexed by $i = 1, \ldots, m$ that is choosing one among several risky projects indexed by $j = 0, 1, \ldots, n$. The choice also requires a plan for

[2] Other key articles are by Myerson[36] regarding optimal auctions, Myerson and Satterthwaite[42] regarding bilateral exchange, and Gresik and Satterthwaite[8] regarding multilateral auctions. See also Gresik,[9,10,11,12] Guesnerie and Laffont,[13] Laffont and Tirole,[19] McAfee,[22] Wilson,[54] and Myerson.[42,38,39,40,41]

[3] See Baron,[4] Hart and Holmström,[14] and Kennan and Wilson[18] for surveys of these literatures. Palfrey[43] surveys the more general literature on implementation via Bayesian games of incomplete information.

[4] See Borch,[5] Mirrlees,[29] Mossin,[33] and Wilson.[49,53]

[5] See Moore and Repullo[31,32] and Palfrey.[43]

allocating the proceeds among the members. If the jth project is chosen, then the group's net income will be $y_j(\theta)$. This income is a random variable defined on a probability space $(T, \mathcal{T}, F)$ with generic element $\theta \equiv (t, \tau)$, where $t \equiv (t_1, \ldots, t_m)$. As described below, the variable t_i represents the ith individual's private information, while τ comprises all other stochastic features that are not observed by any member of the group until after all decisions have been taken; we omit specification of other features that are common knowledge. The group's decision rule is a pair $\langle x, s \rangle$ specifying the probabilities $x \equiv (x_j)$ assigned to the various projects and the sharing rule $s \equiv (s_{ij})$ that allocates income among the individuals depending on the project and its outcome.

We suppose that the ith individual has private information about the realization of the ith component t_i of t; we say that t_i is his *type*. Consequently, the decision rule can depend on reports $\hat{t} \equiv (\hat{t}_1, \ldots, \hat{t}_m)$ about their types submitted by the members. In addition, the sharing rule can depend on the subsequently observed realization of a random variable $z_j(\theta)$ depending on the project chosen; naturally the income y_j is measurable with respect to the outcome z_j, say, via $\overline{y}_j(z_j)$. Thus, the probability $x_j(\hat{t})$ assigned to the jth choice and i's share $s_{ij}(z_j; \hat{t})$ depend on the list $\hat{t}$ of reports, and the shares depend further on the observed outcome. These functions must satisfy the feasibility conditions that

$$x_j(\hat{t}) \geq 0 \quad \text{and} \quad \sum_j x_j(\hat{t}) = 1\,, \quad \text{and} \quad \sum_i s_{ij}(z_j; \hat{t}) \leq \overline{y}_j(z_j) \tag{1}$$

(assuming free disposal of income), for all reports $\hat{t}$, choices j, and likely outcomes z_j.[6] In practice, it may be necessary to impose lower bounds on the individuals' shares, say, $s_{ij} \geq s_i^*$; we mention later how these are included.

Individual i has preferences specified by a state-contingent utility function $u_{ij}(\cdot\,; \theta)$ that depends on the project and his share of the income. The functions $\{u_{ij}(\cdot\,; \cdot)\}$ as well as the probability distribution F, including its various conditional distributions, are assumed to be common knowledge among the members. Consequently, given any profile σ of reporting strategies $\sigma_i : t_i \mapsto \hat{t}_i$, member i's conditional expected utility is

$$\hat{U}_i(t_i; \sigma) \equiv \mathcal{E}\left\{ \sum_j x_j\big(\sigma(t)\big) u_{ij}\big(s_{ij}(z_j(\theta); \sigma(t)); \theta\big) \mid t_i \right\}.$$

Due to the revelation principle, there is no loss of generality in assuming that the decision rule is designed to induce accurate reporting by the members; that is, σ is the identity function I.[7] For this case, define $U_i(t_i) \equiv \hat{U}_i(t_i; I)$.

[6] Some formulations allow the weaker condition $\mathcal{E}\left\{\sum_i s_{ij} - \overline{y}_j\right\} \leq 0$ on the presumption that an outside party, such as a risk-neutral banker, could absorb income risks; cf. McAfee.[22]

[7] An exposition of the revelation principle is in Myerson and Satterthwaite,[42] among others.

To ensure accurate reporting the decision rule must include sufficient incentives. A decision rule is *incentive compatible* if accurate reporting is a Nash equilibrium; that is, if any one member expects others to report accurately, then accurate reporting is one of his optimal responses. We also include *individual rationality* or participation constraints, requiring that

$$U_i(t_i) \geq U_i^*(t_i) \tag{2}$$

for each i and t_i, where $U_i^*(t_i)$ represents a minimal expected utility that member i's type t_i must obtain.[8] The *feasible* decision rules are those that are incentive compatible and individually rational.[9]

By convention, $j = 0$ usually signifies a riskless null project (disband the group) that provides this utility directly: $u_{i0}(s;\theta) \equiv U_i^*(t_i)$ and $y_0(\theta) \equiv 0$. This project is one that *any* member can insist on, which makes meaningful the notion that other projects require unanimous consent of the members. In some formulations it is more accurate descriptively to suppose that each individual i has a personal action (e.g., refusal to trade) that he can take unilaterally to ensure that his utility is $U_i^*(t_i)$. This difference is immaterial since, in any case, all that matters is that each individual i must be assured utility at least $U_i^*(t_i)$. The participation constraint can be strengthened to coalitional rationality, requiring that the decision rule is in the core of the associated cooperative game as in Wilson,[51] but we omit this complication here.

A feasible decision rule is *incentive efficient* if it is common knowledge among the members that no other feasible decision rule would benefit some without harming others. In technical terms, a decision rule is incentive efficient if there exist nonnegative type-contingent welfare weights $\{\lambda_i(t_i)\}$, not all zero for any state t, such that the rule maximizes the welfare measure $\mathcal{W} = \mathcal{E}\left\{\sum_i \lambda_i(t_i)U_i(t_i)\right\}$ on the domain of feasible rules. If these weights are *not* type contingent, then the rule is *ex ante* incentive efficient. Individual rationality and stronger participation constraints such as coalitional rationality can be interpreted as imposing constraints on the allowable welfare weights.

Our objective in the subsequent analysis is to characterize the incentive-efficient decision rules for a general class of models. This requires several regularity assumptions.

1. A feasible decision rule exists (e.g., selection of the null project).
2. The net income $y_j(\theta)$ or $\overline{y}_j(z_j)$ from each project is a real vector of dimension ℓ.

[8]The formulation can be adapted to participation constraints that impose lower bounds either *ex post*, say, $\sum_j x_j u_{ij} \geq U_i^*(t)$ for each t, or *ex ante*, say, $\mathcal{E}\{U_i(t_i)\} \geq U_i^*$.

[9]For expositions of the concepts of incentive compatibility, individual rationality, and *interim* and *ex ante* incentive efficiency, see Holmström and Myerson.[15] They define a stronger concept of durability that we do not address here. These constructions differ from d'Aspremont and Gérard-Varet[2,3] mainly in the imposition of individual rationality constraints.

3. Each individual i's type t_i is a real vector of dimension K_i. The conditional support of t is a rectangle $D = \{t \mid a \le t \le b\}$ that is independent of τ. Let $D = D_1 \times \cdots \times D_m$, where $D_i \equiv \{t_i \mid a_i \le t_i \le b_i\} \subset \Re^{K_i}$ is the support of i's type, and use ∂D_i to denote the boundary of D_i.
4. The joint conditional distribution $F(t \mid \tau)$ and the marginal distributions $F_i(t_i)$ have density functions $f(t \mid \tau)$ and $f_i(t_i)$ that are positive and continuously differentiable with uniformly bounded derivatives on the domains D and D_i. For later reference, define
$$\phi_i(\theta) \equiv \frac{\partial f(t \mid \tau)/\partial t_i}{f(t \mid \tau)} - \frac{\partial f_i(t_i)/\partial t_i}{f_i(t_i)},$$
where $\theta \equiv (t, \tau)$ and the indicated partial derivatives are gradient vectors. Note that $\mathcal{E}\{\phi_i(\theta) \mid t_i\} \equiv 0$.
5. Each utility function u_{ij} is bounded, increasing, and concave with respect to the share s and continuously differentiable with uniformly bounded derivatives with respect to the share and i's type t_i. For later reference, define the gradient vector
$$v_{ij}(s; \theta) \equiv \frac{\partial u_{ij}(s; t, \tau)}{\partial t_i}.$$

These assumptions suffice, via the Lebesgue dominated convergence theorem, to assure that differentiation with respect types can be permuted with respect to expectation over other random variables, which we use repeatedly. Assumptions 3 and 4 exclude the possibility that the support of one member's type is restricted by the observation of others' types. In particular, they exclude the case that some members have inferior information (e.g., t_1 is t_2-measurable) or no private information, but the formulation can be amended to address this special case.[10]

We use the notation included in assumptions 4 and 5 to specify the requirement of incentive compatibility in a convenient form. Incentive compatibility requires for each member i and each type t_i that
$$U_i(t_i) = \max_{\hat{t}_i} \hat{U}_i(t_i; \sigma)\Big|_{\sigma = I},$$
where the possible reports $\hat{t}_i$ are variations of i's strategy σ_i. The Envelope Theorem asserts, therefore, that the gradient vector U_i' of U_i satisfies
$$U_i'(t_i) = V_i(t_i) \qquad \text{where} \qquad V_i(t_i) \equiv \frac{\partial \hat{U}_i(t_i; \sigma)}{\partial t_i}\Big|_{\sigma = I}. \tag{3}$$

[10] Essentially, a member with inferior information need not report his type; examples are in Kennan and Wilson[18] among many others. In subtler formulations than the one here, individuals' overlapping information can be used to check partially the veracity of the report made by each. The extreme case of nonexclusive information is studied by Postlewaite and Schmeidler.[44]

Interchanging the order of differentiation and expectation, direct computation yields[11]

$$V_i(t_i) \equiv \mathcal{E}\left\{\sum_j x_j(t)\left[u_{ij}\left(s_{ij}(z_j(\theta);t);\theta\right)\phi_i(\theta) + v_{ij}\left(s_{ij}(z_j(\theta);t);\theta\right)\right] \mid t_i\right\}.$$

Given the form (3) of the incentive-compatibility constraint as a differential equation for U_i, a single boundary condition is needed for each i to ensure that U_i corresponds to its original definition. For this we use the equality

$$\mathcal{E}\left\{\sum_j x_j(t)u_{ij}(s_{ij}(z_j(\theta);t);\theta) - U_i(t_i)\right\} = 0\,, \tag{4}$$

called the *consistency* constraint, which could alternatively be specified as an inequality ($\geq$) since its only effect is to ensure that each member's utility is no more than the decision rule provides.

2. DERIVATION

Our purpose in this section is to establish a version of Myerson's principle that the net effect of private information is to substitute "virtual" utilities for the members' utility functions in the construction of an incentive-efficient decision rule. We first proceed generally and then derive more specific conclusions for models with simplifying features in Section 3.

The maximization problem that characterizes an incentive-efficient decision rule can be posed as follows. Given contingent welfare weights, the objective is to choose a feasible decision rule that maximizes the welfare measure $\mathcal{W}$. Individual rationality (2) and incentive compatibility (3) impose the constraints $U_i(t_i) \geq U_i^*(t_i)$ and $U_i'(t_i) = V_i(t_i)$ uniformly, and in addition the utility assignments must satisfy consistency (4).

[11]Differentiate $\int\int \sum_j x_j u_{ij}(s_{ij};t,\tau)f(t \mid \tau)\,dt\,dF_0(\tau)/f_i(t_i)$ with respect to t_i. In addition to the first-order necessary condition, a member's second-order necessary condition imposes the further constraint that $U_i''(t_i) - \partial^2\hat{U}_i(t_i;I)/\partial t_i^2$, or equivalently $\partial V_i(t_i)/\partial\hat{t}_i$, is positive semi-definite. For instance, in Myerson and Satterthwaite[42] and Rochet,[47] $\hat{U}_i$ is linear in t_i so U_i must be convex. Guesnerie and Laffont[13] and McAfee and McMillan[20] reduce this condition to a monotonicity constraint on the allocation; such conditions are related to the condition of Bayesian monotonicity required for implementation of general social choice functions; cf. Jackson[16] and Mookerjee and Reichelstein.[30] More generally, accurate reporting must be optimal globally. Although this constraint is binding in some applications, we ignore it here because the net effect is simply to modify the allocation obtained without this constraint to enforce monotonicity; cf. Guesnerie and Laffont,[13] Mussa and Rosen,[34] and Wilson.[56]

For this specification, an associated Lagrangian expression is

$$\mathcal{L} \equiv \mathcal{E}\left\{\sum_i \left[\lambda_i(t_i)U_i(t_i) + \rho_i(t_i)[U_i(t_i) - U_i^*(t_i)] + \mu_i\left[\sum_j x_j u_{ij} - U_i(t_i)\right]\right]\right\}$$
$$+\sum_i \int_{D_i} [U_i'(t_i) - V_i(t_i)] \cdot \psi_i(t_i)\, dt_i\,,$$

where the dot $(\cdot)$ indicates an inner product, and in the expectation some arguments are omitted for simplicity. This expression can be written out in full by using the previous formula for V_i. The scalar constants μ_i, the scalar-valued functions $\rho_i(t_i)$, and the vector-valued functions $\psi_i(t_i)$ are Lagrange multipliers for the corresponding constraints indicating consistency (4), individual rationality (2), and incentive compatibility (3). Note that ρ_i augments the welfare weight λ_i sufficiently to ensure individual rationality.

We interpret the Lagrangian as the objective function for a problem in the calculus of variations in which the functions U_i are to be optimized along with the decision rule. The analysis then has two parts: the first characterizes the optimal decision rule in terms of the multipliers, and the second determines the multipliers.

THE OPTIMAL DECISION RULE

To characterize the optimal decision rule, we collect the relevant terms from the Lagrangian into Myerson's *virtual utilities*, defined as:

$$\check{u}_{ij}(s;\theta) \equiv u_{ij}(s;\theta) - [u_{ij}(s;\theta)\phi_i(\theta) + v_{ij}(s;\theta)] \cdot \alpha_i(t_i)\,,$$

where $\alpha_i(t_i) \equiv [\mu_i f_i(t_i)]^{-1}\psi_i(t_i)$ (assuming $\mu_i > 0$). Then:

- For each project j and each likely outcome z_j, the shares $\{s_{ij}(z_j;t)\}$ maximize $\mathcal{E}\left\{\sum_i \mu_i \check{u}_{ij}(s_{ij};\theta) \mid z_j, t\right\}$ subject to $\sum_i s_{ij} \leq \overline{y}_j(z_j)$. Call this maximum $\check{u}_j(z_j;t)$.

This maximization may be subject to lower bounds on the shares, say, $s_{ij} \geq s_i^*$; if so, and no shares are feasible, then $\check{u}_j(z_j;t) \equiv -\infty$. The maximum $\check{u}_j$ must be interpreted properly as a supremum when no bounds are imposed on the shares.

Note that no separate provision is made for side bets or randomization. This is because all observable events on which such contingent transfers can be based are assumed to be included as part of z_j; in particular, the shares s_{ij} are required to be measurable with respect to the observed event z_j and the reports t.[12] The advantages of risk sharing usually arise from concavity of the utility functions u_{ij}. Here, however, we see that it is concavity of the virtual utilities $\check{u}_{ij}$ that is the source

[12] Myerson[36] exemplifies an alternative approach in which exogenous randomization is included by conducting the analysis in terms of the concave hulls of the virtual utilities.

of risk-averse behavior. If some members' virtual utilities are convex functions of their shares, then they will absorb all income risks.

- Given an optimal sharing rule for each project, the preferred projects are those that maximize $\mathcal{E}\left\{\breve{u}_j(z_j(\theta);t) \mid t\right\}$. Let $x(t) \equiv (x_j(t))$ be a feasible assignment of probabilities to the optimal projects.

This calculation is subtler than it appears. Ordinarily it suffices to select a single project, say, $x_j(t) = 1$, for one optimal project $j = j(t)$, but if no single optimal project satisfies all the individual rationality constraints, then randomization is necessary.

THE LAGRANGE MULTIPLIERS

We follow Mirrlees'[28] method to characterize the multipliers. In the present case, the Divergence Theorem states that

$$\int_{D_i} U_i'(t_i) \cdot \psi_i(t_i)\, dt_i + \int_{D_i} U_i(t_i)\nabla \cdot \psi_i(t_i)\, dt_i = \int_{\partial D_i} U_i(t_i)[\psi_i(t_i) \cdot d\xi]\,.$$

In the second term the divergence (trace of the Jacobian) of ψ_i is $\nabla \cdot \psi_i \equiv \sum_k \partial\psi_{ik}/\partial t_{ik}$ if $t_i \equiv (t_{ik})_{k=1,\ldots,K_i}$; that is, $\nabla = (\partial/\partial t_{ik})$. In the last term, the vector differential $d\xi$ is outward-normal to the boundary ∂D_i, so the integrand is proportional to $U(t_i)\nu_i(t_i)\cdot\psi_i(t_i)$, where $\nu_i(t_i)$ is the unit outward-normal vector at $t_i \in \partial D_i$. The rectangular shape of D_i implies that on the boundary where only the kth component of t_i is constrained, the integrand includes only the kth component of $\psi_i(t_i)$, and with a negative or positive sign according as the component t_{ik} is constrained below or above.

Applying the Divergence Theorem to the terms involving i's type-contingent expected utility yields

$$\begin{aligned}&\int_{D_i} \{[\lambda_i(t_i) + \rho_i(t_i) - \mu_i]U_i(t_i)f_i(t_i) + U_i'(t_i) \cdot \psi_i(t_i)\}\, dt_i \\ &= \int_{D_i} U_i(t_i)\{[\lambda_i(t_i) + \rho_i(t_i) - \mu_i]f_i(t_i) - \nabla \cdot \psi_i(t_i)\}\, dt_i + \int_{\partial D_i} U_i(t_i)[\psi_i(t_i) \cdot d\xi]\,.\end{aligned}$$

Consequently, the right side of this equality must be zero at an optimum choice of U_i. That is, pointwise optimization of $U_i(t_i)$ at each type $t_i \in D_i$ requires

$$\nabla \cdot \psi_i(t_i) = [\lambda_i(t_i) + \rho_i(t_i) - \mu_i]f_i(t_i)\,. \tag{5}$$

In addition, at a boundary type $t_i \in \partial D_i$, the transversality condition requires that $\nu_i(t_i) \cdot \psi_i(t_i) \leq 0$, and $= 0$ if $U_i(t_i) > U_i^*(t_i)$. In particular, the rectangular shape of D_i implies that $\psi_{ik}(t_i) = 0$ on the boundary where t_{ik} is bounded and $U_i(t_i) > U_i^*(t_i)$.

In principle, these conditions are sufficient to determine the variable multiplier ψ_i on the domain D_i given any particular specification of the constant multiplier μ_i. This is obvious when i's type is one-dimensional because then Eq. (5) specifies an ordinary differential equation and the transversality condition selects one of its solutions. The task is subtler when the type is multidimensional. In this case the single partial-differential equation (5) allows many solutions for the K_i components of ψ_i. Recall, however, that V_i must be the gradient of U_i, and therefore the Jacobian of V_i must be the Hessian of U_i; consequently, at each t_i the Jacobian of $V_i(t_i)$ must be a symmetric matrix. Thus, one must select a solution for ψ_i that, via its effect on the decision rule, yields a Jacobian of V_i that is symmetric.[13] This assures that the vector field represented by V_i can be integrated to obtain the function U_i.

Lastly, the consistency condition (4) determines the correct value of the constant scalar multiplier μ_i. This can be stated in a more convenient form by using Eq. (5) to obtain

$$\mathcal{E}\left\{\sum_j x_j[\lambda_i u_{ij} - \mu_i \check{u}_{ij}]\right\} = 0\,. \tag{6}$$

This makes explicit the dependence on μ_i, but one must remember that the decision rule $\langle x, s\rangle$ also depends on (μ_i).

The next section applies these general conditions to models with special structural features.

3. APPLICATIONS

We describe several general categories that illustrate structural features. Kennan and Wilson[18] study the first in contexts of economic exchange and negotiation of legal disputes; and Cremér and McLean[7] and McAfee and Reny[23] study the second in the context of auctions. Guesnerie and Laffont[13] study the third, fourth, and fifth in the context of principal-agent problems; and McAfee,[22] Myerson and Satterthwaite,[42] and Wilson[54] among many others study these in the context of exchange.

COMMON-VALUE MODELS. If each member's utility is independent of the type vector t, then $\check{u}_{ij}(s;\theta) = u_{ij}(s;\tau)[1-\alpha_i(t_i)\cdot\phi_i(\theta)]$. Further, if each member's type is distributed independently of others' types conditional on τ, then ϕ_i depends only on (t_i, τ).

[13] Methods for solving such differential equations and some solved examples in the context of nonlinear pricing are described in Wilson.[56] Recall that if utilities are linear in the type parameter, then the members' second-order conditions require that the Jacobian must also be positive semi-definite.

PRIVATE-VALUE MODELS. Private-value models assume that each member's utility depends only on his own type, which is uninformative about the outcome z_j. In this case, $\check{u}_{ij}(s;t) = u_{ij}(s;t_i)[1 - \alpha_i(t_i) \cdot \phi_i(t)] - \alpha_i(t_i)v_{ij}(s;t_i)$. Note that correlation enters via $\phi_i(t)$.

INDEPENDENT TYPES. If each member's type is distributed independently of others' types and τ, then $\phi_i(\theta) = 0$.

ONE-DIMENSIONAL TYPES. Suppose that each member's type t_i is a scalar, in which case ψ_i is also scalar-valued. For many of the models in the literature, U_i^* is nonincreasing and U_i is increasing.[14] In this case, the individual rationality constraint is not binding at the upper boundary where $t_i = b_i$. Combining this with the condition (5) on the interior yields

$$\psi_i(t_i) = \mathcal{E}\left\{\mu_i - \lambda_i(\tilde{t}_i) - \rho_i(\tilde{t}_i) \;\middle|\; \tilde{t}_i \geq t_i\right\}[1 - F_i(t_i)],$$

where $\rho_i(t_i) = 0$ for those types $t_i > t_i^*$ above the highest type t_i^* for whom the individual rationality constraint is binding. For instance, if the criterion is *ex ante* efficiency, then

$$\psi_i(t_i) = [\mu_i - \lambda_i][1 - F_i(t_i)] \qquad \text{and} \qquad \alpha_i(t_i) = \frac{[1 - \lambda_i/\mu_i][1 - F_i(t_i)]}{f_i(t_i)},$$

for those types above t_i^*.

LINEAR UTILITIES. If the members' utilities are similarly linear in income, say, $u_{ij}(s,\theta) \equiv p(\theta) \cdot s + w_{ij}(\theta)$, then $\check{u}_{ij}(s,\theta) = \check{p}_i(\theta) \cdot s + \check{w}_{ij}(\theta)$ where $\check{p}_i = (\check{p}_{ih})$,

$$\check{p}_{ih} \equiv p_h - \left[p_h\phi_i + \frac{\partial p_h}{\partial t_i}\right] \cdot \alpha_i \qquad \text{and} \qquad \check{w}_{ij} \equiv w_{ij} - \left[w_{ij}\phi_i + \frac{\partial w_{ij}}{\partial t_i}\right] \cdot \alpha_i$$

are analogous virtual representations. In this case, even if the price p is constant, the virtual prices $\check{p}_i$ affect the optimal sharing rule when the types are correlated.

Because this case occurs prominently in applications, we mention its chief simplifying feature when the criterion is *ex ante* efficiency and members' types are independent. In this case $\phi_i \equiv 0$, and therefore $\mu_i = \mu$ for all i, since, otherwise, the Lagrangian would be unbounded. The decision rule that maximizes the group's *ex ante* expected surplus is obtained from uniform constant welfare weights, say, $\lambda_i = 1$ for all i; some applications also address the analogous monopoly formulation in which $\lambda_i = 0$ except for one member, say, $i = 1$, for whom $\lambda_1 = 1$. To determine the single scalar multiplier μ in a way that obviates specification of the sharing rule,

[14] See Guesnerie and Laffont[13] and Wilson[56] for various sufficient conditions and Milgrom and Shannon[27] for necessary and sufficient conditions.

the members' consistency conditions (4) or (6) can be replaced by the aggregate condition

$$\mathcal{E}\left\{\sum_j x_j \textstyle\sum_i [\lambda_i u_{ij} - \mu \check{u}_{ij}]\right\} = 0 \text{ or } \mathcal{E}\left\{\sum_j x_j \sum_i \left[u_{ij} - \beta_i[u_{ij}\phi_i + v_{ij}]\right]\right\} = 0\,, \tag{7}$$

where, if Eq. (5) is interpreted as yielding $\alpha_i(t_i;\mu_i)$ dependent on μ_i, $\beta_i(t_i) = \alpha_i(t_i;\infty)$. For instance, if the types are one-dimensional as above, then

$$\beta_i(t_i) = \frac{\mu}{\mu - \lambda_i}\alpha_i(t_i;\mu) = \frac{1 - F_i(t_i)}{f_i(t_i)}$$

for those types above t_i^*. The aggregate condition (7) asserts that μ is the ratio of the aggregates of the members' actual and virtual expected utilities, which generally exceeds unity due to the informational rents obtained by members. And, it is determined by the condition that the decision rule obtained from μ yields a nil aggregate virtual utility calculated using $\alpha_i(t_i;\infty)$.[15]

MORAL HAZARD. Moral hazard is not explicitly included in the formulation above. We describe briefly one way to include it.[16] Allow each member's utility function u_{ij} to depend also on all members' private actions $a_j \equiv (a_{ij})_{i=1,\ldots,m}$ taken if the non-null action j is selected. Also, allow that the probability distribution of the outcome is conditioned on a_j via a joint distribution of the form $F(t)F_j(\tau \mid t, a_j)$. In this case, i's strategy is augmented by a second component: besides the strategy σ_i specifying his report $\hat{t}_i$ depending on his type t_i, he chooses an action strategy $a_i \equiv (a_{ij})_{j=1,\ldots,n}$ of intended actions for each non-null project, depending on both his actual type t_i and subsequently contingent on the reports $\hat{t}$ if these are observable. Whether or not a member's intended actions are reported, the key feature is that the actual actions taken need not be subsequently observable (i.e., the action a_{ij} taken by i need not be measurable with respect to the observed outcome z_j); consequently, the sharing rule remains a function only of the reports and the subsequently observed outcome. A decision rule is now a triplet $\langle x, s, a\rangle$. Given any specified decision rule, i's type-dependent utility is supposed to reflect maximization with respect to both his report and his intended actions; consequently,

[15] See Myerson and Satterthwaite[42] and Kennan and Wilson[18] for other presentations of this condition. In Section 4 we present it in terms of a normalized version ψ_i° of ψ_i. In Myerson and Satterthwaite, $\mu - 1$ is the multiplier for a single constraint summarizing both consistency and incentive compatibility. Gresik and Satterthwaite[8] and Milgrom[24] are two of several studies of exchange contexts showing that informational rents vanish as the number of participants with (conditionally) independent information increases.

[16] See Mirrlees.[29] Using essentially this formulation, Laffont and Tirole[19] provide a completely solved example in the context of contracting between a principal and agent; however, they also use some features peculiar to the example.

the Envelope Theorem again implies that $U_i'(t_i) = V_i(t_i)$, where, of course, V_i is calculated using the specified decision rule. Thus, seemingly the only modification in the procedure occurs in the optimization of the decision rule, which now includes a choice of the members' action strategies; in particular, one optimizes with respect to the virtual utilities, and each member i's strategy a_i must be measurable with respect to $(t_i, \hat{t})$—or on the presumption that reports are accurate and t-measurable. Nevertheless, it is well known that strong assumptions are required to assure that this approach yields valid necessary conditions for the design problem that can be expressed in terms of the first-order necessary conditions for the members' optimization of their private actions; cf. Rogerson[48] and Jewitt.[17]

4. AUCTIONS

An auction is the special case in which a project and its sharing rule assign transactions to the traders in a market. In this section we describe applications of the preceding results to the context of pure exchange in which there is no production ($y_j \equiv 0$) and no post-trade information ($z_j \equiv 0$) on which to condition contracts. A decision rule $\langle x, s\rangle$ is represented hereafter as an assignment $\langle q, P\rangle$ that specifies the commodity bundle $q_i(t)$ received and the payment $P_i(t)$ made by trader i when t is the list of the traders' reported types. Similarly, represent a trader's utility as $u_i(q_i, -P_i; \theta)$ on the assumption that each trader cares only about his own components of the assignment. To exclude a motive for randomization, assume that u_i is smooth, increasing, and concave with derivatives that increase with θ; further, the components of θ are affiliated random variables with increasing hazard rates (if they are independent) or monotone likelihood ratios (if correlated).[17]

To simplify we assume that utilities are linear in money, say, $u_i(q_i, -P_i; \theta) \equiv u^i(q_i; \theta) - P_i$. Also, each trader's individual rationality constraint is represented by his option to forego participation in the market and thereby obtain the null trade, whose utility is normalized to zero: $U_i^*(t_i) \equiv u_i(0, 0; \theta) \equiv 0$. For the welfare criterion we use *ex ante* efficiency throughout: $\lambda_i(t_i) \equiv \lambda_i$.

Suppose there are ℓ commodities in addition to money, which is one-dimensional. For each trader i the set of feasible trades of commodities is a subset $Q_i \subset \Re^\ell$ and we assume that the set of feasible monetary payments is $\Re$. The set of feasible assignments is restricted by the aggregate feasibility conditions

$$\sum_i q_i = 0 \qquad \text{and} \qquad \sum_i P_i = 0 .$$

[17] Cf. Milgrom and Weber.[26] Affiliation requires nonnegative correlation on every rectangle. Its principal implication is that the conditional expectation of any increasing function of θ is increasing in the conditioning variables. Milgrom and Shannon[27] provide the necessary and sufficient conditions for the requisite monotonicity properties that obviate randomization.

Due to these restrictions, the optimal auction design is obtained by solving an amended version of the Lagrangian problem in Section 2 in which additional multipliers $\pi(t)$ and $\nu(t)$ are used to include the two aggregate feasibility conditions. The necessary conditions in this case are best stated in terms of the normalized multipliers $\rho_i^\circ \equiv \rho_i/[\mu_i - \lambda_i]$ and $\psi_i^\circ \equiv \psi_i/[\mu_i - \lambda_i]$. The analog of Eq. (5) is therefore

$$\nabla \cdot \psi_i^\circ(t_i) = [\rho_i^\circ(t_i) - 1] f_i(t_i) , \tag{8}$$

together with the analogous transversality conditions. The remaining conditions can then be stated in terms of $\beta_i(t_i) \equiv \psi_i^\circ(t_i)/f_i(t_i)$. In particular, $\alpha_i(t_i) \equiv R_i \beta_i(t_i)$ where $R_i \equiv [\mu_i - \lambda_i]/\mu_i$ is called the Ramsey number in theories of nonlinear pricing, as we explain below.

Initially we address the case that each $Q_i = \Re^\ell$. In this case, using the assumption that utilities are linear in money, the necessary conditions for finite optimal choices of the payments and the allocation are:

$$\mu_i[1 - R_i \phi_i(t_i) \beta_i(t_i)] = \nu(t) , \tag{9}$$

$$\mathcal{E}\left\{ u_q^i(q_i(t);\theta) - \left[\frac{\mu_i R_i}{\nu(t)}\right] u_{qt}^i(q_i(t);\theta) \cdot \beta_i(t_i) \;\middle|\; t \right\} = \pi(t)/\nu(t) , \tag{10}$$

where the subscripts q and qt indicate partial differentiation with respect to q_i and t_i. Similar conditions result when utility is nonlinear in money, but they are expressed in terms of the marginal rates of substitution between money and the commodities.

THE ROLE OF NONLINEAR PRICING

We first describe the principle that the design of an efficient auction is a version of nonlinear pricing.[18] This principle is quite general, but it suffices here to illustrate the correspondence for the special case of independent private values. That is,

$$u^i(q_i;\theta) \equiv u^i(q_i, t_i) \qquad \text{and} \qquad F(\theta) \equiv F_1(t_1) \cdots F_m(t_m) .$$

This case has the simplifying feature derived from Eq. (9) that $\mu_i = \mu$ and $\nu(t) = \mu$ for each contingency t. Consequently, Eq. (10) can be written as

$$u_q^i(q_i(t), t_i) - R_i u_{qt}^i(q_i(t), t_i) \cdot \beta_i(t_i) = \frac{\pi(t)}{\mu} . \tag{11}$$

To relate these conditions to nonlinear pricing, we address a simple version of the general problem studied by Mirrlees.[28] Consider a regulated firm that incurs the cost $p \cdot x$ to supply a commodity bundle x to any customer in any of

[18] Bulow and Roberts[6] obtain analogous results in the case that indivisible items are traded. Mirrlees[28] and Wilson[55] present expositions of the theory of nonlinear pricing.

several markets indexed by i. There are many customers in each market, each identified by his type t_i, and the variety of types in market i is given by the distribution function $F_i(t_i)$. The firm's objective is to maximize the aggregate of consumers' surplus subject to the constraints that its net revenue is r_i in each market i. To do this the firm offers an outlay schedule O_i in market i that requires a customer to pay $O_i(x)$ for the bundle x. Customer t_i chooses the bundle $x_i(t_i)$ that maximizes his net benefit $u^i(x, t_i) - O_i(x)$ and pays $P_i(t_i) = O_i(x_i(t_i))$. As shown by Mirrlees[28] and Guesnerie and Laffont,[13] the firm's problem can be posed as follows, using $U_i(t_i) \equiv \max_x u^i(x, t_i) - O_i(x)$. The objective is to maximize the consumers' surplus $\sum_i \int_{D_i} U_i(t_i)\, dF_i(t_i)$ subject to the participation constraints $U_i(t_i) \geq 0$, the incentive constraints $U_i'(t_i) = u_t^i(x_i(t_i), t_i)$, and the revenue constraints $\int_{D_i} [O_i(x_i(t_i)) - p \cdot x_i(t_i)]\, dF_i(t_i) = r_i$, where $O_i(x_i(t_i)) \equiv u^i(x_i(t_i), t_i) - U_i(t_i)$. For this maximization, the firm chooses the assignments $U_i(t_i)$ and $x_i(t_i)$ of the utility and bundle obtained by each type t_i in each market i. To accomplish this, one uses multipliers $\rho_i(t_i)$, $\psi_i(t_i)$, and μ_i for the three constraints to form a Lagrangian objective as above.

We now indicate how solutions to this nonlinear pricing problem provide the solution to the auction design problem. In the nonlinear pricing problem, let $x_i(t_i; p, r_i)$ be type t_i's assigned bundle when the firm's cost vector is p and the revenue requirement in market i is r_i. This allocation satisfies the necessary condition

$$u_q^i(x_i, t_i) - R_i(p, r_i) u_{qt}^i(x_i, t_i) \cdot \beta_i(t_i) = p\,, \tag{12}$$

analogous to Eq. (11), where $R_i(p, r_i)$ is the Ramsey number chosen to ensure that the revenue requirement is met. The parallel between Eqs. (11) and (12) indicates that the auction design can be constructed by adjusting the cost vector p and the revenue requirements r_i in each contingency t so that $p(t) = \pi(t)/\mu$ and $R_i(p(t), r_i(t)) = R_i$. If the traders are identical *ex ante* (namely u^i, F_i, and λ_i are independent of i), then the latter is a simple one-dimensional problem because R_i is independent of i in both the nonlinear pricing and auction design problems, so it suffices to solve for a single revenue requirement $r(t)$ that is the same for all traders.

This construction has the following interpretation. In the auction there are only m traders rather than many customers in several markets, and each trader i has type t_i; moreover, feasibility requires that if the list of their reported types is t, then $\sum_i q_i(t) = 0$ and $\sum_i P_i(t) = 0$. For each specification $r = (r_i)$ of the revenue requirements, the commodity-balance constraint is satisfied by using the appropriate cost $p(t, r)$ in the nonlinear pricing problem. And given the cost $p(t, r)$ the money-balance constraint is met by ensuring constancy of the Ramsey numbers; that is, choose $r(t)$ to solve $R_i(p(t, r), r_i) = R_i$ for each trader i, which yields $p(t) = p(t, r(t))$.[19]

[19]This assures that money balance is feasible but does not necessarily construct the actual payments, which relies on the technique in Gresik and Satterthwaite.[8] The incentive conditions determine only each type's expected payment. To complete the construction, each trader is assigned an additional side payment whose expectation is zero for each type and that depends only

Thus, to design an optimal auction, one first constructs the solutions to a family of nonlinear pricing problems parameterized by the cost vector p and the revenue requirement r, using the distribution F of potential types to describe the variety of customers. When the m actual traders appear at the auction, one can use either of two procedures. In the direct revelation game, they report their types t and are then assigned bundles $q_i(t)$ and payments $P_i(t)$ from the nonlinear pricing schedule (and possibly side payments as in footnote [19]) based on the cost vector $p(t)$ and revenue requirement $r(t)$ that assure feasibility. In fact, for each trader the assigned bundle is an optimal choice from the nonlinear pricing schedule based on that cost and revenue requirement. Consequently, an alternative implementation has each trader i submit a schedule $q_i(t_i; p, r_i)$ indicating his preferred choice from each possible nonlinear pricing schedule (p, r_i). One then chooses the actual values of p and r to implement by adjusting them so that the aggregate net trade is zero. The construction is more complex in general formulations, but the basic correspondence principle persists.

It is worth noting that as the number of traders increases, the distribution of types appearing at the auction converges to the theoretical distribution F, provided each falls within one of a finite number of classes i characterized by u^i and F_i. The correct choices of p and r become perfectly predictable and cannot be affected significantly by any one trader. At any limit point, predictability implies that the Ramsey number $R_i(p, 0)$ is nil in the associated nonlinear pricing problem: the linear outlay schedule $O_i(x) = p \cdot x$ solves the pricing problem with $r_i = 0$, which is sufficient to ensure money balance in every contingency. In parallel in the auction design problem, non-manipulability implies that the imputed costs of the incentive constraints shrink and informational rents disappear.[20] Therefore, if the welfare weights are symmetric (to exclude monopoly power), say, $\lambda_i = 1$, then at the limit $\mu = 1$, $R_i = 0$, and each trader's transaction is determined entirely by the ordinary demand condition for his type: $u^i_q(q_i, t_i) = p$. That is, each limit outcome is a Walrasian equilibrium in which the market is cleared by a fixed, uniform price vector p derived from the population average of the traders' demand functions.

A SIMPLE SYMMETRIC MARKET

For this example and the next, we use a symmetric version of the previous model, including $\lambda_i = 1$ and $\mu_i = \mu$, and suppose there is a single commodity and a single type parameter ($\ell = 1$, $K_i = 1$). For this first example, suppose $u(q_i; t_i) = t_i q_i$; further, each trader is endowed with one unit of the good and can sell or buy up to one unit. Thus, $Q_i = [-1, 1]$ and i's virtual utility is

$$\check{u}^i(q_i, -P_i; t_i) = \check{t}_i q_i - P_i \qquad \text{where} \qquad \check{t}_i = t_i - R\beta_i(t_i)\,, \quad R = [\mu - 1]/\mu\,,$$

on others' reported types (so as to leave the participation and incentive constraints unaltered) such that in each contingency the aggregate of these side payments offsets the sum of the outlays derived from the nonlinear pricing problem. See also McAfee[22] and McAfee and McMillan.[20]

[20] Roberts and Postlewaite[46] prove essentially this in great generality.

and therefore, the aggregate virtual utility is $\check{u}(q;t) = \mu \sum_i \check{t}_i q_i$, which is to be maximized subject to the constraint $\sum_i q_i = 0$. The optimal allocation is obtained by choosing a price $p(t)$ that provides a median for the reported distribution of the virtual types $\check{t}_i$ and then allocating $+1$ unit to those traders (the buyers) for whom $\check{t}_i > p(t)$ and -1 unit to those (the sellers) for whom $\check{t}_i < p(t)$.

This example has the property that if the number m of traders is odd, then the incentive constraints are not binding ($R = 0$) and the outcome is efficient. To see this suppose $P_i(t) = p(t)q_i(t)$ and observe that in this case $p(t)$ is precisely the median of the reported types, and the agent at this median does not trade. Consequently, each trader perceives that when truthful reporting would enable him to (say) buy at a favorable price, false reporting could reduce the price only by making his report the median and thereby excluding trade, or below the median and thereby being assigned to sell at a price below his true valuation. When m is even, this simple property does not hold because false reporting could affect the price without eliminating the opportunity to trade. It also fails if traders are designated *a priori* as either sellers or buyers. However, in the following variant we examine a more elaborate version in which the design of the payment rule is critical.

A SYMMETRIC MARKET WITH LINEAR DEMANDS

For this example we assume initially that $Q_i = \Re$ and each trader's utility is $u(q_i; t_i) = t_i q_i - [1/2]q_i^2$. The virtual utilities are, therefore,

$$\check{u}(q_i, -P_i; t_i) = u(q_i; \check{t}_i) - P_i \qquad \text{and} \qquad \check{u}(q;t) = \mu \sum_i u(q_i; \check{t}_i)\,,$$

which yields the optimal allocation

$$q_i(t) = \check{t}_i - p(t) \qquad \text{where} \qquad p(t) = \frac{1}{m}\sum_{i=1}^{m} \check{t}_i\,.$$

First we check that the incentive constraints are binding when the naive linear payment rule is used: $P_i(t) = p(t)q_i(t)$. If they are not binding, then $R = 0$ and $\check{t}_i = t_i$, so i obtains the net profit $\hat{U}_i(t_i; I) = [1/2]\mathcal{E}\left\{[t_i - p(t)]^2 \mid t_i\right\}$. From this it follows that i gains by reporting a false type that differs from his true type in the direction of the difference between the mean $\bar{t}$ of the distribution of types and his true type; e.g., if his type is high, then he expects to be a net buyer and, therefore, prefers to decrease the price by reporting a type lower than his true type. Thus, the incentive constraint is binding if $t_i \neq \bar{t}$.

For a full analysis it is essential to construct the correct virtual types to determine the optimal assignment. This requires the solution of Eq. (8) to obtain $\psi_i^\circ(t_i)$, and thereby $\beta_i(t_i) = \psi_i^\circ(t_i)/f_i(t_i)$, and then the determination of μ and R. In the present case it is clear *a priori* that $U_i(t_i) > 0$ for both extremes $t_i = a$ and $t_i = b$ at the lower and upper ends of the support of the type distribution, since they are almost sure to trade at a favorable price. Consequently, the transversality condition requires that $\psi_i^\circ(t_i) = 0$ for these two types, and then the differential equation (8) implies that

$$\psi_i^\circ(t_i) = \begin{cases} -F_i(t_i) < 0 & \text{if } a < t_i \le a^*, \\ 1 - F_i(t_i) > 0 & \text{if } b^* \le t_i < b, \end{cases} \tag{13}$$

where $a^* \le t_i \le b^*$ is the interval for which $U_i(t_i) = 0$. In particular, on this interval $\rho_i^\circ(t_i) > 0$, and it is necessarily such that the solution to Eqs. (7) and (8) provides a smooth path between a^* and b^*; the continuity of the path determines μ. Note that $U_i(t_i) = 0$ for a type $t_i \in [a^*, b^*]$ near the mean $\bar{t}$ because when, say, $t_i > \bar{t}$, he could sustain a loss by selling at a price $p(t)$ for which $t_i > p(t) > \check{t}_i$.

This example indicates how convergence to an efficient outcome is obtained as the number of traders increases. As m increases and μ declines to 1, the left and right segments of ψ_i° in Eq. (13) remain invariant, but $\rho_i^\circ \equiv \rho_i/[\mu - 1]$ increases. Consequently, the rising segment over the interval $[a^*, b^*]$ becomes steeper, implying that this interval shrinks and eventually vanishes. Thus, the effect of additional traders is two-fold: (1) the reduction of μ brings the virtual types closer to the actual types and thereby sets the price closer to the Walrasian price; and (2) the interval of those types of each trader who obtain zero expected gain from trade shrinks. At the limit each type's expected utility is $[1/2][t_i - \bar{t}]^2$ as calculated previously.

The characterization is essentially unaltered if traders are designated *ex ante* as either sellers ($Q_i = \Re_-$) or buyers ($Q_i = \Re_+$). For the sellers $\psi_i^\circ(t_i) = -F_i(t_i)$, and for the buyers $\psi_i^\circ(t_i) = 1 - F_i(t_i)$; those sellers with $t_i > a^*$ and those buyers with $t_i < b^*$ obtain $U_i(t_i) = 0$. And it is unaltered if the sets Q_i are further modified to reflect trading of discrete items, as in Wilson.[54]

In fact, however, particular cases of this example, like the one before, have the property that an optimal allocation is efficient. That is, $R = 0$ and $\check{t}_i = t_i$ independently of the number m of traders. As we have seen previously, this must be due to a nonlinear payment rule. In the following, we establish this result for a more general setup with ℓ commodities and $K_i = \ell$ type parameters for each trader; also, $Q_i = \Re^\ell$. However, we address only the special case that the types are uniformly distributed.

EFFICIENCY OF THE ALLOCATION

This example admits an explicit solution when the distribution of each trader's type is uniform or normal on a ball. Using the methods illustrated in Wilson[56] for nonlinear pricing, one can solve completely various specifications in which there are ℓ commodities, each trader's type is ℓ-dimensional, and $u_q(q_i, t_i)$ is the vector $t_i - q_i$. We address here only the case that each t_i is uniformly distributed on the ball on which $z(t_i) \le r$, where $z(t_i) \equiv [1/2] \sum_k t_{ik}^2$.

The clue is to anticipate that $U_i(t_i) = w(z(t_i))$ for some univariate function w.[21] If this is so, then the Envelope Theorem's implication that $U_i'(t_i) = \mathcal{E}\left\{q_i(t) \mid t_i\right\}$ can be inserted into the formula for the optimal allocation to obtain

$$\frac{m-1}{m}[t_i - R\beta_i(t_i)] - w'(z(t_i))t_i = \frac{1}{m}\sum_{j\neq i}\kappa_j\,,$$

where $\kappa_j = \mathcal{E}\left\{t_j - R\beta_j(t_j)\right\}$. Differentiation of this relationship, and using Eq. (8), yields the differential equation

$$\frac{m-1}{m}[\ell + R] - 2w''(z)z - \ell w'(z) = 0\,,$$

on the domain for which $\rho_i^\circ(t_i) = 0$, so

$$w'(z) = \frac{m-1}{m}\left[1 + \frac{R}{\ell}\right] + \frac{C_0}{z^{\ell/2}}\,.$$

To determine the constant of integration C_0, we use the transversality condition, which in this case requires $\psi_i^\circ(t_i)\cdot t_i = 0$ when $z(t_i) = r$ on the presumption, as above, that $U(t_i) > 0$ on the boundary. By symmetry around the origin of the conditional distribution of t_j given $z(t_j)$, each $\kappa_j = 0$, so on the boundary the formula for the optimal allocation yields $w'(r) = [m-1]/m$, which determines C_0. In sum, we obtain

$$w'(z) = \frac{m-1}{m}[1 - Ry(z)] \qquad \text{and} \qquad \beta_i(t_i) = y(z)t_i\,,$$

$$\text{where} \quad y(z) \equiv \left[\frac{1}{\ell}\right]\left[\left(\frac{r}{z}\right)^{\ell/2} - 1\right]\,.$$

These formulas apply outside an interior ball of types on which the participation constraint is binding. Its boundary $z(t_i) = z^\circ$ is identified by the condition that w is smooth across the boundary: because $w(z) = 0$ inside, this requires $w'(z) = 0$ or equivalently $y(z) = 1/R$ and $\check{t}_i = 0$ inside $z(t_i) \le z^\circ$. Then $w(z)$ is determined by the boundary condition that $w(z) = 0$ for $z \le z^\circ$, so for $z \ge z^\circ$ (and omitting the special case $\ell = 2$):

$$w(z) = C_1 + \frac{m-1}{m}z\left[1 + R\frac{1+2y(z)}{\ell-2}\right]\,,$$

$$C_1 = -\frac{m-1}{m}\left[\frac{\ell+R}{\ell-2}\right]z^\circ \qquad \text{and} \qquad z^\circ = r\left[\frac{R}{(\ell+R)}\right]^{2/\ell}\,.$$

The last step uses the aggregate form of Eq. (4) to determine R or equivalently μ. Although one can calculate Eq. (4) explicitly, the important feature is that w

[21]This insight is due to Mark Armstrong.[1]

depends on m only via the multiplicative factor $[m-1]/m$, and this is also true of $\mathcal{E}\{u(q_i(t), t_i)\}$ for each trader i. Consequently, the dependence on the number of traders drops out and the root $R = 0$ is always a solution.[22] Thus, the optimal allocation is efficient in each contingency t.

The remarkable aspect of this conclusion is that the allocation is efficient but, except at the limit as $m \to \infty$, it cannot be achieved with the linear payment rule assumed in Walrasian models of markets. From d'Aspremont and Gérard-Varet[2] (Theorem 6), we know that when $R = 0$ a nonlinear payment rule that achieves efficiency is

$$P_i(t) = -\mathcal{E}\left\{\sum_{j\neq i} u_j(q_j(t), t_j) \mid t_i\right\} + \frac{1}{m-1}\sum_{j\neq i}\mathcal{E}\left\{\sum_{k\neq j} u_k(q_k(t), t_k) \mid t_j\right\}.$$

CORRELATED TYPES AND COMMON-VALUE AUCTIONS

The construction above assumes that the traders' types are independent. Even when they are correlated, possibly because they include information about a common-value component of their utilities, the optimal auction assures that the allocation is efficient in every contingency. The construction of the optimal payment scheme in this case includes penalties and rewards for each trader, contingent on others' reported types, that are designed to ensure truthful reporting, even if the trader's expected gain from trade is nil. However slight they might be, correlations among the types can be exploited to ensure truthful reporting by making the penalties and rewards sufficiently large.

This can be seen indirectly in our formulation by noting that the optimal payments $P_i(t)$ minimize $\sum_i P_i \mu_i \overline{\xi}_i(t)$ subject to $\sum_i P_i = 0$, where, if $\xi_i(\theta) \equiv 1 - R_i\phi_i(\theta)\beta_i(t_i)$,

$$\overline{\xi}_i(t) \equiv \mathcal{E}\{\xi_i(\theta) \mid t\} = 1 - R_i\overline{\phi_i}(t)\beta_i(t_i) \qquad \text{and} \qquad \overline{\phi_i}(t) \equiv \mathcal{E}\{\phi_i(\theta) \mid t\}.$$

Note that $\overline{\phi_i}$ depends nontrivially on all the types if they are correlated, even if they are conditionally independent given τ, provided the dependence on τ is nontrivial. Consequently, any finite solution to the minimization problem requires that $\mu_i\overline{\xi}_i(t) = \nu(t)$, where again $\nu(t)$ is a multiplier for the money-balance constraint. This condition can be satisfied in general only if $\mu_i R_i \beta_i = 0$ uniformly.[23] This implies that each $\mu_i = \nu(t)$ and, therefore, $\mu_i = \mu \equiv \max_i \lambda_i > 0$. Further, if $\lambda_i = \mu$, then $R_i = 0$; or if $\lambda_i < \mu$, then $\beta_i = 0$, which implies $\psi_i^\circ = 0$, $\rho_i^\circ = 1$, and $U_i = 0$. Thus, those traders accorded inferior welfare weights obtain none of the expected gains from trade, which are captured entirely by those with the maximal welfare

[22] Using *Mathematica* I have verified for small values of ℓ that this is the only *real* root.

[23] This assumes $m > 1$. Singular exceptions occur, but the ones I have studied for $m = 2$ violate affiliation, and for $m > 2$ the degree of singularity escalates.

weight. For instance, if only one trader has the maximal welfare weight, then only he profits from the auction.[24]

In any case, the commodity allocation is efficient because $\alpha_i(t_i) = R_i\beta_i(t_i) = 0$ for every trader, and therefore, the virtual utility of each allocation agrees with the actual utility: the condition for an optimal allocation is

$$\mathcal{E}\left\{u_q^i(q_i(t);\theta) \mid t\right\} = p(t),$$

where $p(t)$ is a price that ensures feasibility. For instance, in the symmetric common-value case, each $u^i(q_i;\theta) = u(q_i;\tau)$, and the only efficient allocation is the null trade $q_i(t) = 0$, which is analogous to the no-trade theorem of Milgrom and Stokey.[25] Similarly, if we add a seller who has no information and wants simply to dispose of an inelastic supply S at a nonnegative price, then each $q_i(t) = 0$ or S/m as the price $p(t)$ is negative or positive. This is true regardless of the welfare weight accorded the seller, but if the seller's weight is greater than the others, then he obtains all the gains from trade via the payments.

More realistic payment rules can be derived by imposing the participation constraint in its *ex post* form; that is, each trader's expected profit is nonnegative in every contingency t. The formulation can be adapted to this modification by adjoining the *ex post* participation constraint. Then the condition for optimality of the payments requires only that $\mu_i\bar{\xi}_i(t) \leq \nu(t)$, where $\nu(t) = \max_i \mu_i\bar{\xi}_i(t)$ and the difference $\mu_i\zeta_i(t) = \nu(t) - \mu_i\bar{\xi}_i(t) \geq 0$ is the Lagrange multiplier used for the *ex post* participation constraint. The condition for the optimality of the allocation is then

$$\mu_i\mathcal{E}\left\{u_q^i(q_i(t);\theta)[\xi_i(\theta) + \zeta_i(t)] - R_i u_{qt}^i(q_i(t);\theta)\beta_i(t_i) \mid t\right\} = \pi(t),$$

where again the price $\pi(t)$ ensures feasibility of the commodity allocation.

To illustrate the implications of this modification, consider the symmetric common-value model where $u^i(q_i;\theta) = u(q_i;\tau)$ and the types are conditionally independent and identically distributed given τ. For this model the optimality condition for the allocation can be phrased as

$$\nu(t)\mathcal{E}\left\{u_q^i \mid t\right\} - \mu_i R_i \mathrm{Cov}\{u_q^i, \phi(t_i;\tau) \mid t\}\beta_i(t_i) = \pi(t),$$

where $u_q^i \equiv u_q(q_i;\tau)$ and Cov indicates covariance. For example, suppose that each $\lambda_i = 1$, so $\mu_i = \mu$ by symmetry, and $u_q^i = \tau - q_i$. Also, suppose θ has a normal distribution for which the marginal distribution of τ has mean $\bar{\tau}$ and variance σ^2,

[24] This conforms to the results obtained by Cremér and McLean[7] who examine correlation among the bidders' private values, and McAfee, McMillan, and Reny[21] in the case of a common value, for an auction (of discrete items) designed to maximize the seller's expected profit. They show that, in the limit as the bounded domain of payments is enlarged, the seller obtains all the gains from trade in every contingency t. McAfee and Reny[23] construct the payment rules, which are quite complicated. This conclusion depends on utilities being linear in money; Robert[45] shows that it is false if utilities are strictly concave or payments are bounded. It is also false if an *ex post* participation constraint is invoked.

and the conditional distribution of t_i has mean τ and variance 1. This specification implies the conditional mean and variance

$$\mathcal{E}\{\tau \mid t\} = \frac{\overline{\tau}/\sigma^2 + \sum_i t_i}{1/\sigma^2 + m} \qquad \text{and} \qquad \mathrm{Var}\{\tau \mid t\} = \frac{1}{1/\sigma^2 + m}.$$

The optimal allocation is

$$q_i(t) = \left[\frac{\mu R}{\nu(t)}\right] \mathrm{Var}\{\tau \mid t\}[\overline{\beta}(t) - \beta_i(t_i)],$$

where $\overline{\beta}(t) = \sum_i \beta_i(t_i)/m$ and

$$\nu(t) = \mu + \mu R \cdot \max_i \left\{[t_i - \mathcal{E}\{\tau \mid t\}] - \frac{t_i - \overline{\tau}}{\sigma^2 + 1}\right\} \beta_i(t_i),$$
$$\pi(t) = \nu(t)\mathcal{E}\{\tau \mid t\} - \mu R \cdot \mathrm{Var}\{\tau \mid t\}\overline{\beta}(t).$$

Note from the calculation of $\nu(t)$ that typically only one trader is not constrained by the *ex post* participation constraint. From the fact that $R \to 0$ as $m \to \infty$, one derives $\nu(t) \to \mu$, $\pi(t)/\mu \to \tau$, and $q_i(t) \to 0$ almost surely, which is efficient. Alternatively, suppose $\sigma \to \infty$ so that prior information becomes negligible, and the marginal variance of each type also becomes infinite. Then Eq. (8) indicates that β_i becomes constant (and therefore zero), so again the limit is the no-trade allocation. Thus, the complexity of this allocation rule is due mainly to the significant role of prior information compared to the traders' private information.

5. CONCLUSION

The formulation in Section 1 and the analysis in Section 2 unify a literature that encompasses applications ranging from labor contracting to auctions. In fact, these applications are not so disparate as they seem when encumbered by institutional detail. The aim in each case is to construct an efficient allocation subject to participation and incentive constraints, albeit in some cases with welfare weights favoring one party. Typically this involves some collective decision, some risk sharing, etc. The critical constraint in each case is incentive compatibility, which is required to induce truthful revelation of private information—or in an implementation, to ensure that the allocation can be realized by equilibrium strategies. The results conform to those obtained by different methods in studies of efficient trading procedures and, at the other extreme, also to those obtained in studies of contracting where risk sharing and imperfect monitoring of productive inputs are key features. An important advantage of this formulation is that correlation among participants' types is no impediment to the analysis, and the results extend naturally to multidimensional types (although computations remain difficult).

The main disadvantage is heavy reliance on the formulation as a direct revelation game. The results provide clues about the possible forms of implementation as, say, auctions in the case of trading contexts, but they do not provide explicit characterizations of the efficient game form when participants' actions are bids and offers. This is especially limiting when one envisions implementations with important dynamic features, such as a double auction with oral outcry of ask and bid prices and acceptances of offered transactions. In practice, therefore, the results are useful mainly as standards against which to compare the performance of subjects in experimental auctions. In this sense, the prospect of including risk aversion, correlation, and other realistic features is the main advantage over existing results.

ACKNOWLEDGMENTS

Research support provided by NSF grants SES 8908269 and 9207850.

REFERENCES

1. Armstrong, Mark. Private communication, 1992.
2. d'Aspremont, Claude, and Louis-André Gérard-Varet. "Incentives and Incomplete Information." *J. Public Econ.* **11** (1979): 25–45.
3. d'Aspremont, Claude, and Louis-André Gérard-Varet. "Bayesian Incentive Compatible Beliefs." *J. Math. Econ.* **10** (1982): 25–45.
4. Baron, David. "Design of Regulatory Mechanisms and Institutions." In *Handbook of Industrial Organization*, edited by Richard Schmalensee and Robert Willig, Volume II, Chapter 24, 1347–1447. Amsterdam: Elsevier, 1989.
5. Borch, Karl. "Equilibrium in a Reinsurance Market." *Econometrica* **30** (1962): 424–444.
6. Bulow, Jeremy, and D. John Roberts. "The Simple Economics of Optimal Auctions." *J. Pol. Econ.* **97** (1989): 1060–1090.
7. Cremér, Jacques, and R. P. McLean. "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions." *Econometrica* **56** (1988): 1247–1257.
8. Gresik, Thomas A., and Mark A. Satterthwaite. "The Rate at Which a Simple Market Becomes Efficient as the Number of Traders Increases: An Asymptotic Result for Optimal Trading Mechanisms." *J. Econ. Theory* **48** (1989): 304–332.
9. Gresik, Thomas A. "*Ex Ante* Incentive Efficient Trading Mechanisms Without the Private Valuation Restriction." *J. Econ. Theory* **55** (1991): 41–63.
10. Gresik, Thomas A. "The Efficiency of Linear Equilibria of Sealed-Bid Double Auctions." *J. Econ. Theory* **55** (1991): to appear.
11. Gresik, Thomas A. "*Ex Ante* Efficient, *Ex Post* Individually Rational Trade." *J. Econ. Theory* **55** (1991): to appear.
12. Gresik, Thomas A. "Efficient Bilateral Trade with Statistically Dependent Beliefs." *J. Econ. Theory* **55** (1991): to appear.
13. Guesnerie, Roger, and Jean-Jacques Laffont. "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm." *J. Public Econ.* **25** (1984): 329–369.
14. Hart, Oliver, and Bengt Holmström. "The Theory of Contracts." In *Advances in Economic Theory: Fifth World Congress*, edited by Truman Bewley, Chapter 3, 71–156. Cambridge, UK: Cambridge University Press, 1987.
15. Holmström, Bengt, and Roger Myerson. "Efficient and Durable Decision Rules with Incomplete Information." *Econometrica* **51** (1983): 1799–1819.
16. Jackson, Matthew. "Bayesian Implementation." *Econometrica* **59** (1991): 461–478.
17. Jewitt, Ian. "Justifying the First-Order Approach to Principal-Agent Problems." *Econometrica* **56** (1988): 1177–1190.
18. Kennan, John, and Robert Wilson. "Bargaining with Private Information." *J. Econ. Lit.* **30** (1992): to appear.

19. Laffont, Jean-Jacques, and Jean Tirole. "Using Cost Observation to Regulate Firms." *J. Pol. Econ.* **94** (1986): 614–641.
20. McAfee, R. Preston, and John McMillan. "Multidimensional Incentive Compatibility and Mechanism Design." *J. Econ. Theory* **46** (1988): 335–354.
21. McAfee, R. Preston, John McMillan, and Philip Reny. "Extracting the Surplus in the Common Value Auction." *Econometrica* **57** (1989): 1451–1459.
22. McAfee, R. Preston. "Efficient Allocation with Continuous Quantities." *J. Econ. Theory* **53** (1991): 51–74.
23. McAfee, R. Preston, and Philip J. Reny. "Correlated Information and Mechanism Design." *Econometrica* **60** (1992): 395–421.
24. Milgrom, Paul R. "A Convergence Theorem for Competitive Bidding with Differential Information." *Econometrica* **47** (1979): 679–88.
25. Milgrom, Paul, and Nancy Stokey. "Information, Trade, and Common Knowledge." *J. Econ. Theory* **26** (1982): 17–27.
26. Milgrom, Paul, and Robert Weber. "A Theory of Auctions and Competitive Bidding." *Econometrica* **50** (1982): 1089–1122.
27. Milgrom, Paul, and Christina Shannon. "Monotone Comparative Statics." Report 11, Stanford Institute of Theoretical Economics, 1991.
28. Mirrlees, James. "Optimal Tax Theory: A Synthesis." *J. Public Econ.* **6** (1976): 327–358.
29. Mirrlees, James. "The Optimal Structure of Incentives and Authority Within an Organization." *Bell J. Econ.* **7** (1976): 105–131.
30. Mookerjee, Dilip, and Stefan Reichelstein. "Implementation via Augmented Revelation Mechanisms." *Rev. Econ. Studies* **57** (1990): 453–476.
31. Moore, John, and Rafael Repullo. "Subgame-Perfect Implementation." *Econometrica* **56** (1988): 1191–1220.
32. Moore, John, and Rafael Repullo. "Nash Implementation: A Full Characterization." *Econometrica* **58** (1990): 1083–1100.
33. Mossin, Jan. *Theory of Financial Markets.* Englewood Cliffs, NJ: Prentice-Hall, 1973.
34. Mussa, Michael, and Sherwin Rosen. "Monopoly and Product Quality." *J. Econ. Theory* **18** (1978): 301–317.
35. Myerson, Roger. "Incentive Compatibility and the Bargaining Problem." *Econometrica* **47** (1979): 61–73.
36. Myerson, Roger. "Optimal Auction Design." *Math. Op. Res.* **6** (1981): 58–63.
37. Myerson, Roger. "Mechanism Design by an Informed Principal." *Econometrica* **51** (1983): 1767–1798.
38. Myerson, Roger. "Two-Person Bargaining Problems with Incomplete Information." *Econometrica* **52** (1984): 461–488.
39. Myerson, Roger. "Multistage Games with Communication." *Econometrica* **54** (1986): 323–358.
40. Myerson, Roger. "Mechanism Design." In *Allocation, Information, and Markets,* edited by J. Eatwell, M. Milgate, and P. Newman. London: Macmillan, 1991.

41. Myerson, Roger. *Game Theory*, Chapter 10. Cambridge, MA: Harvard University Press, 1991.
42. Myerson, Roger, and Mark Satterthwaite. "Efficient Mechanisms for Bilateral Trade." *J. Econ. Theory* **29** (1983): 265–281.
43. Palfrey, Thomas R. "Implementation in Bayesian Equilibrium: The Multiple Equilibrium Problem in Mechanism Design." In *Advances in Economic Theory: Invited Papers for the Sixth World Congress of the Econometric Society*, edited by Jean-Jacques Laffont. Cambridge: Cambridge University Press, 1992.
44. Postlewaite, Andrew, and David Schmeidler. "Implementation in Differential Information Economies." *J. Econ. Theory* **39** (1986): 14–33.
45. Robert, Jacques. "Continuity in Auction Design." *J. Econ. Theory* **55** (1991): 169–171.
46. Roberts, D. John, and Andrew Postlewaite. "The Incentive for Price Taking Behavior in Large Exchange Economies." *Econometrica* **44** (1976): 115–28.
47. Rochet, Jean-Charles. "The Taxation Principle and Multi-Time Hamilton-Jacobi Equations." *J. Math. Econ.* **14** (1985): 113–128.
48. Rogerson, William. "The First-Order Approach to Principal-Agent Problems." *Econometrica* **53** (1985): 1357–1368.
49. Wilson, Robert. "The Theory of Syndicates." *Econometrica* **36** (1968): 119–132.
50. Wilson, Robert. "A Bidding Model of 'Perfect' Competition." *Rev. Econ. Studies* **4** (1977): 511–518.
51. Wilson, Robert. "Information, Efficiency, and the Core of an Economy." *Econometrica* **46** (1978): 807–816.
52. Wilson, Robert. "Auctions of Shares." *Quart. J. Econ.* **93** (1979): 675–689.
53. Wilson, Robert. "A Note on Revelation of Information for Joint Production." *Social Choice and Welfare* **1** (1984): 69–73.
54. Wilson, Robert. "Incentive Efficiency of Double Auctions." *Econometrica* **53** (1985): 1101–1116.
55. Wilson, Robert. "Strategic Analysis of Auctions." In *Handbook of Game Theory*, edited by R. Aumann and S. Hart, Volume 1, Chapter 8. Amsterdam: North-Holland, 1992.
56. Wilson, Robert. *Nonlinear Pricing*. New York: Oxford University Press, 1992.

III. Evidence

John Rust,* John H. Miller,† and Richard Palmer‡
*University of Wisconsin, †Carnegie-Mellon University, and ‡Duke University *†‡ Santa Fe Institute; e-mail: dat@sfi.santafe.edu

SIX
Behavior of Trading Automata in a Computerized Double Auction Market

This paper reports the results of a series of tournaments held at the Santa Fe Institute beginning in March, 1990 in which computer programs played the roles of buyers and sellers in a synchronized double auction market. We show that despite the decentralized nature of the trading process and traders' incomplete information about supply and demand, transaction-price trajectories for a heterogeneous collection of computer programs typically converged to the competitive equilibrium, resulting in allocations that were nearly 100% efficient. We also show that a very simple trading strategy is a highly effective and robust performer in these markets. A simple rule-of-thumb was able to outperform more complex algorithms that used statistically based predictions of future transaction prices, explicit optimizing principles, and sophisticated "learning algorithms."

The Double Auction Market, Eds. D. Friedman and J. Rust,
SFI Studies in the Sciences of Complexity,
Proc. Vol. XIV, Addison-Wesley, 1993

1. INTRODUCTION

This paper reports the results of a series of computerized double auction tournaments held at the Santa Fe Institute beginning in March 1990. The tournament consisted of over 30 computer programs (*automata traders*) playing the roles of buyers and sellers in a simplified synchronized double auction (DA) market. The tournament was organized with several objectives in mind: (1) to get new insights on the form of effective trading strategies, (2) to compare the performance of automata traders and human traders, and (3) to create an artificial market to help us better understand the operation of the "invisible hand" in real-world DA markets.

The remarkable efficiency properties of DA markets have been documented in numerous laboratory experiments using human subjects. By assigning subjects *tokens* with fixed *redemption values* and *token costs*, well-defined supply and demand curves can be constructed. The intersection of these curves defines the price and quantity at which neoclassical economic theory predicts trading will occur, the *competitive equilibrium* (CE) solution. The complication is that in most experimental markets each trader only knows their own token values: no single trader has enough information to determine the market supply and demand curves in order to compute the CE. The nearly universal finding of more than three decades of human experiments is that despite the presence of incomplete information and the small number of traders, transaction prices and quantities quickly converge to CE. The resulting market allocations are highly efficient: traders are typically able to exploit close to 100% of the potential profits.

Although the textbook "supply equals demand" model may provide a good prediction of *closing* prices and quantities in DA markets, it fails to explain the dynamics by which this happens. A more sophisticated theory is required to show how the trading process aggregates traders' dispersed information, driving the market towards CE. The essence of the problem was clearly stated by Friederik Hayek nearly 50 years ago:

> "The problem is in no way solved if we can show that all the facts, *if* they were known to a single mind, would uniquely determine the solution; instead we must show how a solution is produced by the interactions of people each of whom possesses only partial knowledge. To assume that all the knowledge to be given to a single mind in the same manner in which we assume it to be given to us as the explaining economists is to assume the problem away and to disregard everything that is important and significant in the real world." (Hayek,[16] p. 530)

Since the use of a computer tournament to gain insights into human trading behavior is somewhat unorthodox, Section 2 briefly reviews current theories of DA markets. Although these theories have provided important insights into the nature of trading strategies and price formation, it is fair to say that none of them has provided a satisfactory resolution of "Hayek's problem." In particular, current theories assume a substantial degree of implicit coordination by requiring that traders have common knowledge of each other's strategies (in game-theoretic models), or by assuming that all traders use the same strategy (in learning models). Little is known theoretically about price formation in DA markets populated by heterogeneous traders with limited knowledge of their opponents. Although experimental studies have provided considerable empirical evidence on the nature of trading *behavior* under these conditions, they have failed to cast light on trading *strategies* which are essentially unobservable.

In order to observe strategies directly, we sponsored a tournament in which entrants submitted trading programs playing the roles of buyers and sellers in a computerized DA market. To attract good programs we offered $10,000 in prizes, paid out in proportion to profits earned by entrants' programs over the course of the tournament. In return, we obtained a heterogeneous collection of trading programs to populate a unique laboratory for studying decentralized price formation. Section 3 describes the rules of the tournament and the structure of our "synchronized DA," a modified version of the traditional continuous DA market designed to simplify the task of programming strategies and guarantee equal trading opportunities. Section 4 presents the results of the cash tournament held at the Santa Fe Institute in March 1990 and subsequent non-cash "scientific" and "evolutionary" tournaments held in 1991. We find that the top-ranked programs yield a fairly "realistic" working model of a DA market in the sense that their collective behavior is consistent with the key "stylized facts" of human experiments. We also find that a very simple strategy is a highly effective and robust performer in these markets. This strategy was able to outperform more complex algorithms that use statistically based predictions of future transaction prices, explicit optimizing principles, or sophisticated "learning algorithms." The basic idea behind the approach can be described quite simply: *wait in the background and let others do the negotiating, but when bid and ask get sufficiently close, jump in and "steal the deal."* However, the results of our evolutionary tournaments show that when too many other traders try to imitate this strategy, market efficiency can fall precipitously due to negative information externalities. Specifically, if too many traders "wait in the background," little information is generated until just before the end of the trading period. This tends to produce "closing panics" as traders rush to unload their tokens in the final seconds of the trading period, resulting in failure to execute all potentially profitable transactions.

Long-run stability in the trading environment seems to require the presence of active bidders that provide a flow of information to "lubricate the market." However, most of the active bidding strategies seem to be too impatient, exposing themselves to a high risk of mistakes and consequent exploitation by the background traders. Although a few of the more complex trading programs appear to be resistant to short-run exploitation, none of them appear strong enough to resist the parasitic effects of the background traders in the long run. We show that a market dominated by background traders can be "quasi-stable" if a small but steady stream of short-lived "noise traders" enters the market. The noise traders provide a flow of information and source of new capital to keep the market running despite being nearly totally dominated by background traders. However, since price volatility is very high in such a market, it is likely to present an attractive opportunity for exploitation by new strategies. Section 5 concludes with some observations on how one might find such strategies.

2. REVIEW OF PREVIOUS APPROACHES TO ANALYZING DA MARKETS

Modern economic theory has attempted to explain the apparent disequilibrium behavior in DA markets as actually being the equilibrium outcome of a game of incomplete information. The "maintained hypothesis" is that observed trading behavior is a realization of a Bayesian Nash equilibrium (BNE) of this game. Given the immensity of the strategy space (especially in continuous-time formulations), it has proven extremely difficult to characterize the equilibria of these games. The clearest characterizations have been obtained by Satterthwaite and Williams,[31,32] chapter 4, for a class of static DA games known as the *k-double auction*. They have established that equilibrium bidding stategies in the k-DA converge to *truthtelling* as the number of traders gets large, which implies that prices and quantities converge to CE. However in a continuous DA it is easy to see that truthful revelation is a very poor strategy: if a buyer places a bid equal to their true redemption value and another trader accepts that bid, the bidder will clearly earn zero profit.[1]

To the best of our knowledge, the only characterization of equilibrium in a dynamic DA market is due to Wilson.[41] His equilibrium, described as a "waiting game Dutch auction" (WGDA) by Cason and Friedman (chapter 8),

> "... offers a concrete explanation of the mechanism by which the dispersed information about traders' valuations is manifested in the prices at which transactions are consummated. The mechanism, according to the present hypothesis, is multilateral sequential bargaining in which the traders are endogenously matched for transactions via a signalling process using delay as the primary signal." (p. 412)

Cason and Friedman[6] have shown that beyond the prediction of high *ex post* trading efficiency, many of the other predictions of Wilson's model are inconsistent

[1]The static k-DA market is also known as a uniform-price, sealed-bid auction. The dynamic DA market is typically identified with continuous real-time trading according to the rules of standard experimental DA markets. For definitions and discussions of static vs. dynamic DA, see Friedman, chapter 1 of this volume.

with the behavior of human traders in laboratory experiments. Wilson's model predicts that trade will occur in the efficient order,[2] whereas in human experiments the rank correlation between the order of transactions with the efficient order is typically much less than 100%. Indeed, Wilson's model predicts that in equilibrium all *ex post* efficiency losses will be due to unrealized intra-marginal trades, i.e., the market may trade too few tokens but never too many. However, in experimental settings, a significant fraction of efficiency losses are due to extra-marginal trades. This is due to the fact that in human experiments, buyers and sellers are not matched for transactions as predicted by Wilson's waiting-game equilibrium: it is frequently the case that extra-marginal traders succeed in "bumping" intra-marginal traders.[3] Furthermore, bidding behavior seems to be poorly described as a sequence of Dutch auctions called exclusively by the current bidder or asker. In human experiments there is often stiff competition for the "right" to hold the current bid or ask, which is frequently "stolen" by other more eager traders. Finally, Cason and Friedman show that in human experiments transaction price changes are significantly negatively autocorrelated. Wilson's model predicts zero autocorrelation in price changes since equilibrium transaction prices must follow a martingale to preclude intertemporal arbitrage.

Given enough freedom in the specification of traders' beliefs and risk aversion, Ledyard[20] has shown that essentially any set of undominated strategy profiles can be "rationalized" as a BNE outcome of the DA trading game.[4] Thus, it may be possible to construct game-theoretic models wherein differential risk aversion provides an "explanation" for extra-marginal efficiency losses observed in human experiments. However, although we know that models can be "rigged" to match any set of stylized facts, there is no guarantee that they will be theoretically plausible. Perhaps the least plausible element of any game-theoretic model is the assumption that players have common knowledge of each other's beliefs and strategies. This presumes an unreasonably high degree of implicit coordination amongst the traders, begging Hayek's question of how coordination is achieved in a decentralized market in the first place. Game theory also assumes that there is no *a priori* bound on traders' ability to compute their BNE strategies. However, even traders with infinite, costless computing capacities may still decide to deviate from their BNE strategies if they believe that limitations of other traders force them to use sub-optimal strategies. Since traders can only observe outcomes, they will never be

[2] The efficient order is the trade sequence which maximizes surplus, i.e., the first trade occurs between the buyer with the highest redemption value and the seller with the lowest token cost, the second trade occurs between the buyer and seller with the next most valuable tokens, and so on.

[3] This suggests that differences in traders' "impatience" may also be a function of other factors we might call "aggressiveness" (or "stupidity"?) which may have no direct relation to the magnitude of their token values.

[4] Easley and Ledyard[9] provide an argument in footnote 13 & p. 76 of their paper that this result will hold even if traders are assumed to be risk neutral.

certain of exactly which strategies their opponents are using. This learning problem is of such a high dimensionality relative to the limited number of observations available within typical trading periods that it may not pay to try to adopt a sophisticated Bayesian updating strategy.[5] Game-theoretic solutions can also be "non-robust" in the sense that equilbrium solutions depend critically on seemingly inessential details of the trading rules and the common-knowledge assumptions about the form of the probability distribution of traders' token values.[6] However, behavior of human subjects does not appear to be dramatically affected by minor changes in DA trading rules or the presence or absence of common knowledge about the distribution from which tokens are drawn.[7] In a game-theoretic model it would be impossible to even define the concept of equilibrium without such prior information.

In response to the difficulties of using game theory to analyze and explain experimental findings in dynamic DA markets, economists have begun to formulate explicit disequilibrium trading theories based on simple yet plausible rules-of-thumb. Examples of this approach include Easley and Ledyard,[9] Friedman,[11] and Garcia.[13] The results of these studies suggest that rationality is not a necessary condition for observing efficient outcomes and convergence to CE in DA markets. Gode and Sunder[15] provided a particularly striking demonstration of this result. They showed that markets populated by "zero-intelligence" (ZI) strategies exhibit very high *ex post* efficiencies, and the corresponding price trajectories frequently

[5] It has been shown[8,10] that Bayesian updating can be inconsistent in infinite-dimensional parameter spaces. In certain cases the prior distribution can completely overwhelm the data in the sense that the posterior will not converge to the "truth" even given an infinite number of observations. Intuitively, this will also be the case when the dimensionality of the object being learned is large relative to the number of observations.

[6] For example, the seemingly innocuous change in the clearing rules that transform the 1-DA to the MDA described in Satterthwaite and Williams[33] yields an entirely different set of equilibria, which are much more difficult to analyze. In particular they show that trader's strategies may change substantially (for example, truthtelling is no longer a dominant strategy for sellers: in the MDA equilibrium strategies may involve asking less than their token costs) and the resulting game may have no pure strategy equilibria.

[7] For example, Kagel and Vogt's experimental results[19] do not reveal a significant difference in behavior between the 1-DA and the MDA. In extreme cases, however, prior information can affect DA outcomes but not always in the way one would expect. For example, in experiments using the "swastika" configuration for supply and demand curves, Smith[36,37] has shown that common knowledge of the realized values of the tokens (i.e., complete information about supply and demand) actually *retards* convergence to CE. Apparently the existence of complete information allows some subjects to try to achieve allocations that are better than their CE profit allocations, leading to conflicts that inhibit convergence and reduces efficiency, a result that may also be due to the fact that "when agents know each other's payoffs, it provides scope for interpersonal utility comparisons which impinge on behavior" (Smith,[37] pp. 357–358). Thus, common knowledge seems to be neither necessary nor sufficient for attaining CE outcomes.

converged to CE.[8] These findings strongly suggest that the nice properties of DA markets may have more to do with the properties of the institution itself than the rationality of traders *per se.*[9]

Although ZI traders are collectively rational, they are individually irrational and, therefore, unlikely to provide a good model of human trading behavior. Specifically, we show that ZI traders, like *truthtellers*, will be rapidly exterminated by even slightly more sophisticated trading strategies. Their collective behavior in successive trading periods is necessarily *IID,* whereas humans exhibit strong inter-period learning effects. In addition, even though price trajectories frequently converge to equilibrium, overall price volatility in DA markets populated by ZI traders is unrealistically high.[10] Similar results hold for markets populated by truthtellers. Truthtelling can be viewed as a limiting form of the ZI strategy when the set of prices over which it randomizes collapses to the true token value. Since trade in a synchronized DA market populated entirely by truthtellers is necessarily fully efficient, it follows that ZI traders converge to 100% efficiency as the set of prices over which they randomize converge to a unit mass at their true token values. The fact that ZI traders are so efficient, even when large random deviations are allowed, is somewhat surprising. This result clearly depends on the ability to "recontract" in dynamic DA markets. Gode and Sunder[15] show that in the best case with only one intra-marginal buyer and seller, roughly only 50% efficiency is achieved in the first step of a synchronized DA.[11]

However, strategies that are more individually rational than ZI may display less collective rationality. One of the lessons from Axelrod's[3] prisoner's dilemma tournament is that sometimes players can be "too smart for their own good" in the sense that sophisticated and self-interested behavior can be deterimental to achieving good cooperative outcomes. The question is whether this is true in DA markets as well: is it the case that sophisticated optimizing strategies make individual traders better off, but reduce market efficiency? This seems to be true in DA markets as well: clever strategies can exploit unsophisticated "nice" strategies

[8] A ZI seller with token cost C asks an amount $C + \tilde{U}$, where $\tilde{U}$ is uniformly distributed over its support S. Similarly a *ZI* buyer with redemption value R bids amount $R - \tilde{U}$. At each step t of trading, a ZI trader uses *IID* draws $\tilde{U}$ to construct its bids and asks, accepting the first profitable opposing bid or ask that comes along. Thus, ZI traders are "minimally rational" in the sense that they do not attempt to optimize or learn from past observations, although they do avoid trading at a loss by always bidding below their redemption values or asking above their token costs.

[9] We should also mention related work by Marimon, McGratten, and Sargent[21] who studied the behavior of a collection of Holland's[18] *classifier systems* in a dynamic exchange economy. Although their economy is a more complicated dynamic market than the essentially static DA market studied here (in particular, it can have multiple isolated equilibria), they also found that their artificial agents eventually converged to an equilibrium of the system.

[10] There are aspects of the behavior of ZI traders that are consistent with human behavior, such as significantly negatively autocorrelated transaction price changes and low correlations between the actual order of trade and the efficient order. For details, see Cason and Friedman.[6]

[11] This suggests that ZI traders will create significant inefficiencies in one-shot sealed-bid DA's such as the k-DA. This is confirmed by simulations in Kagel and Vogt[19] which show that in the k-DA, ZI traders acheive efficiencies in the range of only 30–50%.

such as *truth teller* or ZI, but if everyone uses these strategies market efficiency falls. We suspect that an analog of the Satterthwaite-Williams result carries over to dynamic DA markets: namely, that the gains from behaving strategically are negligible even in markets with very small numbers of traders. However, in order to define the gains to behaving strategically, we also have to define an appropriate notion of what it means to behave passively or "non-strategically." Identifying non-strategic behavior with truthtelling will not work in dynamic DA markets since truthtelling, like ZI, is easily exploitable. In a large and efficient market, there is reasonably well-defined notion of *pricetaking* behavior: namely, placing all bids or asks at the market price and only accepting a bid or offer that is at least as good as the market price. In an efficient market it seems intuitively clear that pricetaking should be close to a dominant strategy.[12] However, it is less clear exactly what it means to be a pricetaker in thin markets with small numbers of traders where initial transaction prices are highly volatile and potentially far from equilibrium, a situation that is typical of the first few periods of most experimental DA markets as well as many markets observed in the field.

3. STRUCTURE OF THE DA TOURNAMENT

Most of the trading programs used in this study were submitted in response to advertisements for a "Double Auction Tournament" held at the Santa Fe Institute in March 1990. Cash prizes totalling $10,000 were offered to a maximum of 100 entrants in proportion to the trading profits earned by their programs over the course of the tournament. In addition to prize money and wide publicity, a substantial effort was devoted to make the programming and debugging of trading strategies as easy as possible. This included development of the *Santa Fe Token Exchange* (SFTE) which opens at the start of each hour for token trading over the worldwide Internet computer network.[13]

The computerized DA was implemented via a *message-passing protocol* which specifies the form of allowable messages that programs can send, such as bids, asks, and buy and sell orders. Entrants were provided with a simple "skeleton" trading

[12] Roberts and Postlewaite[28] were the first to formally establish such a result in the context of a static complete-information Walrasian exchange economy. The results of Satterthwaite and Williams[31,32] and Rustichini, Satterthwaite, and Williams[30] cited in Section 2 can be interpreted as an extension of the Roberts-Postlewaite result to a non-Walrasian exchange economy with incomplete information.

[13] Many entrants reported that the SFTE was useful for refining their trading programs in advance of the actual tournament. We also distributed "free-ware" to allow entrants who did not have Internet access to set up their own local token exchanges.

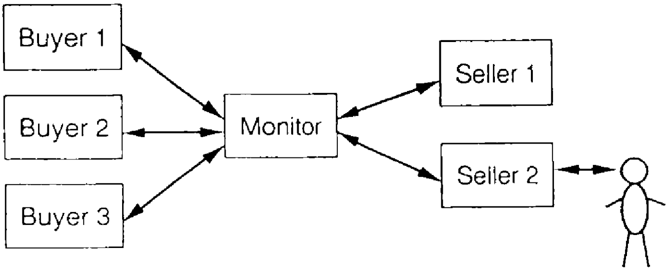


FIGURE 1 Interplayer communication via the monitor.

program (written in C, Fortran, or Pascal) which handled all the message-passing housekeeping, allowing them to focus on the logic of their strategies, rather than on programming details. A central *monitor program* coordinates the trading process by communicating with all of the trading programs, executing their buy and sell orders and relaying their bids and asks to the other traders. Trading programs (which could also be interfaces to human traders) communicate only with the monitor and not directly with each other as illustrated in Figure 1. It is important to note that the monitor programis only a clerk: it is not an "auctioneer" and has no market-clearing authority.[14]

The structure of our computerized DA market is very similar to the continuous-time experimental DA markets described in Section 2. The major differences are (1) time is discretized into alternating *bid/ask* (BA) and *buy/sell* (BS) steps, and (2) transactions are cleared according to AURORA rules described below. The DA market opens with a BA step in which all traders are allowed to simultaneously post bids and asks. After the monitor informs the traders of each others' bids and asks, the holders of the *current bid* (highest outstanding bid) and *current ask* (lowest outstanding ask) enter into a BS step.[15] During the BS step, either player can accept the other player's bid or ask. If an acceptance occurs, a transaction is executed.[16] A *trading period* is simply a set of S alternating BA and BS steps.

The discretization of time was adopted to simplify the programming of trading strategies and improve the synchronization of communications between players and the monitor in a multiprocessing or network-computing environment where delays may vary from player to player and moment to moment. In a continuous-time environment, "faster" traders have an inherent advantage. This speed advantage may arise due to communication delays (e.g., simultaneous messages sent from a trader in Japan and Chicago may arrive at different times at a central computer in

[14]The monitor does enforce trading rules, and can impose upper- and lower-price limits. It also has the authority to censor illegal or late messages, although no cpu time limits were imposed in the actual tournament.

[15]If a current bid (ask) does not exist, then all buyers (sellers) enter into the BS step.

[16]If both parties accept each other's offers, the monitor randomly chooses between the current bid and ask to determine the transaction price.

New York) or due to processing delays (e.g., machines may be able to recognize and respond to certain conditions faster than humans). By discretizing time and setting sufficiently wide response-time limits, we can effectively guarantee that all traders have equal trading opportunities. In the limit, our implementation of a discrete-time DA market is not restrictive since a continuous-time trading environment can be arbitrarily well approximated by a discrete-time environment with very many short trading intervals.[17]

The AURORA rules were inspired by similar rules used by the AURORA computerized trading system developed by the Chicago Board of Trade. AURORA rules stipulate that only the holder of the current bid or current ask are allowed to trade. We adopted these rules as a substitute for ad hoc tie-breaking rules which are necessary in discrete-time trading environment when several traders are able to simultaneously accept an outstanding bid or ask.[18] In early human experiments using random tie breaking rather than AURORA rules, we found that traders often expressed frustration that trade execution seemed more a matter of luck than strategy due to the fact that other traders would repeatedly win random tie breaks for the acceptance of an outstanding bid or ask. On the other hand, experimentalists have criticized the AURORA rules on the grounds that it makes it harder for traders to remain in the background since it forces them to "show their hand" by posting a winning bid or ask before being allowed to trade. We have found, however, that these rules do not place a significant constraint on background traders: they can still stay quietly in the background for most of the trading period, jumping in the moment they detect an attractive bid or ask. Indeed, this is precisely the strategy followed by the winner of the tournament.

An individual DA *game* is divided into one or more *rounds*, and each round is further divided into one or more *periods*. A single period of the DA game consists of a fixed number of alternating BA and BS steps as described above. The reason for structuring games to have multiple rounds and periods within rounds is to control players' abilities to learn about their opponents. Tokens and redemption values are fixed within each period of a given round, but are allowed to change between rounds. Thus DA games with many periods allow players to learn the value of each others' *tokens*, while DA games with many rounds allow players to learn about each others' *strategies*.

At the start of a DA game, the monitor broadcasts *public information* to the traders, including the number of buyers and sellers and their identities, the number of rounds, periods, and time steps, the number of tokens each agent will have, and the joint distribution F from which the traders' token values are drawn. Next, the monitor sends each trader a packet of *private information*, namely, their realized

[17]The same point applies to price units in our DA market, which were rounded to the nearest integer.

[18]Presumably, the Chicago Board of Trade had a very different motivation for considering the AURORA rules. Since they have the effect of making all transactions publicly observable, they may have been designed partly in response to trading abuses that were uncovered in the Chicago Exchanges in the late 1980s.

token values. Since public information is provided by a simultaneous broadcast to all players, it serves as a means of ensuring that players have common knowledge about all relevant game parameters. The joint distribution F was communicated to players using a four-digit `gametype` variable. Token values are represented by T_{jk} where j indexes the trader, and k indexes the token assigned to the trader. Tokens are randomly generated according to

$$T_{jk} = \begin{cases} A + B + C_k + D_{jk}, & \text{if } j \text{ is a } \textit{buyer}; \\ A \qquad\; + C_k + D_{jk}, & \text{if } j \text{ is a } \textit{seller}, \end{cases} \tag{3.1}$$

where[19] $A \sim U[0, R_1]$, $B \sim U[0, R_2]$, $C_k \sim U[0, R_3]$, and $D_{jk} \sim U[0, R_4]$. Notice that when $R_1 = R_2 = R_3 = 0$, we have the standard independent private-values model where tokens are independently uniformly distributed on the interval $[0, R_4]$. A `gametype` equal to 0 indicates an environment where redemption values were generated by an unspecified process.

The best way to understand what goes on in a DA market is to study the monitor output for the sample tournament game in Figure 2. The figure shows the first period of a DA game with two rounds and three periods per round. In this case there are four buyers and four sellers, and each trader is assigned four tokens. The implied supply and demand curves yield a unique CE price of 691 at a quantity of 11.[20] The "+" next to each trader's token indicates an intra-marginal token, a "−" indicates an extra-marginal token, and an "=" denotes a token value equal to the CE price. The first BA step yields a current bid of 435 held by B1 (buyer 1) and a current ask of 1128 held by S2 (seller 2). Neither B1 nor S2 chose to accept the other's bid or ask at BS step 1, so the bidding begins again at BA step 2. The first transaction occurs in BS step 4 when both B3 and S1 simultaneously accept each other's offers. Here, a random tie-break results in S3 selling its first token (denoted by capital A) at B3's bid of 717. Immediately after the transaction the current bid and ask are set to zero and a new BA step starts up in period 5. The game continues this way until the final BS step is reached in step 25. The box at the end of the monitor output provides a summary of the period's trading activity: there were ten transactions yielding a total profit of 604, which is 91% of the total surplus of 663. In this case, the source of the inefficiency was due to two events. First, one intra-marginal trade was not consummated. Second, there were two extra-marginal trades made by B3 which displaced an equal number of

[19]Each of the four digits of the **gametype** variable correspond to $\{R_1, ..., R_4\}$ according to the base-3 coding, $R_i = 3^{k(i)} - 1$ where $k(i)$ is the ith digit of **gametype**.

[20]The discrete nature of the market often implies that a nonunique equilibrium point emerges, in which case we have a range of equilibrium prices or quantities.

```
DA game 1  Fri May 24 02:14:05 1991
------------------------------------
protocol:         5  monitor:      443  gametype:     6453
nrounds:          2  nperiods:       3  ntimes:         25
minprice:         1  maxprice:    2000  ntokens:         4
ran1:           728  ran2:          80  ran3:          242
ran4:            26  deadsteps:    100  timeout:        30
id  name             id  name
--  ------------     --  ----------
B1  silverbuffalo    S1  burchard
B2  staecker         S2  pricetaker
B3  perry            S3  breton
B4  anon2            S4  anderson
+----------------------------------------------------------------------------+
|                             Round  1, period 1                             |
+-------+-------------------------+-------------------------+----------------+
| token |  B1    B2    B3    B4   |  S1    S2    S3    S4   |  Equilibrium   |
+-------+-------------------------+-------------------------+----------------+
|   a   | 754+  760+  761+  751+  | 651+  666+  646+  661+  |  691 to 691    |
|   b   | 722+  708+  717+  719+  | 661+  675+  659+  665+  |  av:  691.0    |
|   c   | 691=  705+  690-  702+  | 680+  683+  680+  693-  |  trades: 11    |
|   d   | 681-  678-  689-  691=  | 779-  776-  788-  774-  |                |
+-------+-------------------------+-------------------------+----------------+
|t  step|  B1    B2    B3    B4   |  S1    S2    S3    S4   | cbid coff price|
+-------+-------------------------+-------------------------+----------------+
|1  BA  | 435*  276   368   345   | 1550  1128* 1182  1999  |  435 1128      |
|     BS|                         |                         |  435 1128      |
|2  BA  |       440   574*        | 1127  1090   966* 1065  |  574  966      |
|     BS|                         |                         |  574  966      |
|3  BA  |  @    579   676*        |  803*  942   827   930  |  676  803      |
|     BS|                         |                         |  676  803      |
|4  BA  |  @    681   717*        |  652$  793   770   791  |  717  652 !    |
|     BS|             a>A         |  a>A                    |            717 |
|5  BA  |       216   606*  345   | 1075   713*  957  1999  |  606  713      |
|     BS|                         |                         |  606  713      |
|6  BA  |  @    611   635*        |  661*  712    X    703  |  635  661      |
|     BS|                         |                         |  635  661      |
|7  BA  |  @    640   642*        |              656*       |  642  656      |
|     BS|             b<B         |               <A        |            656 |
 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
|23 BA  |       557         651*  |        778   681   666* |  651  666      |
|     BS|                   c<C   |                     <B  |            666 |
|24 BA  |       232         550*  |  809         681* 1999  |  550  681      |
|     BS|                         |                         |  550  681      |
|25 BA  |       596         651*  |               "         |  651  681      |
|     BS|                         |                         |  651  681      |
+-------+-------------------------+-------------------------+-----Market-----+
|Trades |  1/2   2/3   4/2   3/3  |  3/3   3/3   2/3   2/2  |     20/21      |
|Profit |  27    48    125   117  |  141   62    45    39   |      604       |
|Eqlbrm |  94    100   96    99   |  81    49    88    56   |      663       |
|Effncy |  29%   48%  130%  118%  | 174%  127%   51%   70%  |      91%       |
+-------+-------------------------+-------------------------+----------------+
```

FIGURE 2 Sample monitor output.

intra-marginal trades. This example illustrates a typical feature of efficiency losses in dynamic DA markets, namely the coexistence of extra-marginal trades along with unconsummated intra-marginal trades.[21]

Tournament entrants were told that their programs would be placed in an unspecified number of alternative *environments*. Each environment is a complete specification of all relevant parameters of the DA game listed in Table 1. Participants were told potential ranges for each of the parameters, but were not given any specific advance information about how the actual environments would be selected. The actual DA tournament consisted of playing a large number of DA games in ten separate environments presented in Table 1. Each of the ten environments were allocated \$1,000 prize money, and separate conversion factors were calculated to translate token profits into dollar earnings. The conversion factor $c(i)$ for environment i is the ratio $1000/TS(i)$, where $TS(i)$ is the total surplus available in environment i. Due to the lack of 100% efficiency, actual dollar payments in the tournament amounted to \$8,937. Overall, we ran a total of 2,233 games in the ten separate environments, comprising 13,398 individual periods of play.

One can see from Table 1 that the tournament subjected programs to a wide range of trading conditions. The base case (BASE) was an environment similar to the one used in pre-tournament trials at the SFTE. Other environments include duopoly and duospony (BBBS and BSSS), a degenerate surplus distributions where all players receive the same token values shifted by a common random constant (EQL), an independent private values environment where each trader's token is an *IID* draw from a uniform distribution (RAN), a single-period environment that prevented players from learning from previous market outcomes (PER), a "high-pressure" environment where the traders' time allotment was very short (SHRT), and an environment where each trader was only assigned a single token (TOK). Our intention was to force programs to compete under a broad range of conditions in order to provide a rigorous and comprehensive test of their effectiveness.

To insure that tournament earnings were not due to a series of lucky token draws, we developed a sampling scheme that guaranteed that all trading programs had equal surplus endowments with probability 1.[22] Once a random set of token values was drawn according to the sampling scheme given in Eq. (3.1), trading programs were randomly selected to play in a set of N games (where $N = 30$ is the total number of entrants) subject to the constraints that no program played a copy

[21]The monitor output contains a number of other symbols. @ denotes a bid equal to the current bid, X and Y denotes a bid below the current bid and the minimum allowed price, respectively (both illegal), and Z denotes a bid above the maximum allowed price. \$ denotes a bid above the current ask (and current bid), & denotes a bid above the current ask but not current bid, ~ denotes a token traded at a loss, and ! denotes a crossing of the current bid and ask.

[22]In the case of two trading programs that were only programmed to play one side of the market, a *Skeleton* stand-in trader was substituted in the games they refused to play. There are slight variations in actual token endowments caused by the fact that one program occasionally "died" midway through a trading period, resulting in forfeiture of its potential surplus in the remaining periods of the game.

TABLE 1 DA Trading Environments.

Parameter	Environment									
	BASE	BBBS	BSSS	EQL	LAD	PER	SHRT	SML	RAN	TOK
gametype	6453	6453	6453	0	0	6453	6453	6453	0007	6453
minprice	1	1	1	1	1	1	1	1	1	1
maxprice	2000	2000	2000	2000	2000	2000	2000	2000	3000	2000
nbuyers	4	6	2	4	4	4	4	2	4	4
nsellers	4	2	6	4	4	4	4	2	4	1
ntokens	4	4	4	4	4	4	4	4	4	1
nrounds	2	2	2	2	2	6	2	2	2	2
nperiods	3	3	3	3	3	1	3	3	3	3
ntimes	75	50	50	75	75	75	25	50	50	25
games	1624	1624	1624	1624	1624	1624	1624	3428	1624	1624
games/player	56	56	56	56	56	56	56	112	56	56
periods/player	336	336	336	336	336	336	336	336	336	336
conversion ratio ($\times 10^{-4}$)	6.11	8.95	9.73	3.48	3.57	6.97	7.04	6.32	1.04	20.6

of itself in the same game and all programs played all positions (B1, B2, S1, S2, etc.) an equal number of times. After this set of N games was completed, the scheme was repeated with a new set of token values. This sampling process guarantees that differences in the trading profits earned by the traders can be ascribed to differences in their trading ability, since each program received the same endowment of tokens and encountered roughly the same collection of opponents in a large number of replications of the DA game.

4. RESULTS OF DA TOURNAMENTS

We received 30 programs for the first (cash) tournament held in March, 1990. Table 2 summarizes the entries, listed by the name of the participant(s) who submitted the program.[23] Of the 30 entries, 15 were from economists, 9 from computer scientists, 3 from mathematicians, and the remaining 3 were from an investment broker,

[23]In cases where trading programs were developed by teams of individuals and we were unable to determine the primary author, we substituted a program nickname supplied by the authors. We also received two anonymous entries.

a professor of marketing, and a joint entry from two cognitive scientists. Several of the entries were outgrowths of research papers describing formal models of DA trading behavior.[24] Several of the entries emerged from working groups that co-developed sets of strategies, in some cases pre-testing them in "local tournaments" using our double auction software. These groups include seven entries from the Economic Science Lab (ESL) at the University of Arizona, three from the University of Minnesota, and two each from the University of Colorado (Economics) and Carnegie-Mellon University (Computer Science).[25] The table also includes four entries from SFI including the ZI, truthtelling, and pricetaking strategies discussed in Section 2, as well as a "skeleton" strategy provided to entrants as a simple example of a working trading program. Due to potential conflict of interest, none of the latter programs were entered in the cash tournament held in March 1990 although they were used as experimental controls in subsequent "scientific tournaments." All of the entrants programmed their strategies by replacing the bid/ask and buy/sell subroutines of the skeleton program with their own code. Although versions of the skeleton program were available in C, Fortran, and Pascal, almost all of the entries (26 out of 30) were programmed in C. Only two were written in Fortran, and two in Pascal.

We found that the most useful way of comprehending the variety of strategies in Table 2 was to classify them along the following dimensions:

simple *vs.* complex,
adaptive *vs.* nonadaptive,
predictive *vs.* non-predictive,
stochastic *vs.* non-stochastic,
optimizing *vs.* non-optimizing.

Rust, Miller, and Palmer[29] describe how these categories are defined and provide a detailed analysis of individual trading programs. In general, we found that although there was a substantial range in program complexity, the majority of the programs appeared to encode the entrant's "market intuition" using simple rules of thumb. Since these rules are "hard-wired," most of the programs are also classified as nonadaptive. Exceptions include a neural network program submitted by cognitive scientists Dallaway and Harvey, and an "adaptive cellular curve fitter" submitted by mathematician Paul Burchard. Besides using private information about token values, most of the programs relied on only small number of public information variables, the current bid, ask, and elapsed time being the most important. Only ten programs made use of the prior information about the distribution

[24] For example, Kennet-Friedman entry is based on Friedman's[11] model of DA trading as a Bayesian game against nature (BGAN), and the Ledyard-Olson entry is based on Easley and Ledyard's model described in Easley and Ledyard.[9]

[25] Individuals contributing from the Arizona group include Shawn LaMaster, Steve Rassenti, Roland Michelitsch, Kevin McCabe, Vernon Smith, Corinne Bronfman, and Mark Van Boening. Individuals contributing from the Colorado group include Greg Fullerton, Mark Cronshaw, Jamie Kruse, and B. J. Lee.

of token values provided by the `gametype` variable. Although 20 programs made use of the number of buyers and sellers in the DA game, only two programs (Ledyard-Olson and Staecker) used this information in an explicit way, e.g., by including separate "monopoly subroutines." Most of the programs did not attempt to keep track of the behavior of individual opponents or make statistical predictions of future market quantities. An exception was the program of Mark Staecker, developed as part of a senior honors thesis at the University of Western Ontario, which predicted the next high bid, low ask, and equilibrium price using market-level statistics from previous periods. Based on these predictions, Staecker's program decides if a transaction is likely to occur at the next BS step and if so, uses its predictions to place an "attractive" bid or ask in the next BA step.

The top-ranked program was submitted by economist Todd Kaplan of the University of Minnesota. It was one of the shortest programs submitted and is classified as *simple, nonadaptive, non-predictive, non-stochastic,* and *non-optimizing.* The second-ranked program was submitted by computer scientist Mark Ringuette from Carnegie-Mellon University. Despite the fact that they were independently developed, both strategies are remarkably similar. The strategies can be described in one line as *wait in the background and let the others do the negotiating, but when bid and ask get sufficiently close, jump in and steal the deal.* These programs succeed in "stealing the deal" by bidding an amount greater than or equal to the *previous* current ask. Ringuette's program differs from Kaplan's by randomly overbidding the previous current ask. When time is running out or when a long time has elapsed since making its last trade, Ringuette's program defaults to a modified version of the Skeleton bidding strategy whereas Kaplan's program places a bid equal to the smaller of the current ask or its current token value. In practice this implies that Kaplan's program eventually defaults to truthtelling mode when confronting patient opponents who delay making "serious" bids and asks.

4.1 RESULTS OF MARCH 1990 TOURNAMENT

Table 3 presents the dollar payoffs earned by the eligible trading programs in the March 1990 tournament, broken down by environment.[26] The top program, Kaplan, earned a total of \$408, \$14 higher than the second place program of Ringuette. The gaps separating third, fourth, and fifth place were \$7.45, \$10.51 and

[26] The NN program of Dallaway and Harvey was disqualified from the March 1990 tournament because it consistently incurred large losses. The authors submitted a revised version for the subsequent scientific tournament which performed satisfactorily.

TABLE 2 Taxonomy of DA trading programs.

Author/Nickname	Institution	F	L	C	A	S	P	O	CPU
Anon-1	Anonymous	CS	C	-	2	-	1	-	86
Anon-2	Anonymous	CS	C	-	3	-	1	-	89
Jacobson	Carnegie Mellon	CS	C	-	3	X	1	-	86
Ringuette	Carnegie Mellon	CS	C	-	2	X	1	-	85
Golden Buffalo	Colorado	E	C	-	2	X	2	-	88
Silver Buffalo	Colorado	E	C	-	2	X	2	-	88
Lin	Portland State	E	C	-	2	X	1	-	89
Perry	Portland State	M	C	-	3	X	2	-	88
Anderson	Minnesota	E	C	-	2		1	-	88
Breton	Minnesota	E	C	-	3	X	1	-	88
Bromiley	Minnesota	MK	F	-	2		1	-	90
Kaplan	Minnesota	E	C	-	3		1	-	86
Pricetaker	SFI	EM	C	-	2	X	1	-	88
Skeleton	SFI	EM	C	-	2	X	1	-	84
Truthteller	SFI	EM	C	-	1		1	-	84
ZI	SFI	EM	P	-	1	X	1	-	82
Exp	Arizona ESL	E	C	-	2		2	-	86
Free	Arizona ESL	E	C	-	2		2	-	87
Gamer	Arizona ESL	E	C	-	1		1	-	84
Max	Arizona ESL	E	C	X	2		2	X	157
Max-R	Arizona ESL	E	C	X	2		2	X	261
Slide	Arizona ESL	E	C	-	2	X	2	X	96
Terminator	Arizona ESL	E	C	-	2	-	2	-	89
Bolcer	UC Irvine	CS	P	-	2	-	2	-	86
Burchard	Princeton IAS	M	C	X	5	X	2	X	99
Dallaway/Harvey	Sussex	CS	C	X	5	-	1	X	91
Kennet/Friedman	Tulane/UCSC	E	P	X	2	-	2	X	181
Kindred	Duke	CS	C	-	2	X	1	-	89
Ledyard/Olson	Cal Tech	E	C	X	3	X	2	-	187
Leinweber	MJT Advisors	B	C	-	2	X	1	-	87
Lee	British Columbia	CS	C	-	2	X	1	-	88
Staecker	Western Ontario	CS	C	X	3	-	2	X	88
Utgoff	Massachusetts	CS	C	-	2	X	2	-	86
Wendroff/Rose	Los Alamos NL	M	F	-	3	-	2	-	88

Legend: F Field—B = broker, CS = computer science, E = economics, M = math, physics, MK = marketing.
L Programming Language—C, F = Fortran, P = Pascal.
C Complex—X if program is "complex" as defined in Rust, Miller, and Palmer.[29]
A Adaptive—5-level ranking defined in Rust, Miller, and Palmer[29]: 1 = least adaptive, 5 = most adaptive.
S Stochastic—X if program makes use of random number generator.
O Optimizing—X if program uses an explicit optimization principle.
P Predictive—3-level ranking defined in Rust, Miller, and Palmer[29]: 1 = doesn't predict, 2 = predicts market variables.
CPU Ratio of CPU time consumption to average for all programs.

TABLE 3 Dollar Payoffs in March 1990 Double Auction Tournament.

Trading program	Overall	BASE	BBBS	BSSS	LAD	EQL	PER	RAN	SHRT	SML	TOK
Kaplan	408	42	42	41	43	42	44	41	38	42	33
Ringuette	394	41	37	40	45	38	40	37	33	37	46
Staecker	387	41	39	37	43	36	41	35	35	38	42
Anon-2	376	33	36	40	40	39	33	34	39	36	46
Ledyard	367	34	38	37	38	36	36	35	35	37	41
Perry	366	34	41	36	40	36	32	38	35	35	40
Breton	360	35	33	37	36	37	35	37	35	37	40
Anderson	358	33	39	38	34	37	31	36	38	35	39
Anon-1	354	36	37	35	36	35	37	35	35	35	32
Burchard	344	34	35	34	34	34	35	39	36	34	28
Terminator	342	35	27	40	36	34	36	33	31	32	38
Golden Buffalo	340	34	35	35	35	35	32	37	34	30	33
Lee	337	33	34	32	35	35	35	37	35	36	27
Leinweber	333	34	34	32	39	36	32	36	35	34	22
Silver Buffalo	330	34	30	32	33	38	36	33	33	25	37
Slide	330	35	34	32	34	33	36	32	33	32	29
Jacobson	329	34	28	34	31	37	31	32	32	30	39
Bromiley	316	32	31	35	30	33	35	31	31	25	32
Max	299	31	32	33	30	28	32	25	22	32	34
Max-R	294	28	34	29	31	27	30	26	25	29	35
Utgoff	286	32	26	25	32	35	32	25	26	29	23
Kindred	271	30	24	33	32	34	31	30	27	25	4
Free	242	27	32	21	23	25	21	27	27	22	17
Gamer	230	25	20	22	24	24	25	21	25	21	22
Wendroff	228	25	20	21	25	24	23	26	25	15	25
Lin	224	21	20	23	22	25	17	25	25	19	27
Exp	210	16	17	23	26	26	20	11	14	23	34
Kennet	164	16	6	14	23	14	19	21	17	12	23
Bolcer	148	13	19	8	18	17	16	16	13	10	18
Total	$8967	897	881	898	945	928	903	893	870	847	905
Surplus	$10000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

$8.63, respectively. While these differences in earnings may not seem economically significant, they are statistically very significant. Kaplan's earnings are over 2.5 standard deviations higher than Ringuette's second place earnings, and the gaps separating first from second, third, and fourth places are 3.8, 5.6, and 7.1 standard deviations, respectively. The average standard deviation in profits of $5.75 was only slightly higher than the $5.40 standard deviation in surplus allocations, calculated over 3,360 individual periods of play in the ten environments. Recall that our procedure for generating tournament games guarantees that the token endowments of all traders are identical with probability 1. Given the large number of periods of play, an appeal to the law of large numbers allows us to be very confident that differences in traders' earnings reflect true differences in profitability rather than randomness due to player matchings and stochastic elements in the programs themselves.

The player rankings are also highly consistent across the ten environments. The average Spearman rank correlation between overall tournament payoffs and payoffs in each of the ten environments is 77%, ranging from a high of 95% in environment SML to a low of 72% in environment TOK. Kendall's W-statistic, which measures the degree of concordance in all the rankings, is highly significant at 79%, allowing us to easily reject the hypothesis that player rankings in different environments are independent. It is striking that Kaplan's program took first place in seven out of ten environments, coming in second place in environment EQL and third place in environment SHRT. The only place where Kaplan's program did not do well was the environment TOK where traders were endowed with only a single token. At the bottom end of the spectrum, the BGAN (Bayesian game against nature) program of Friedman and Kennet was consistently one of the worst performers in all ten environments. With earnings of $164.30, the BGAN is over ten standard deviations below the earnings of the next highest competitor.[27]

While we are very confident of our ability to distinguish the best and worst programs, we are much less confident about the relative rankings of the middle group of programs. One can see from Table 3 that after the large gaps separating the fourth and fifth place entries (Anon-2 and Ledyard-Olson), the differences in payoffs of the next group of programs are within one standard deviation of each other. The next significant difference in payoffs is a $10 gap separating the ninth place entry of Anon-1 from the tenth place entry of Burchard. Even after 3360 periods, it's clear that we would need many more observations to be confident of the relative rankings of programs between fifth and ninth place. In general, it's impossible to make any reliable performance distinctions if we can only observe traders over a small number of periods. The average dollar earnings of 9.4 cents per period of play is dominated by the per-period standard deviation in profits of 10.0 cents. Most of the latter variation is attributable to the 9.2 cent standard deviation in surplus arising from traders' random token endowments. Thus, a computerized trading environment is virtually a necessity if one wants to reliably discriminate good traders from bad. It appears that it would be infeasible to make the same sorts of distinctions in markets with human traders given that it takes hundreds or even thousands of periods of play before one can be sure that differences in relative performance are statistically significant.

Total tournament payouts at the bottom of Table 3 provide a convenient measure of trading efficiency, since conversion ratios from token profits to dollar payoffs

[27]BGAN may have suffered as a result of problems converting the program from PC Turbo Pascal to Sun Pascal. Although the converted program compiles without error, there were errors in several subroutines that generated under- and overflow errors at run time suggesting a possible incompatibility in its calls to certain functions. The low dollar ranking of the program of Bolcer should be disregarded since the program was only programmed to play the role of seller. In terms of dollar payoffs per game played, Bolcer's program is about equivalent in performance to the programs *Max* and *Terminator* which placed 19th and 20th in overall earnings.

were based on realized surplus rather than on realized profits. Thus, the total dollar payouts of \$8,967 correspond to an 89.7% efficiency ratio.[28] Efficiency was highest in the environment LAD, where fixed supply and demand curves were shifted by a random constant in such a way as to always yield a unique equilibrium price and quantity, and in EQL where traders where given symmetric token endowments (shifted by a random constant). However, the learning problem in both of these environments is nontrivial since traders were not given any prior information on the distribution of token values (i.e., `gametype` was set equal to 0), and therefore had no way of knowing that they had equal tokens or that there would be a unique equilibrium price and quantity.

It is probably also not a surprise that the least efficient environment was SML where there were only two buyers and two sellers. Human experiments reveal that the competitive properties of the DA market start to break down when there are so few traders.[29] The SML environment is but a step away from the most extreme situation of bilateral bargaining, which (when time constrained) is known to have inefficient outcomes owing to a high frequency of disagreement. The increased frequency of disagreement in the SML environment shows up in the distribution of trader's profits: even though surplus endowments are 0 only 5% of the time, traders walk away with 0 profits over 20% of the time. Other environments with relatively low efficiencies include SHRT (where there was a time constraint on trading) and BBBS and BSSS (duopoly and duopsony markets, respectively).

Overall efficiency levels appear to be somewhat lower than that observed in the later periods of experimental markets with human traders. We suspected that the low-trading efficiencies of the bottom ten programs were responsible for most of the aggregate inefficiencies, suggesting that running a tournament that excluded these programs would result in a more "realistic" and efficient market. Before doing so, we gave all entrants an opportunity to revise their programs in light of the results of the March 1990 tournament. A second series of scientific tournaments were conducted in May, 1991 with seven revised entries.[30] The scientific tournament also included several new programs written by the authors, including *Skeleton*, *Pricetaker*, and *ZI*.

[28] Payouts are net of profits earned by a *Skeleton* "stand-in" for the programs of Dallaway-Harvey and Bolcer which only played the roles of buyer and seller, respectively. If we were to include profits earned by *Skeleton*, aggregate market efficiency would be slightly higher.

[29] Although perhaps not as badly as we might have expected, *a priori*. See Clauser and Plott,[7] chapter 12 of this volume.

[30] We received revised entries from Dallaway-Harvey, Ledyard-Olson, and Perry, and four revised entries from the Arizona ESL group, *Max*, *Max-R*, *Slide*, and *Terminator*.

4.2 RESULTS OF SCIENTIFIC AND TOP 17 TOURNAMENTS

Overall player rankings in the "scientific tournament" were quite similar to the original tournament. Aggregate efficiency increased slightly from 90% to 92%, ranging from a low of 88% in SHRT to 95% in EQL. Once again the programs of Kaplan and Ringuette were the clear winners, with Kaplan's program trading at an overall efficiency level of 121%, significantly higher than Ringuette's efficiency of 116%. Staecker's program, trading at an efficiency level of 109%, came in third place just as in the original tournament. Two independent copies of *Skeleton* placed fifth and eighth place, which is somewhat surprising given that *Skeleton* was provided to all participants in advance of the tournament and thus should have represented a fairly easy target to beat.[31] On the lower end of the scale, the programs *Gamer*, *Exp*, and *Lin*—which were among the poorest performers in the original tournament remained the poorest performers in the scientific tournament, trading at an average efficiency level of under 70%. *ZI* also performed quite poorly trading at an efficiency level of 75%, confirming our discussion in Section 2. The revised neural-network trader of Dallaway and Harvey performed somewhat better than *ZI*, trading exclusively in the role of buyer at an efficiency level of 85%. *Pricetaker* turned out to be the median trader at rank 17 with an average efficiency level of 95%. This is somewhat below the 100% efficiency ranking we would have expected for our implementation of a naive pricetaking strategy.

In order to see whether the leading position of Kaplan and Ringuette is a result of general superiority or merely a relative superiority in their ability to exploit the lowest ranked traders, we ran a third tournament consisting of the top 17 players from the scientific tournament. The results of the "Top 17" tournament are summarized in Table 4. We see that once again, Kaplan and Ringuette remain the leaders even after elimination of the lowest ranked traders. Ringuette does slightly better than Kaplan in terms of overall earnings, although the difference is not statistically significant.[32] Kaplan's program comes in first place in five of the ten environments whereas Ringuette's program is first in four environments, and the two programs are tied for first place in environment SML. The fact that Kaplan and Ringuette were able to maintain their high efficiency ratios in this tighter, more competitive market suggests that they are, in fact, generally superior to all of the other programs.

The only program that improved significantly in the Top 17 tournament was Anderson, which moved from twelfth place to fourth place. The trading efficiencies of the remaining programs generally declined significantly, especially Staecker, Lee,

[31]Two copies of *Skeleton* were included in the scientific tournament as a further check on the statistical reliability of our rankings. The relative performance of the two copies is rather close in all environments except TOK, where the two copies traded at efficiency levels of 41% and 84%, respectively.

[32]Note that overall earnings in the scientific tournament were computed by summing *token* profits as opposed to dollar earnings in the March 1990 tournament. This change in implicit weighting scheme may account for the change in overall rankings of Kaplan and Ringuette.

TABLE 4 Trading efficiencies in the "Top 17" tournament.

Trading program	Overall	BASE	BBBS	BSSS	LAD	EQL	PER	RAN	SHRT	SML	TOK
Ringuette	117	121	103	116	120	113	119	120	91	126	130
Kaplan	114	110	109	122	103	120	129	111	118	126	116
Anon-1	108	109	110	110	111	108	103	106	113	117	109
Anderson	105	100	104	102	115	106	104	103	102	107	101
Staecker	101	111	103	107	99	103	108	96	97	117	126
Burchard	101	94	92	92	91	107	84	112	95	81	84
Perry	98	103	104	100	97	106	106	95	87	91	97
Anon-2	97	99	102	98	108	95	93	93	88	107	124
Lee	96	90	88	81	96	105	90	101	88	90	88
Ledyard/Olson	96	90	90	89	98	104	94	97	90	96	103
Golden Buffalo	94	90	94	88	101	91	91	97	81	80	118
Breton	92	102	94	91	94	92	98	91	81	94	105
Leinweber	90	86	83	88	87	89	87	98	85	67	33
Skeleton	89	91	89	78	89	100	84	93	93	62	39
Jacobson	89	97	91	92	96	80	87	87	88	94	104
Silver Buffalo	82	87	84	74	87	91	81	79	80	71	102
Pricetaker	71	81	81	95	77	68	87	57	87	98	96
Market	97	98	95	96	98	99	97	96	92	96	98

Skeleton, Breton, and *Pricetaker*. However, despite these declines, average market efficiency increased to 97%. The latter efficiency levels are as high as efficiency levels observed in comparable human experiments. Indeed we found that when we participated as human traders in the Top 17 market, it was difficult to consistently trade at higher than 100% efficiency, whereas we found it relatively easy to consistently trade at higher than 100% efficiency in markets that included the lowest ranked trading programs. This suggests that the Top 17 market may serve as a good working model of a "competitive market" such as observed in experimental markets with human traders.

4.3 AGGREGATE BEHAVIOR OF COMPUTERIZED TRADERS: SOME "STYLIZED FACTS"

A more detailed analysis of tournament results reveals that the top-ranked trading programs do in fact yield a fairly realistic working model of a DA market in the sense that their collective behavior is consistent with the following stylized facts of human DA markets: (1) convergence to CE, (2) high *ex post* efficiency levels, (3) reductions in transaction-price volatility and efficiency losses in successive trading periods reflective of apparent "learning" effects, (4) coexistence of extra-marginal and intra-marginal efficiency losses, (5) low-rank correlations between the realized order of transactions and the "efficient" order, and (6) negatively autocorrelated transaction-price changes. Even in markets that include the less efficient, lower ranked traders, transaction-price trajectories appear to be very similar to those observed in human markets. Indeed, we typically find that transaction prices and quantities converge close to the CE in the very first trading period.

Figure 3 shows a typical outcome, game BASE012 of the scientific tournament. The figure plots the induced supply and demand curves and the transaction price trajectories in each of the three trading periods. All three trajectories converged to the CE, generating nearly 100% efficient outcomes. In this case the market traded at 100% efficiency in the first trading period, compared to 98% in the second and third periods. Despite the high *ex post* efficiency, the rank correlation coefficient between the order of the buyer's and seller's trades and the "efficient order" is very low, corresponding to what we observe in human experiments. For example, the rank correlation for buyers and sellers in period 1 is 60% and 48%, respectively, falling to just 10% in period 3. The information in the right border of Figure 3 shows the times and traders involved in each transaction made in period 3. For example, the 1st token was traded in BS step 3 when buyer B3 accepted the offer of seller S4 (at a price of 429), and the eighth token was traded in BS step 51 when buyer B2 accepted the offer of seller S1 (at a price of 436). If trades were made in the efficient order as predicted by Wilson's WGDA theory described in section 2, then the first token should have been traded by B2 and S4 and the eighth token should have been traded by B2 and S2. The automata traders frequently trade extra-marginal tokens, something that is commonly observed in human experiments but is also ruled out by game theoretic models such as Wilson's WGDA. For example, in Figure 3 we see that in period 3 buyer B4 succeeded in buying three tokens, "bumping" buyer B1 who only succeeded in buying one token (in an efficient allocation, each buyer and seller trade their two most valuable tokens). As a result, the traders failed to exploit 27 units of potential surplus in period 3.[33]

For comparison, Figure 4 presents typical price trajectories in a market with 100% ZI traders. The nature of the ZI strategy implies that transaction-price sequences in successive periods are *IID*, so these traders cannot exhibit the type of learning behavior that is characteristic of human DA markets. It is evident that

[33]The four-way decomposition of lost surplus, EM, IM, BS, and SS on the right-hand border of Figure 3 is explained in Section 4.4 and the appendix. The left border of Figure 3 presents other statistics on the trading process, broken out by period. These include the correlation coefficient of transaction-price changes (C), *ex post* efficiency (E), time of last transaction (T), Spearman rank correlations of the order of the transactions with the efficient order for buyers (B) and sellers (S), and the maximum and average absolute percentage deviation of transaction prices from the midpoint equilibrium price (A), (M).

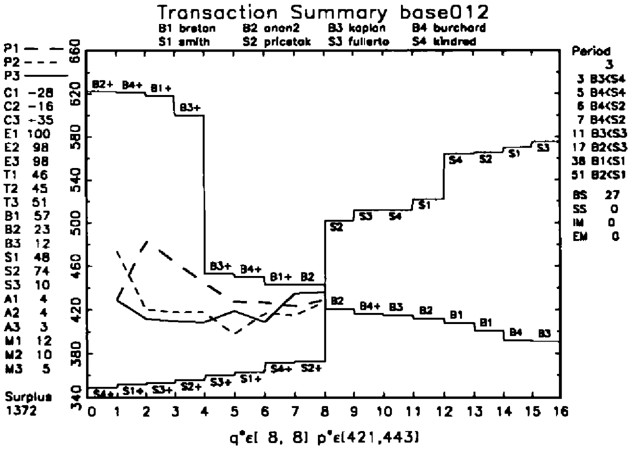


FIGURE 3 Price trajectories in DA game BASE012.

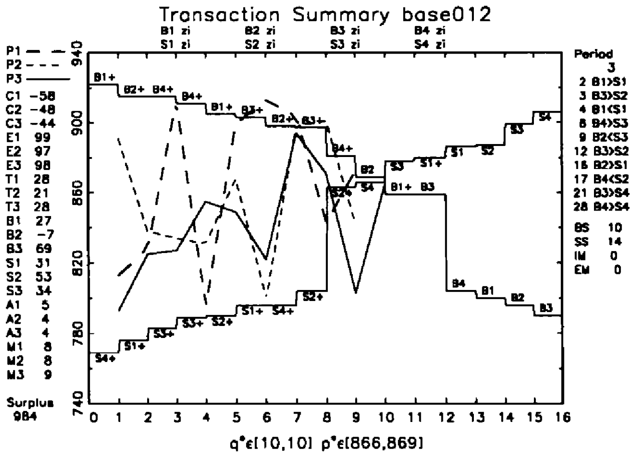


FIGURE 4 Price Trajectories for ZI Traders.

TABLE 5 Summary statistics for successive trading periods, P1, P2, and P3.

Statistic (standard deviation)	Top 17 Traders 2327 Games			All Traders 4274 Games			ZI Traders 2295 Games		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
BRANK	47.8	46.4	46.0	48.5	46.9	46.5	57.6	57.4	56.6
		(.9)			(.6)			(.9)	
SRANK	48.0	47.5	47.2	48.9	48.8	48.5	53.1	52.2	52.5
		(.9)		(.6)		(.6)		(.9)	
CORRCHG	-26.9	-24.0	-23.9	-29.9	-27.9	-27.2	-48.1	-46.6	-47.4
		(1.1)			(.8)			(1.0)	
COEFVAR	8.1	6.5	6.2	10.7	8.7	8.6	12.7	12.7	12.7
		(.2)			(.1)			(.2)	
DEV, MAX	19.1	15.4	14.6	23.4	19.8	19.4	27.1	27.0	27.2
		(.3)			(.3)			(.5)	
DEV, AVERAGE	.42	.15	0.7	-.71	-1.24	-1.20	3.23	3.46	3.17
		(.22)			(.18)			(.22)	
DEV, LAST	.21	.23	0.02	-.03	-.68	-.59	.73	1.15	.60
		(.19)			(.18)			(.24)	
ABS-DEV, AVERAGE	9.9	8.0	7.7	11.4	9.9	9.7	11.9	12.0	11.9
		(.2)			(.2)			(.2)	
ABS-DEV, LAST	5.0	4.6	4.4	6.6	6.5	6.5	6.7	6.8	6.9
		(.2)			(.2)			(.2)	
HIT RATE	23.5	26.3	27.2	18.5	18.7	18.9	19.0	18.0	19.7
		(.9)			(.6)			(.8)	
HIT RATE 2	48.8	52.3	53.2	38.6	39.2	39.4	36.9	36.6	37.6
		(1.0)		(.7)		(.7)		(1.0)	
HIT RATE 5	57.9	62.4	62.6	48.2	49.4	48.6	52.3	51.4	51.2
		(1.0)		(.8)		(.8)		(1.0)	
HIT RATE 10	82.6	86.2	86.8	77.0	76.7	76.8	82.4	81.1	81.2
		(.7)		(.6)		(.6)		(.8)	
PCT2ND	40.2	34.3	33.4	27.6	26.1	25.7	5.7	5.7	5.9
		(.4)			(.3)			(.2)	
EFF, AV	94.8	96.7	96.9	90.6	91.1	91.5	97.0	97.0	96.9
		(.1)			(.3)			(.1)	
EFF, TOTAL	95.6	96.8	97.0	92.3	92.1	91.8	97.9	97.8	97.9

Legend: BRANK Rank correlation of buyers' transactions with efficient order
SRANK Rank correlation of sellers' transactions with efficient order
CORRCHG Correlation coefficient of transaction price changes
COEFVAR Coefficient of variation of transaction prices
DEV,MAX Maximum % deviation from midpoint equilibrium price
DEV,AV Average % deviation from midpoint equilibrium price
DEV,LAST % deviation of last transaction price from midpoint eq. price
ABS-DEV,AV Average absolute % deviation from midpoint equilibrium price
ABS-DEV,LAST Absolute % deviation of last transaction price from midpoint eq. price
HIT RATE % price trajectories where $q_t \in [\underline{q}, \overline{q}]$ and $p_t \in [\underline{p}, \overline{p}]$
HIT RATE 2 % price trajectories where $q_t \in [\underline{q}, \overline{q}]$ and $p_t \in [.98\underline{p}, 102\overline{p}]$
HIT RATE 5 % price trajectories where $q_t \in [\underline{q}, \overline{q}]$ and $p_t \in [.95\underline{p}, 105\overline{p}]$
HIT RATE 10 % price trajectories where $q_t \in [\underline{q} - 1, \overline{q} + 1]$ and $p_t \in [.9\underline{p}, 1.1\overline{p}]$
PCT2ND % of trades occurring in 2nd half of trading period
EFF,AV Average of Profits/Surplus for individual games
EFF,TOTAL Total Profits/Total Surplus
Where: p_t is last transaction price; $[\underline{p}, \overline{p}]$ is equilibrium price interval; and $[\underline{q}, \overline{q}]$ is equilibrium quantity interval

overall transaction-price volatility is significantly higher for ZI traders in all periods. The zig-zag pattern of transaction-price trajectories implies serial correlation coefficients in the neighborhood of -50% predicted by Cason and Friedman.[6] Finally, notice that ZI traders are very impatient: they complete all their transactions in the first third of the trading period.

The behavior displayed in Figures 3 appears qualitatively similar to outcomes of human experiments. To demonstrate that these examples are not atypical, Table 5 presents an array of trading statistics averaged over all tournament environments. In order to highlight learning effects, we break out the statistics by trading period. To help put these statistics in perspective, we also present results for a tournament with 100% ZI traders.

Table 5 suggests that the Top 17 traders exhibit significant inter-period learning effects: as we move from the first period (P1) to the last (P3) we see that trading efficiency increases, the fraction of price trajectories hitting the CE target increases, and price variability decreases (whether measured by the coefficient of variation or by the percentage deviation from the midpoint equilibrium price). The only statistics that show no systematic improvement in successive periods are the rank correlation coefficients of buyers' and sellers' transaction sequences *vis-à-vis* the efficient trading order and the serial correlation coefficient for transaction-price changes. The rank correlations for the Top 17 traders are significantly less than the values for ZI traders and less than the 100% predictions of Wilson's WGDA theory, but are somewhat above the 0–30% range for the human experiments analyzed in Cason and Friedman.[6] The autocorrelation of transaction-price changes for the Top 17 traders averaged -25% which is significantly less than the -50% value for ZI traders and the 0% value predicted by the WGDA theory, but roughly consistent with the values reported in Cason and Friedman[6] for experienced human subjects.[34] In terms of efficiency, the Top 17 and ZI markets are roughly comparable at 97%, which is actually somewhat higher than the 93% average efficiency for DA markets with inexperienced subjects reported in Cason and Friedman,[6] but roughly comparable to the 97.8% average efficiency in first five periods of DA experiments with experienced subjects reported in Table 2.d of McCabe, Rassenti and Smith.[23]

Learning effects are much less pronounced when we include the lowest ranked trading programs and, of course, are completely absent in markets with 100% ZI traders. Notice that while ZI traders attain the highest efficiency levels, price volatility (measured by any of the statistics in the second panel of Table 5) is significantly higher than the Top 17 traders. ZI traders are noticeably less patient, exchanging only 6% of their tokens in the second half of the trading period compared to over 33% for the Top 17. The statistics on the "hit rates" show that despite the high efficiency of ZI markets, less than 20% of all price trajectories actually converge to equilibrium, compared to nearly 27% in the third period in comparable markets

[34] Cason and Friedman report autocorrelations of -50% for inexperienced subjects, suggesting that trader experience tends to push autocorrelations towards 0%, although the results are inconclusive since they are based on a relatively small number of observations.

with Top 17 traders. If we widen the target slightly and count any price trajectory that is within 5% of the equilibrium price interval, then the hit rate increases to over 60% for the Top 17 traders compared to just over 50% for the ZI traders. Hit rates for all traders in the scientific tournament are slightly lower than the ZI traders.

TABLE 6 Analysis of Efficiency Losses.

Case		Scientific Tournament					Top 17 Tournament				
Environment	Period	% cost Surplus	EM	IM	SS	BS	% cost Surplus	EM	IM	SS	BS
BASE	1	93.7	10.1	35.8	17.3	36.8	97.0	35.8	33.7	13.2	17.3
	2	93.4	9.6	43.0	12.4	34.9	97.8	41.0	31.0	12.2	15.9
	3	93.4	10.2	41.4	13.9	34.5	98.2	49.0	19.4	14.4	17.2
BBBS	1	91.2	4.2	45.2	2.1	48.5	93.7	10.0	53.1	4.8	32.1
	2	90.3	3.2	48.7	2.1	45.9	95.9	13.7	43.7	4.1	38.5
	3	90.6	3.2	47.4	2.0	47.4	95.9	13.8	34.2	4.9	47.1
BSSS	1	91.3	0.8	55.0	42.0	2.2	94.2	6.8	36.5	50.7	6.0
	2	92.2	0.7	51.0	46.1	2.2	96.0	9.6	13.4	70.4	6.6
	3	91.6	0.3	57.0	41.3	1.5	96.0	11.1	16.0	67.5	5.4
EQL	1	94.8	0.0	82.9	5.7	11.5	98.3	0.0	53.1	26.0	20.8
	2	94.9	0.0	89.7	3.2	7.1	98.8	0.0	50.0	24.3	25.7
	3	95.1	0.0	89.3	4.2	6.5	98.9	0.0	50.8	19.7	29.5
LAD	1	94.0	8.8	56.6	13.8	20.8	97.8	21.0	26.2	28.6	24.2
	2	93.9	6.4	53.4	18.6	21.6	98.3	23.4	17.2	26.2	33.2
	3	93.7	6.2	53.9	16.8	23.1	98.4	29.5	14.3	25.5	30.7
RAN	1	91.8	21.0	30.8	16.9	31.3	95.6	44.2	13.3	20.5	22.0
	2	91.5	14.6	36.7	11.4	37.3	96.4	49.9	13.2	14.9	22.1
	3	90.8	11.6	38.5	10.6	39.3	96.6	52.6	15.9	12.4	19.1
SML	1	91.1	7.8	72.7	9.1	10.4	93.0	8.3	69.2	16.9	5.5
	2	91.3	7.9	72.5	7.6	11.9	96.3	15.9	50.0	24.6	9.5
	3	90.8	6.8	74.0	6.7	12.4	97.1	16.3	43.7	23.5	16.5
TOK	1	95.0	11.7	85.5	1.4	1.3	97.5	0.0	95.7	2.3	2.0
	2	93.2	2.9	95.6	0.5	1.0	99.1	0.0	94.2	5.8	0.0
	3	94.0	2.3	96.3	0.5	0.9	98.2	0.0	95.9	2.7	1.4
SHRT	1	88.3	3.2	62.7	11.4	22.6	89.3	4.5	70.5	15.5	9.5
	2	88.1	3.0	62.6	10.6	23.8	92.9	4.5	70.5	12.8	11.7
	3	87.6	2.5	64.5	10.3	22.8	93.4	6.4	67.1	15.3	11.2

The other major difference between ZI and the Top 17 traders is that the rank order correlations of the trading sequences are significantly higher, and the serial correlation coefficient of transaction-price changes are significantly more negative.

We are presently collecting detailed data sets that will allow us to go beyond the simple stylized facts outlined above and conduct more precise statistical comparisons of the behavior of human and computer traders. Our conjecture is that humans will display much more dramatic inter-period learning effects than the Top 17 programs. It is also probable that we will find significant differences in the stochastic properties of price trajectories in later periods of the game, as well as differences in the timing of bids, asks, and transactions.

4.4 ANALYSIS OF DA EFFICIENCY LOSSES

One of the stylized facts of human DA markets is that a major fraction of efficiency losses are due to trades of extra-marginal tokens. In order to quantify the magnitude of these losses, it's useful to distinguish between four types of inefficiencies that can occur in DA markets:

IM: value of lost surplus of non-traded intra-marginal tokens (i.e., those that lie to the left of the equilibrium quantity, q^*) when the actual number of trades q is less than q^* (0 otherwise).

EM: value of lost surplus due to trade of extra-marginal tokens (i.e., those that lie to the right of q^* on the supply and demand curves) when the actual number of trades q is greater than q^* (0 otherwise).

BS: value of lost surplus due to trades of extra-marginal buyers' tokens that displaced potential trades of an equal number of buyers' intra-marginal tokens.

SS: value of lost surplus due to trades of extra-marginal seller's tokens that displaced potential trades of an equal number of seller's intra-marginal tokens.

Table 6 presents an "inefficiency audit" that summarizes this four-way decomposition of efficiency losses. The last column of each section of table presents the ratio of total profits to total surplus in each period of play, and the remaining columns present a percentage breakdown of lost surplus due to each of the four sources, EM, IM, SS, and BS. In general, the audit reveals that total extra-marginal efficiency losses—the sum of EM, SS, and BS—constituted a substantial fraction of total efficiency losses in all environments except TOK. As noted above, this finding is consistent with the results of human experiments that show that efficiency losses are frequently a result of trading too many tokens rather than too few tokens. Extra-marginal efficiency losses were identically zero only in the EQL and TOK environments. In EQL this result is to be expected given the nature of the token distribution which resulted in supply and demand curves with large steps, each four units wide, a unique equilibrium price, and a four-unit range of market-clearing quantities. In the single token TOK environment, if an EM efficiency loss occurs, it is more likely that some trader has taken a loss on a transaction. We can see that in the scientific tournament EM losses in the TOK environment are nonzero, reflecting the fact that some traders (principally the programs *Free* and *Kindred*) were trading at a loss. Once these lower ranked programs were removed in the Top 17 tournament, EM efficiency losses were virtually eliminated.

Overall, Table 6 shows that the largest single source of inefficiency is IM, indicating that generally too few rather than too many tokens were traded. This effect is most noticeable in environments TOK and SHRT. Such a result is to be expected for the SHRT environment due to the small number of trading steps. However, the large value of IM in the TOK environment is surprising, given that one would expect it would be much easier to trade a single token instead of four.[35] Intra-marginal efficiency losses were also large in the duopsony and duopoly environments BSSS and BBBS. In BBBS this may possibly reflect sellers' attempts at "collusion" in order to restrict output in an attempt to share joint monopoly profits. Note, however, that the other large source of efficiency losses is due to excessive competition on the long side of the market. Thus, in environment BSSS where there are three sellers for each buyer, the large value of the SS efficiency losses indicate that the six sellers became engaged in "price wars" as they competed to sell their tokens to the two buyers.

Table 6 shows that the level of intra-marginal efficiency losses are typically significantly lower in the Top 17 tournament, reflecting the fact that the majority of the lower ranked trading programs did poorly as a consequence of failing to trade all of their potentially profitable tokens. On the other hand, the relatively higher levels of extra-marginal efficiency losses in the Top 17 tournament provide another indication that this is indeed a relatively more competitive and aggressive market.

4.5 RESULTS OF THE "EVOLUTIONARY TOURNAMENT"

A limitation of the previous tournaments is that trading programs were not allowed to play against themselves. In order to provide a simple model of imitation and growth processes, we decided to conduct an "evolutionary tournament" based on ideas from evolutionary biology.[3,22][36] The idea is that in real-world markets the best traders will attempt to "clone" their strategies in order to gather a larger market share. These processes will tend to lead to expansion in the number of traders using effective trading rules and declines in the number using poor trading

[35]This result may indicate possible programming problems if the majority of entrants developed their programs on the assumption that traders would normally be endowed with four tokens. Although tournament rules explicitly noted the possibility of single-token environments, all pre-tournament games run on SFTE involved four-token environments. Entrants may have assumed that single-token environments would constitute a negligible share of tournament profits and focused their attention on the four-token case.

[36]Actually, from a strict biological perspective, it would be more accurate to call it an "ecological tournament" since the set of species (trading programs) is fixed and only their relative proportions are allowed to change over time.

strategies. However, the changing market composition may also present opportunities: some trading rules may actually perform better in a tighter, more competitive market.

In an evolutionary tournament each trading program is assigned a measure of "fitness," and an initial population of programs evolves over time in accordance with the principle of survival of the fittest. Specifically, let the fitness level of program i in game t be given by its *capital stock* $K_i(t)$. A trader's fitness evolves over time according to the law of motion:

$$K_i(t) = K_i(t-1) + \Pi_i(t) - S_i(t) \tag{4.1}$$

where $\Pi_i(t)$ is trader i's profit in game t, and $S_i(t)$ is the surplus (i.e., token value) assigned to trader i in game t. Thus, our measure of fitness corresponds to trading efficiency: fitness increases when trading efficiency exceeds 100% and declines otherwise. A real-world interpretation is that traders are buying and selling shares of stock. In game t the trader purchases shares of stock at an initial cost of $S_i(t)$, closing out his account at the end of the day by selling off his shares for $\Pi_i(t)$. The competitive hypothesis that price and quantity converge to CE can be reinterpreted as the "efficient markets hypothesis" that the expected value of end-of-day holdings $\Pi_i(t)$ equals the initial purchase price $S_i(t)$. However, if markets are not completely efficient, then superior traders should be able to make positive expected profits, tending to increase their capital stocks over time.

The concept of survival of the fittest is operationalized by letting the fraction $p_j(t)$ of traders using strategy j in game t be proportional to the relative capital share of the type j traders:

$$p_j(t) = \frac{K_j(t)}{\sum_{i=1}^{I} K_i(t)}. \tag{4.2}$$

We might think of Eq.(4.2) as reflecting the Dean Witter philosophy, "one investor at a time." Thus, if Dean Witter manages $p_j(t)$% of the stock of investment capital, then on any given day we would expect that approximately $p_j(t)$% of the traders in the market will be employed by Dean Witter.

We begin the evolutionary tournament by endowing all traders with equal initial capital shares, $K_i(0) = \overline{K}$, $i = 1, \ldots, I$. In a DA market with L buyers and M sellers, we begin the evolution by taking L *IID* draws (with replacement) from the population of buyers and M *IID* draws from the population of sellers to form the market using the multinomial distribution (4.2). Given a randomly selected set of traders, we play a DA game from the BASE environment with a randomly selected set of tokens. After trading is complete, the capital stocks of the L buyers and M sellers are updated according to Eq. (4.1), new selection probabilities are computed according to Eq. (4.2), and a new set of players are drawn for game $t = 1$. It is also easy to construct a market where a constant fraction of noise traders, say, p%, enters the market in each game t. For each of the L buyers and M sellers in the DA market, we draw *IID* Bernoulli random variables with parameter p. If the outcome of the mth Bernoulli random variable B_m is 0, then the mth seller is selected from the pool of I "permanent" traders according to Eq. (4.2); otherwise if $B_m = 1$, then the mth seller is selected at random from a fixed set of noise-trader programs (and similarly for buyers).[37]

We have not proved any results about the limiting behavior of the evolutionary tournament such as whether or not the limiting set of traders form a *stable set* along the lines of Maynard Smith's notion of "evolutionary stable strategies" (ESS). Indeed, it is not even clear that one can define the precise conditions under which the evolution of capital stocks constitutes an ergodic stochastic process. It is easy to see that in the case of a closed market with no inflow of noise traders, the aggregate capital stock is indeed an ergodic stochastic process: any market with nonzero efficiency losses must eventually hit an absorbing state of zero capital with probability 1. However, our computer simulations suggest that the I-dimensional stochastic process of capital shares is non-ergodic. Indeed, our computer simulations indicate that the long-run outcome of competition among our fixed set of trading programs is unstable; i.e., the stable set is empty.

Figures 5 and 6 clearly illustrate this result in the case of an evolutionary tournament conducted under the BASE environment. We evolved capital stocks for buyers and sellers separately to test whether there are any asymmetries in the traders' performance. All 66 traders (33 buyers and 33 sellers) were equally endowed with an initial capital stock of 80,000. We stopped the tournament after 28,000 games when the buyers' capital stock had dwindled to less than 8% of its original value, as can be seen from the line labelled "market" in Figure 5. The striking feature about both figures is that after about 5,000 games, Kaplan's program emerged as the clear leader, dominating both the buyers' and sellers' market after 20,000 games. The programs of Ringuette and Staecker have the second and third largest capital shares, but notice that after 20,000 games their influence began to diminish, losing out to the dominant competition of the Kaplan traders. This effect is especially pronounced among the buyers, and after 22,000 games the capital stocks of Ringuette and Staecker have been reduced to less than their initial allocations, allowing Kaplan to achieve near total domination of the market. However, this is precisely when Kaplan's program began to head into a precipitous decline, losing more than half of its capital stock in the succeeding 6,000 games.

The reason for Kaplan's fall is clear: the success of a "wait in the background" strategy depends on being in a market populated with active bidders. If all traders attempt to try to wait in the background, little information will be generated as each trader waits for the others to make the first move. By the end of the period, there will be a trading panic as all traders attempt to unload their tokens. Thus, there

[37]There are two ways to handle the profits (or losses) earned by noise traders. One way is to update capital stocks of the noise traders according to Eq. (4.2), treating them as a subset of the set of permanent traders. This effectively provides a lower bound of p/N on any trader's participation probability, where N is the number of noise traders. The other way is not to update the capital stocks for the programs which were selected as noise traders. This latter method may still indirectly increase the capital stocks of the permanent traders to the extent that they are able to systematically extract surplus from the noise traders.

is a much higher likelihood that the period will expire with unexploited surplus left on the table. This leads to a sharp fall in aggregate efficiency, precipitating the market "crash" that is evident in Figure 6.

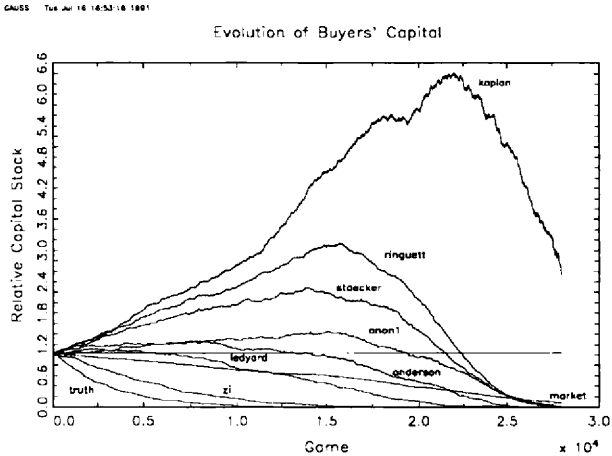


FIGURE 5 Results of the "evolutionary tournament": buyers.

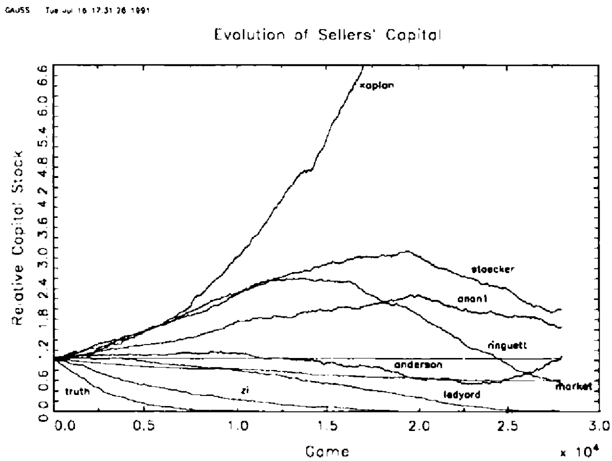


FIGURE 6 Results of the "evolutionary tournament": sellers.

This suggests that it cannot be collectively stable for all traders to adopt a "wait in the background" strategy: doing so creates a serious *information externality* that prevents the market from converging to CE. However, in Kaplan's case the negative impact of the information externality is actually not the primary reason for his precipitous decline. The primary reason is that his program switches into "truthtelling mode" once it succeeds in monopolizing the market. Specifically, if a long time has elapsed since the last trade, or if time remaining in the trading period is running out, Kaplan's program places a bid equal to the minimum of the current ask and $T - 1$ where T is the value of the next untraded token. In practice, the current ask is likely to exceed $T - 1$ after a long period of inactivity (otherwise the program would have already jumped in), implying that Kaplan's program effectively becomes a truthteller by bidding $T-1$. Long periods of inactivity—a sign of impasse between buyers and sellers—are quite likely to occur when at least one side of the market is dominated by background traders like Kaplan or Ringuette. It follows that once it succeeds in monopolizing the market, Kaplan's program necessarily switches into truthtelling mode.

Although truthtelling is a very bad strategy in a market populated by even slightly smarter strategies (as is clearly evident from the fact that *truthteller* provides the lower envelope for the capital trajectories in Figures 5 and 6), truthtelling can be collectively stable if all traders adopt it simultaneously since a market with 100% truthtellers is necessarily 100% efficient. Indeed, that is what we found when we ran a tournament with 100% Kaplan traders in the BASE environment. This suggests that Kaplan's two-stage wait-in-the-background/truthtelling strategy might seem like a clever way to gain and maintain a monopoly position.[38] However, closer examination of Figures 5 and 6 reveal the inherent danger of this approach. By game 22,000, Kaplan's program had virtually monopolized the buyers' market, but it still had not completely monopolized the sellers' market.[39] Thus, Kaplan's buyers tended to get locked into a waiting game with each other, causing them to switch into truthtelling mode, resulting in a net windfall to the sellers (who were still predominantly in background mode). This scenario is clearly reflected in the fact that the steady decline in seller's capital decelerated after game 22,000, so that by game 28,000, sellers still had nearly 60% of their initial capital stock. Since Kaplan's program dominated the seller's market, it gained the most, multiplying its initial capital stock over 12 times to 984,735.

Given the overall similarity in the programs of Kaplan and Ringuette, why does Kaplan dominate in the long run? Initially, both Kaplan and Ringuette succeed in exploiting the other programs, as can be seen by their relatively equal capital shares in the first 5,000 games of the evolutionary tournament. However, once Kaplan

[38] It is unlikely that Kaplan designed his program with this idea in mind since the rules for awarding cash prizes in the March, 1990 tournament are inconsistent with the use of the kinds of evolutionary tournaments which we subsequently designed.

[39] By game 28,000 the top sellers were Kaplan (64%), Staecker (10%), Anon-1 (8%), Perry (7%), Anderson (6%), and Ringuette (3%). In contrast, Kaplan controlled nearly 98% of the buyers' capital.

and Ringuette start to codominate the market, Ringuette's efficiency levels fall from nearly 120% to well below 100%, allowing Kaplan to take the lead. It is not easy to identity the precise cause of this outcome, although parts of the reasons have been discussed in our comparison of Kaplan and Ringuette in Rust, Miller, and Palmer.[29] It appears that the principal reason for Ringuette's decline is that his program is more impatient to trade than Kaplan's, switching from the role of a background trader to an active bidder much sooner than Kaplan does. On the other hand, when Ringuette's program encounters a period of inactivity, it invokes a much more aggressive active bidding strategy than Kaplan, namely a slightly modified version of *Skeleton.* However, since *Skeleton* consistently loses out to Kaplan when the latter is in "background mode," it should not be surprising that Ringuette's program also loses given that it typically switches to "skeleton mode" long before Kaplan's program has switched to truthtelling mode. Specifically, Ringuette's program switches to skeleton mode whenever the number of elapsed steps since the last trade exceeds the smaller of 12, or 60% of the remaining steps in the game.[40] In comparison Kaplan's program switches to truthtelling mode whenever the number of time steps since the last transaction exceeds one half of the number of remaining steps, or whenever five steps have elapsed since the last transaction and the number of steps since his program last traded exceeds two thirds of the remaining steps. Thus, at the beginning of a DA game dominated by Kaplan and Ringuette, Ringuette's program will switch to skeleton mode after BS step 12 whereas Kaplan's program will still be playing the role of a background trader.

In biological terms, our evolutionary tournaments have suggested the following conclusions: (1) Kaplan's strategy appears to be able to successfully invade a population of non-Kaplan's (including Ringuette), and (2) a collection of 100% Kaplan strategies is not collectively stable. These conclusions suggest that the outcome of "closed" evolutionary tournaments (i.e., ones that exclude subsequent entry of noise traders) may be characterized by cycles of "booms" and "crashes" in the populations of Kaplan's. In a boom period, Kaplan's program invades and overtakes a population of non-Kaplan's. However, a crash begins once Kaplan's program attains a near monopoly, dying off on account of negative information externalities which cause it to switch into truthtelling mode. In the latter case, Kaplan's relative capital share will shrink until a sufficient number of active bidders are present in the market to enable Kaplan's program to remain in background sufficiently frequently to counterbalance the losses incurred in truthtelling mode. We conducted evolutionary tournaments with noise traders and found that only a small fraction of active bidders—5 or 10 percent—is necessary to achieve stability in capital shares. However, there is little evidence that any of the other programs can successfully invade a population of Kaplan's that are operating in background mode. Otherwise we would have observed growth in their capital shares as Kaplan's program started

[40] Ringuette apparently also intended to return a bid from skeleton whenever there were fewer than 1/8 of the total number of steps remaining in the game; however, due to an apparent programming error, this option is never invoked.

to monopolize the market. Except for a slight upturn in Anderson's capital share in Figure 6, there is no evidence that this is happening.[41] By remaining in the background, Kaplan's program is able to capitalize on the mistakes of the active bidders, lifting its efficiency well above 100% and ensuring its growth while pushing the active bidders' efficiency well below 100% and ensuring their decline. The biological analogy is that Kaplan's program is a parasite that invades and eventually destroys its host. In the absence of an active bidding strategy that can successfully invade (or at least coexist) with a population of Kaplan's, there would appear to be no mechanism to stabilize the resulting cycles of booms and crashes. However, given the passive nature of a "wait in the background" strategy, it is difficult to see how one could exploit it. Since this strategy is essentially parasitic, we might pose the key open question in biological terms: are there strategies and environments that are resistant to invasion by Kaplanites?

5. CONCLUSIONS

In this paper we have studied the behavior of a collection of computer programs playing the roles of buyers and sellers in a discretized version of a dynamic double auction market. One of our objectives was to use this market in an attempt to understand the operation of the "invisible hand." We found that despite the decentralized nature of the trading process and traders' incomplete information about supply and demand, the transaction-price trajectories of a heterogeneous collection of computer programs typically converged to the competitive equilibrium, resulting in allocations that were nearly 100% efficient. Our findings complement and extend previous theoretical and experimental insights by Easley and Ledyard,[9] Friedman,[11] Gode and Sunder,[15] and Wilson,[41] taking us one step closer to resolving Hayek's problem, namely to "show how a solution is produced by the interactions of people each of whom possesses only partial knowledge." Specifically, the fact that convergence occurs in markets where traders use simple rules-of-thumb suggests that it is the DA *institution*, rather than the rationality of the traders *per se*, that is responsible for the emergence of competitive outcomes. The DA trading rules, particularly the "New York Rules" governing the improvement of standing bids and offers, appear to act as a "funnel" that guides the uncoordinated actions of a heterogeneous collection of decision rules towards the CE. Other institutional features, including the discrete vs. continuous nature of trading process, do not appear to play a significant role in generating competitive outcomes. In particular, our imposition of the "AURORA rules" restricting which traders are eligible to accept the standing bid

[41]The upturn in Anderson's share of sellers' capital appears to coincide with Kaplan's monopolization of the buyers' market. Thus, the upturn seems more likely to reflect Anderson's ability to capture some of windfall gains provided by Kaplan's buyers when they entered truthtelling mode than Anderson's ability to successfully invade a population of Kaplan's.

or ask does not appear to impose a significant constraint on trading opportunities or prevent the market from converging to CE.

Our second objective was to compare the behavior of human and automata traders. Overall, we found that the top-ranked trading programs appear to yield a "realistic" working model of a DA market in the sense that their collective behavior is consistent with the key "stylized facts" observed in human DA experiments: (1) convergence to CE, (2) high *ex post* efficiency, (3) reductions in transaction-price volatility and efficiency losses in successive trading periods reflective of apparent "learning" effects, (4) coexistence of extra-marginal and intra-marginal efficiency losses, (5) low-rank correlations between the realized order of transactions and the "efficient" order, and (6) negatively autocorrelated transaction-price changes. More detailed statistical comparisons of human and computer traders will have to await the completion of matching human experiments to be conducted in our discretized DA market. The fact that our collection of program traders seem to behave similarly to human traders may not seem surprising if programmers are merely encoding their "market intuition" into their computer programs. However, given the complexity of the DA environment and the sophistication of human intelligence, it is not obvious that human behavior in these markets can be captured by a few simple decision rules.

Our final objective was to characterize the form of effective trading strategies in DA markets. We studied a collection of over 30 computer programs ranging in complexity from simple rules-of-thumb to sophisticated adaptive/learning procedures employing some of the latest ideas from the literature on artificial intelligence and cognitive science. In order to evaluate the programs, we conducted an extensive series of computer tournaments involving hundreds of thousands of individual DA games, covering a wide range of trading environments and compositions of trading partners. To our surprise, a single program emerged as the clear winner in nearly all of the tournaments and trading environments. The winning program, submitted by economist Todd Kaplan of the University of Minnesota, was one of the simplest programs that we studied, and can be characterized as *nonadaptive, non-predictive, non-stochastic,* and *non-optimizing.* The basic idea behind the program is to *wait in the background to let others do the negotiating, but when bid and ask get sufficiently close, jump in* and *"steal the deal."* The program makes no use of prior information about the joint distribution of token values, and relies on only a few key variables such as its privately assigned token values, the current bid and ask, its number of remaining tokens, and the time remaining in the current period. The fact that one can design an effective trading program relying on only a few sufficient statistics confirms Hayek's observation about "the remarkable economy of knowledge that is required in order to take the right action in a competitive market."

It appears that the success of Kaplan's strategy is due to the fact that in an efficient market, if the current bid and ask are close, then it is likely to be the case that either (1) bid and ask are close to the equilibrium price interval, or (2) the current bid or ask are close as a result of a mistake in which one of the holders' failed to place their bid or ask at a sufficiently favorable price. Kaplan's program attempts to "steal the deal" by placing a bid equal to the previous asking price, but only if it can make a profit at that price. As a result Kaplan's program tends to earn at least a normal profit if case (1) holds, and a supernormal profit if another trader has made a mistake. Since the decision of how much to bid is much more difficult than the binary buy/sell decision, it is not surprising that mistakes in bidding are a primary source of poor trading performance. By staying out of the bidding game, Kaplan's program is able to avoid making bidding mistakes on its own account while capitalizing on bidding mistakes of others.

Another reason for the relatively poor performance of the complex, adaptive, optimizing, and predictive strategies is the inherent difficulty of making accurate inferences in a noisy marketplace given only a limited number of observations on one's opponents. The randomness in traders' token endowments is the dominant source of uncertainty in any particular DA game. The additional variation in profits induced by mistakes or stochastic elements in the trading strategies is insignificant in comparison. As a result, one needs a very large number of observations on trading outcomes to be able to reliably distinguish good traders from bad. It follows that it is virtually impossible to try to recognize and exploit the idiosyncrasies of one's individual trading partners unless one is interacting with the same group over a very long horizon. The low signal/noise ratio of realized trading profits combined with the high dimensionality of the space of possible trading histories and trading environments implies that programs based on general learning principles (such as neural networks and genetic algorithms) require many thousands of DA training games before they are able to trade even semi-effectively.[42] Nearly all of the top-ranked programs were based on a fixed set of intuitive rules-of-thumb that encoded the programmer's prior knowledge of trading process. This finding suggests that our hopes of using computerized agents endowed with general principles of artificial intelligence to evaluate alternative institutional designs may be too ambitious.

Given the simplicity, robustness, and effectiveness of the "wait in the background" strategy, it seems likely other traders would attempt to imitate it, leading to growth in the relative numbers of these sorts of background traders. On the other hand, less profitable traders should gradually exit the market due to competitive pressures. In order to study the long-run equilibrium of such market, we conducted an "evolutionary tournament" in which the fraction of each type of trader was proportional to its share of the total capital stock. The capital of each trader was updated after each DA game, increasing or decreasing by the difference between realized profits and the trader's surplus allocation in the game. Thus, the capital stocks and relative numbers of each type of trader grew or shrank depending on whether it traded at greater than or less than 100% efficiency. Starting from equal initial capital endowments, the background traders succeeded in exploiting and driving out the active bidders, nearly monopolizing the market. However, the background traders create a negative "information externality" by waiting for their

[42]To quote from the entry by Dallaway and Harvey: "Given that we are doing the equivalent of evolving monkeys that can type Hamlet, we think the monkeys have reached the stage where they recognize that they should not eat the typewriter. If we could have a four billion year time extension before handing in the entry, we are completely confident of winning."

opponents to make the first move. If all traders do this, little information will be generated and the market would be unable to function efficiently. In order to avoid such a deadlock, Kaplan's program defaults to a "truthtelling" bidding strategy if a sufficiently long time has elapsed since a trade has occurred. Although a collection of 100% truthtellers is necessarily 100% efficient, it can be easily exploited by even slightly more sophisticated strategies.

The long-run stability of this market depends on whether it is open or closed to new entrants. A closed market tends to be unstable, exhibiting cycles of booms and crashes in the population of background traders. In a boom period the background traders invade and overtake a population of active bidders. A crash begins when the background traders achieve a near monopoly, because the negative information externalities that they create cause Kaplan's program to switch into truthtelling mode. Only in a knife-edge case where Kaplan's traders are able to simultaneously monopolize both sides of the market is a stable equilibrium achieved with 100% truthtelling. But the background traders will typically succeed in monopolizing one side of the market before the other, resulting in its precipitous decline as a result of systematic exploitation by the background traders on the other side of the market.

However, in an open market, active bidding by a steady flow of short-lived noise traders succeeds in stabilizing the pattern of booms and crashes in the number of Kaplan traders. Only a small fraction of noise traders, comprising less than 10% of the market, is necessary to keep Kaplan's traders in background mode sufficiently frequently to counterbalance the losses they incur in truthtelling mode. Kaplan's traders make up at least 90% of the market in the long run since the growth of a competitive fringe causes Kaplan's traders to shift into background mode, exploiting and eventually halting the growth of the fringe.

Although the noise traders facilitate long-run stability in market shares, the limiting market is still quite unstable in other respects. In particular, transaction-price volatility is unrealistically high—a consequence of the fact that the Kaplan traders are frequently in truthtelling mode. It is unlikely that this situation could persist in the presence of truly adaptive traders, since they would eventually discover best replies that exploit the fact that Kaplan's program eventually switches into truthtelling mode. Since the trading programs submitted to our initial DA tournament were designed to do well in a sequence of short-run encounters with heterogeneous opponents rather than in long-run interactions with homogeneous opponents, it is not surprising that none were successful in exploiting this particular idiosyncrasy.

The open question is whether strategies exist that are capable of dominating Kaplan's "wait in the background" strategy over a nontrivial range of environments. If we were to run another DA tournament, it seems likely that entrants would attempt to beat Kaplan by developing more sophisticated delay and "endgame" strategies rather than reverting to truthtelling mode after a fixed amount of time. Thus, even though Wilson's WGDA equilibrium appears to be inconsistent with the behavior of humans and computer programs, we view our results as confirming the importance of his insight on the role of delay as a key ingredient of an effective trading strategy. The main complicating factor is that traders generally don't have any prior knowledge about the strategies used by their opponents, and it may be very difficult to learn those strategies unless one is interacting with the same group for a very long period of time. If this is the case, then simple rules-of-thumb such as Kaplan's may enable one to capture the key features of an effective strategy in a anonymous market consisting of short-run encounters with heterogeneous (and impatient) opponents, whereas more complicated adaptive/learning procedures may do better in a market where one repeatedly trades over a long period of time with a fixed set of opponents. In future work we plan to investigate whether one can develop hybrid rules that graft adaptive/learning procedures onto simple, effective rules-of-thumb, using the rule-of-thumb as a fall-back, but creating the possibility that it might be improved in light of trading experience.[43] Our hope is to characterize strategies that are undominated over a broad range of environments and consistent with long-run market stability.

ACKNOWLEDGMENTS

We are grateful for the generosity of the Santa Fe Institute and its Economics program for providing the facilities, salary, and administrative and computational support that made this research possible. John Rust is grateful to the Alfred Sloan Foundation for providing tournament prize money, and National Science Foundation grants SES-8721199 for computer hardware and SES-9010046 for funding of matching human experiments at the Experimental Science Laboratory at the University of Arizona (Vernon Smith, Co-PI). International Business Machines provided funding for tournament organizational expenses. John Miller would like to acknowledge an equipment grant from Sun Microsystems. The general idea of the computerized double auction tournament emerged from discussions in a March 1988 meeting at the Sante Fe Institute, including Phil Anderson, Ken Arrow, Brian Arthur, John Holland, Tim Kehoe, Richard Palmer, John Rust, Tom Sargent, and Eugenia Singer. We also acknowledge helpful discussions with Robert Axelrod, Charles Plott, Vernon Smith, and Shyam Sunder, and extend our thanks to particular individuals whose expertise made this tournament possible, including Michael Angerman, Marcella Austin, Ronda Butler-Villa, Steven Pope, Ginger Richardson, Dan Schneidewend, Andi Sutherland, George Tsibouris, and Della Ulibarri. We are grateful for helpful comments from Dan Friedman, Vernon Smith, and seminar participants at the Universities of Oslo and Stockholm, and the London School of Economics. Finally we would like to thank all participants of the DA tournament for their willingness to commit time and effort in developing an ingenious collection of trading programs.

[43] For example, if a learning rule were "grafted" onto Kaplan's strategy, it might be smart enough to recognize that other traders were taking advantage of its truthtelling mode, adjusting its delay and bidding rules to insure its long-run survival.

APPENDIX: EXPLANATION OF THE FOUR-WAY BREAKDOWN OF EFFICIENCY LOSSES

Efficiency losses in any DA game can be decomposed into the following four categories:

IM: value of lost surplus of non-traded intra-marginal tokens (i.e., those that lie to the left of the equilibrium quantity, q^*) when the actual number of trades q is less than q^* (or 0 otherwise).

EM: value of lost surplus due to trade of extra-marginal tokens (i.e., those that lie to the right of q^* on the supply and demand curves) when the actual number of trades q is greater than q^* (or 0 otherwise).

BS: value of lost surplus due to trades of extra-marginal buyers' tokens that displaced potential trades of an equal number of buyers' intra-marginal tokens.

SS: value of lost surplus due to trades of extra-marginal seller's tokens that displaced potential trades of an equal number of sellers' intra-marginal tokens.

In order to define these quantities, we first define IMB and IMS as the sum of the values of all intra-marginal tokens *not traded* by buyers and sellers, respectively, during the period. Let EMB and EMS denote the sum of the values of all extra-marginal tokens *traded* by buyers and sellers during the period. Then we have:

$$\text{LOST SURPLUS} = \text{SURPLUS} - \text{PROFIT} = \text{IMB} - \text{EMB} + \text{EMS} - \text{IMS}. \quad (1)$$

Let NIMB and NIMS denote the number of intra-marginal tokens that failed to be traded by buyers and sellers, respectively. Let NEMB and NEMS denote the number of extra-marginal tokens traded for the two respective sides of the market. Then we also have the identity

$$q - q^* = \text{NEMB} - \text{NIMB} = \text{NEMS} - \text{NIMS}. \quad (2)$$

Clearly when $q = q^*$, there are no *net* intra-marginal or extra-marginal trades, so that $EM = IM = 0$. Then the total amount of lost surplus can then be unambiguously divided into the two categories BS $=$ IMB $-$ EMB and SS $=$ EMS $-$ IMS. However if $q > q^*$, then we face the problem of how to allocate lost surplus due to trades of extra-marginal tokens among the three categories EM, BS, and SS, and if $q < q^*$, we face a similar problem of allocating lost surplus among the categories IM, BS, and SS. Define NBS, NSS, NEM and NIM as follows:

$$\begin{aligned}\text{NEM} &= \max(q - q^*, 0),\\ \text{NIM} &= \max(q^* - q, 0),\\ \text{NBS} &= \min(\text{NIMB}, \text{NEMB}),\\ \text{NSS} &= \min(\text{NIMS}, \text{NEMS}).\end{aligned} \quad (3)$$

Then it is easy to see that the following identities hold:

$$\begin{aligned} \text{NIMB} &= \text{NIM} + \text{NBS}, \\ \text{NEMB} &= \text{NEM} + \text{NBS}, \\ \text{NIMS} &= \text{NIM} + \text{NSS}, \\ \text{NEMS} &= \text{NEM} + \text{NSS}. \end{aligned} \tag{4}$$

Definition (3) provides an unambiguous way of decomposing the total number of inefficient trades into the four categories BS, SS, EM, and IM. There is no unambiguous way of decomposing the value of lost surplus, however. For example, in the DA game illustrated in Figure 1, buyer B2 traded two extra-marginal tokens, and there are three intra-marginal tokens that B1 and B2 failed to trade, and one intra-marginal token that S3 failed to trade. Thus, NEM=0, NIM=1, NBS=2, and NSS=0, which implies that EM=0 and SS=0. To compute the value of BS and IM, we need to determine which of the three intra-marginal buyer's tokens were "bumped." We assume that any one of these tokens is equally likely to have been bumped, and thus we value each "bumped" buyer's token at IMB/3 and compute BS as the value of B3's two extra-marginal tokens less 2/3IMS. IM is computed as IMB/3 less the value of S3's untraded token. More generally EM, IM, BS, and SS can be defined as follows:

$$\begin{aligned} \text{EM} &= \text{NEM}\,(\overline{\text{EMS}} - \overline{\text{EMB}}), \\ \text{IM} &= \text{NIM}\,(\overline{\text{IMB}} - \overline{\text{IMS}}), \\ \text{BS} &= \text{NBS}\,(\overline{\text{IMB}} - \overline{\text{EMB}}), \\ \text{SS} &= \text{NSS}\,(\overline{\text{EMS}} - \overline{\text{IMS}}), \end{aligned} \tag{5}$$

where $\overline{\text{EMS}} \equiv \text{EMS}/\max(\text{NEMS}, 1)$ is the average value of extra-marginal sellers' tokens, and $\overline{\text{IMS}}$, $\overline{\text{EMB}}$, and $\overline{\text{IMB}}$ are defined similarly. Using identities (1) through (4), it is easy to verify that the definitions of EM, IM, BS, and SS insure that they are always nonnegative and that the following identity holds:

$$\text{LOST SURPLUS} = \text{EM} + \text{IM} + \text{BS} + \text{SS}. \tag{6}$$

REFERENCES

1. Abreu, D., and A. Rubinstein. "Finite Automata Play the Repeated Prisoner's Dilemma." *Econometrica* **57(4)** (1988): 345–352.
2. Andreoni, J., and J. H. Miller. "Auctions with Adaptive Artificial Agents." Working Paper 90-01-004, Santa Fe Institute, 1990.
3. Axelrod, R. *The Evolution of Cooperation*. New York: Basic Books, 1984.
4. Blum, M. "A Machine-Independent Theory of the Complexity of the Recursive Functions." *J. ACM* **14(2)** (1967): 322–336.
5. Bollerslev, T., and I. Domowitz. "Some Effects of Restricting the Electronic Order Book in an Automated Trade Execution System." This volume, chapter 8.
6. Cason, T. N., and D. Friedman. "An Empirical Analysis of Price Formation in Double Auction Markets." This volume.
7. Clauser, L., and C. Plott. "On the Anatomy of the 'Nonfacilitating' Features of the Double Auction Institution in Conspiratorial Markets." This volume.
8. Diaconis, P., and D. Freedman. "The Inconsistency of Non-Parametric Bayesian Analysis." *Annals of Statistics* **14** (1986): 1–26.
9. Easley, D., and J. Ledyard. "Theories of Price Formation and Exchange in Double Oral Auctions." This volume.
10. Freedman, D. "On the Inconsistency of Bayes Estimates in Infinite-Dimensional Spaces." *Annals of Statistics* **34** (1963): 1386–1403.
11. Friedman, D. "A Simple Testable Model of Double Auction Markets." *J. Econ. Behav. Org.* (1991): 47–50.
12. Friedman, D. "The Double Auction Institution: A Survey." This volume.
13. Garcia, C. B. "On the Dynamics of Prices to Equilibrium in a Double Auction Market." Manuscript, University of Chicago Graduate School of Business, 1980.
14. Gode, D. K., and S. Sunder. "Allocative Efficiency of Markets with Zero Intelligence (ZI) Traders: Market as a Partial Substitute for Individual Rationality." *J. Pol. Econ.* (1991): forthcoming.
15. Gode, D. K., and S. Sunder. "Lower Bounds for Efficiency in Double Auctions." This volume.
16. Hayek, F. "The Use of Knowledge in Society." *Amer. Econ. Rev.* **35(4)** (1945): 519–530.
17. Hertz, J. A., A. S. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley, 1991.
18. Holland, J. *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press, 1975.
19. Kagel, J. H., and W. B. Vogt. "Buyer's Bid Double Auctions: Preliminary Experimental Results." This volume.
20. Ledyard, J. "The Scope of the Hypothesis of Bayesian Equilibrium." *J. Econ. Theor.* **39** (1986): 59–82.

21. Marimon, R., E. McGratten, and T. Sargent. "Money as Medium of Exchange in an Economy with Artificially Intelligent Agents." *Comp. Sci. Econ. Mgmt.* **5** (1990): 109–123.
22. Maynard Smith, J. *Evolution and the Theory of Games*. Cambridge, MA: Cambridge University Press, 1982.
23. McCabe, K., S. Rassenti, and V. Smith. "Lakatos and Experimental Economics." Manuscript, University of Arizona, 1989.
24. Myerson, R. "Optimal Auction Design." *Mathematics of Operations Research* **6** (1981): 58–73.
25. Myerson, R., and M. A. Satterthwaite. "Efficient Mechanisms for Bilateral Trade." *J. Econ. Theor.* **29**: 265–281.
26. Narendra, K. S. *Adaptive and Learning Systems: Theory and Applications*. New York: Plenum Press, 1986.
27. Palmer, R. G., J. Rust, and J. H. Miller. "Double Auction Tournament Participant's Manual." Santa Fe Institute, 1990.
28. Roberts, J., and A. Postlewaite. "The Incentives for Price-Taking Behavior in Large Exchange Economies." *Econometrica* **44** (1976): 115–128.
29. Rust, J., J. Miller, and R. Palmer. "Characterizing Effective Trading Strategies: Insights from a Computerized Double Auction Tournament." *J. Econ. Dynamics and Control* (1993): forthcoming.
30. Rustichini, A., M. A. Satterthwaite, and S. R. Williams. "Convergence to Price-Taking Behavior in a Simple Market." Discussion Paper 914, Center for Mathematical Studies in Economics and Management Science, Northwestern University, Evanston, IL, 1990.
31. Satterthwaite, M. A., and S. R. Williams. "Bilateral Trade with the Sealed-Bid k-Double Auction: Existence and Efficiency." *J. Econ. Theor.* **48** (1989): 107–133.
32. Satterthwaite, M. A., and S. R. Williams. "The Rate of Convergence to Efficiency in the the Buyer's Bid Double Auction as the Market Becomes Large." *Rev. Econ. Studies* **56** (1989): 477–498.
33. Satterthwaite, M. A., and S. R. Williams. "The Bayesian Theory of the *k*-Double Auction." This volume.
34. Selten, R., M. Mitzkewitz, and G. Uhlich. "Duopoly Strategies Programmed by Experienced Players." Manuscript, Bonn University, Germany, 1990.
35. Smith, V. L., A. Williams, K. Bratton, and M. Vannoni. "Competitive Market Institutions: Double Auctions Versus Sealed Bid-Offer Auctions." *Amer. Econ. Rev.* **72** (1982): 58–77.
36. Smith, V. "Experimental Econmics: Induced Value Theory." *Amer. Econ. Rev.* **66(2)** (1976): 274–279.
37. Smith, V. "Relevance of Laboratory Experiments to Testing Resource Allocation Theory." In *Evaluation of Econometric Models*, edited by J. Kmenta and J. R. Ramsey, 345–377. New York: Academic Press, 1980.
38. Wheeler, R. M. Jr., and K. S. Narendra. "Decentralized Learning in Finite Markov Chains." *IEEE Transactions on Automatic Control* **AC–31–6** (1986): 519–526.

39. Williams, S. R. "Existence and Convergence of Equilibria in the Buyer's Bid Double Auction." *Rev. Econ. Studies* **58(2)** (1990): 351–374.
40. Wilson, R. B. "Incentive Efficiency of Double Auctions." *Econometrica* **53-5** (1985): 1101–1116.
41. Wilson, R. B. "On Equilibria of Bid-Ask Markets." In *Arrow and the Ascent of Modern Economic Theory*, edited by G. Feiwell, 375–414. New York: MacMillan, 1987.

Dhananjay K. Gode and Shyam Sunder
Graduate School of Industrial Administration, Carnegie-Mellon University, Schenley Park, Pittsburgh PA 15213

SEVEN
Lower Bounds for Efficiency of Surplus Extraction in Double Auctions

The double auction is a robust institution for efficient extraction of consumer and producer surplus in a variety of market environments. This conclusion is based on evidence gathered in a large number of laboratory markets populated by profit-motivated human traders, usually students (see Smith,[14] Plott,[8] and Friedman[4] in this volume for literature reviews). The efficiency of surplus extraction in double auctions may derive from the characteristics of this institution, from characteristics of trader behavior, or perhaps from interactions between the two. Virtually all economic experiments of the past have focused on the effect of varying economic institutions or environments on the performance of the market, holding profit-motivated behavior of traders unchanged. Since profit maximization is a maintained hypothesis in this literature, there has been an unchallenged inclination that it plays an important role in the tendency of double auctions to extract most of the consumer and producer surplus.

Gode and Sunder[5] substituted profit-motivated human traders by "budget-constrained zero-intelligence machine traders" in their double auction experiments. These traders are simple computer programs that generate random bids (or asks) subject to a no-loss constraint. The focus of that experiment was on examining the effect of controlled variation in trader behavior on the efficiency of the market. They found that imposition of a no-loss constraint (prohibiting traders from buying above their redemption values or selling below their cost) is sufficient to attain over 98% efficiency, even if traders, stripped of all rationality, submit random bids and asks. These results suggest that surplus extraction may largely be a property of the double auction institution, independent of the trader behavior.

This paper is an attempt to determine the lower bounds for expected surplus extraction efficiency of an idealized double auction populated with budget-constrained zero-intelligence (ZI) traders. The ultimate goal of this effort is to gain insights into the factors responsible for the high efficiency of some markets such as double auctions. The double auction appears to be too complex a game to yield a clear game-theoretic solution; its properties are easier to analyze with traders who act randomly.[1] We use the technique of using zero-intelligence traders to map the sensitivity of the lower bound of the expected efficiency to market parameters, continuously clearing procedures, relative number of extra-marginal traders, and rounds of bidding. The worst-case expected efficiency turns out to be 81% for "synchronized" double auctions and 75% for "continuously clearing" double auctions (these variations of double auction will be defined in the paper). The expected surplus extraction efficiency of continuous markets is lower because they permit a higher chance for extra-marginal traders to displace the intra-marginal traders. As the relative proportion of extra-marginal traders declines, expected efficiency of double auction converges to the close neighborhood of 100%. Some 50–100% of surplus is extracted in the first round of bidding in a synchronized double auction, declining sharply in subsequent rounds.

These analytical and simulation results confirm, and provide a better understanding of, the empirical and simulation results presented in Gode and Sunder.[5] A key insight is the crucial role of the tradeoff between the probability of an efficiency-reducing transaction and the magnitude of the resultant efficiency reduction in defining the environment for which the expected efficiency of double auction attains its lower bound. Second, the extra-marginal shapes of supply and demand functions affect the expected efficiency of double auction through their effect on both the probability as well as the magnitude of efficiency reduction, even though the theoretical equilibrium prediction is independent of the location of extra-marginal units.

The first section of the paper specifies a simple environment (i.e., demand and supply), trader behavior, and an idealized form of the institution labelled "synchronized" double auction. Expected efficiency of this auction with zero-intelligence traders is derived in Section 2. Section 3 examines the sensitivity of this result to the relative proportion of extra-marginal traders using computer simulations. Section 4 derives the efficiency results for a "continuously clearing" version of double auction which is a closer approximation of field and laboratory versions of this institution. Section 5 maps the time profile of surplus extraction in a synchronized double auction. Finally, Section 6 summarizes the results, and discusses its implications for the source of efficiency of double auctions.

[1]See Easley and Ledyard[2] for another alternative approach to study of double auctions. They specify boundedly rational, but non-strategic rules of trader behavior to examine the institutional properties.

THE MODEL

DEMAND AND SUPPLY

Each buyer has the right to buy up to one unit that has a given redemption value between 0 and 1. In the base case we consider first, there is one buyer with redemption value 1 and an infinite number of buyers with redemption value of β ($0 < \beta < 1$). Similarly, each seller has the right to sell up to one unit with a given variable cost between 0 and 1. There is one seller with cost of 0 and an infinite number of sellers with cost $1 - \alpha$ ($0 < \alpha < 1$). Figure 1 shows the resulting demand and supply functions, an equilibrium price range from β to $(1 - \alpha)$, and an equilibrium quantity of 1.

Two features of the demand and supply configuration we have chosen deserve comment. The presence of only one intra-marginal trader on each side of the market precludes the possibility of competition within the intra-marginal traders on each

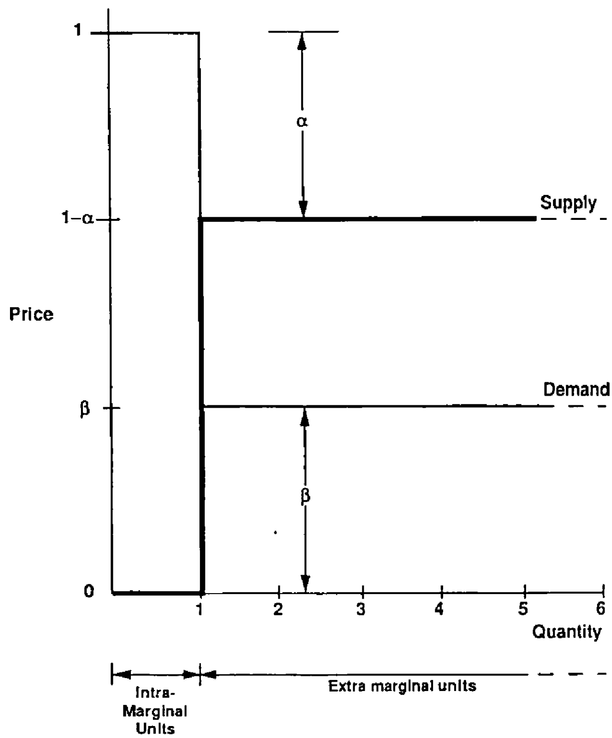


FIGURE 1 Demand and Supply Functions.

side. Competition amongst the intra-marginal buyers (sellers) only serves to exclude the extra-marginal buyers (sellers) from the market by raising the bids above (lowering the asks below) the redemption values (costs) of the extra-marginal buyers (sellers) more quickly. Thus competition among intra-marginal traders lowers the chances of displacement of intra-marginal traders by extra-marginal traders. Since such displacement is the only source of inefficiency in markets with ZI traders, we have used only one intra-marginal trader on each side to arrive at the lower bounds for efficiency. Efficiencies with intra-marginal competition would be higher.

The use of flat supply and demand functions in the extra-marginal region makes it easier to examine the effect of changing the costs (redemption values) of extra-marginal sellers (buyers) on efficiency. Presence of extra-marginal units at multiple redemption values or costs adds complexity without further insights. Finally, the use of a single intra-marginal trader and multiple extra-marginal traders at the same value or cost allows us to study the impact of changing the number of extra-marginal traders relative to the number of intra-marginal traders.

"ZERO INTELLIGENCE" TRADERS

Whenever a buyer has an opportunity to make a bid (to be defined by market rules discussed below), it generates random bids distributed independently and uniformly between 0 and the redemption value of its current unit. The imposition of an upper limit of redemption value on bids amounts to a budget constraint and prevents buyers from buying things they cannot afford to pay for. Similarly a seller generates (whenever it has an opportunity to do so) random offers distributed independently and uniformly between the cost of its current unit and 1. This ensures that the traders would not incur a loss. The lower limit of cost imposed on sellers' asks also has the effect of imposing a similar budget constraint on sellers. The support of the bids/asks generated by these traders does not change in response to the level of highest outstanding bid (current bid) or lowest outstanding ask (current ask).

MARKET RULES FOR SYNCHRONIZED DOUBLE AUCTION

In the base case we assume that the market operates as a *synchronized double auction*.[2] All buyers and sellers are simultaneously solicited for bids and asks until each provides a bid or ask which does not violate its budget (i.e., the no-loss) constraint. The highest bid and the lowest ask are designated as the current bid and the current ask, respectively. If the current bid and the current ask cross, these units are traded in a binding transaction.[3] If they don't cross, solicitation from all

[2] See Rust, Miller, and Palmer,[10] in this volume, for an implementation of the synchronized double auction of which this is an idealized description.

[3] The price at which such transactions are booked can, depending on the market rules chosen, lie anywhere in the range between the crossed bid and ask. Since we are not concerned with the behavior of prices in this article, we leave this market rule unspecified.

traders is repeated. The bid/ask improvement rule is applied to calculate the new current bid/ask. This means that the current bid can be updated in a later round only by a higher subsequent bid and the current ask can be updated only by a lower subsequent ask. A transaction cancels all unaccepted bids and asks. This process is repeated until expiration of the prespecified time allowed for the period.

This particular idealization of double auction is labelled "synchronized" because in each round of bid/ask solicitation every trader's bid or ask is on the board before the highest bids and the lowest asks are allowed to close a transaction. In a later section of the paper, we analyze "continuously clearing" double auctions that bear greater resemblance to laboratory and field institutions.

EXPECTED EFFICIENCY

Expected efficiency of a synchronized double auction populated by budget-constrained zero-intelligence traders (with one intra-marginal buyer and seller each, and an infinity of extra-marginal buyers and sellers) is given by:

$$\begin{array}{ll} 1-(a-\alpha-\beta)\left(\dfrac{\alpha^2}{1-\alpha(1-\alpha)}+\dfrac{\beta^2}{1-\beta(1-\beta)}\right) & \text{for} \quad \alpha+\beta\leq 1\,, \\ 1 & \text{for} \quad \alpha+\beta>1\,. \end{array} \tag{1}$$

(See Appendix for Proof.)

The upper panel of Figure 2 shows expected efficiency as a function of the redemption values of extra-marginal buyers (β) and costs of extra-marginal sellers ($1-\alpha$). In the lower panel, the vertical scale has been expanded to show clearly the shape of the surface. Efficiency is 100% if $\alpha=\beta=0$ or if $\alpha+\beta\geq 1$. When $\alpha=\beta=0$, the extra-marginal buyers' costs are equal to the values of intra-marginal sellers, and the value of extra-marginal buyers' costs are equal to the values of intra-marginal sellers, making it impossible for the extra-marginal traders to enter the market. Given a sufficient number of rounds to submit bids and asks, intra-marginal traders transact their respective units to yield 100% efficiency.

If the demand and supply do not intersect in this market (i.e., $\alpha+\beta\geq 1$), all surplus is necessarily extracted in this market. Efficiency-reducing transactions can take place only when the cost of extra-marginal sellers exceeds the redemption value of extra-marginal buyers (i.e., $\alpha+\beta<1$).

The minimum possible value of expected efficiency in α,β-plane is 80.84%, attained at two points—($\alpha=0,\beta=0.639$) and ($\alpha=0.639,\beta=0$). The minimum is driven by two considerations. Usually, extra-marginal units cannot displace intra-marginal units because the budget constraint prevents these ZI traders from submitting bids that are high enough (or asks that are low enough) to transact. However, variability of transaction prices means that some transactions do take place at prices which are accessible to the extra-marginal traders. The extra-marginal units closer to the equilibrium price have a greater chance of displacing the intra-marginal units because bids are bounded above by the demand function (and asks are bounded below by the supply function) due to imposition of the budget constraint. When such displacement does occur, the units closer to the equilibrium price cause only a small loss of surplus extracted. Therefore, the *expected* loss of surplus due to displacement of intra-marginal units by extra-marginal units close to equilibrium is relatively small. On the other hand, extra-marginal units far away from the equilibrium price can have a big impact on surplus extracted whenever they are able to displace intra-marginal units in trading. But the chances of such displacement become increasingly remote as the distance of such units from the equilibrium price increases. Again, these units, too, have little effect on expected efficiency of the double auction. The maximum reduction in expected efficiency derives from intermediate units with a moderate effect on the magnitude of surplus extraction and only a moderate chance of displacing the intra-marginal units.

This minimum expected efficiency of 80.84% understates the expected efficiency one may expect to observe in a synchronized double auction on average. If we assume that parameters α and β for a particular auction are realizations of random variables drawn from uniform and independent distributions between 0 and 1, every point of the surface shown in Figure 1 would be equally likely. The average height of this surface and, therefore, the average expected efficiency over α, β-plane is 95.8%. This efficiency is only a few percentage points below the efficiencies observed in double auctions populated with profit-motivated human traders. This result seems to support Gode and Sunder's[5] conjecture that extraction of virtually all the surplus is a characteristic of the double auction, independent of trader motivation, intelligence, or learning. If allocative efficiency is to be equated with smartness, then smartness must be attributed to this institution itself; traders need not be smart to attain high efficiency in double auctions.

SENSITIVITY TO THE NUMBER OF EXTRA-MARGINAL TRADERS

Expression (1) and Figure 2 have been derived for the extreme case of a single intra-marginal trader and an infinity of extra-marginal traders on either side of the market. Since we do not have analytical expressions for finite values of N (the number of extra-marginal traders on either side of the market), the firm dark line for $N \to \infty$ in Figure 3 has been computed from Expression (1) and represents

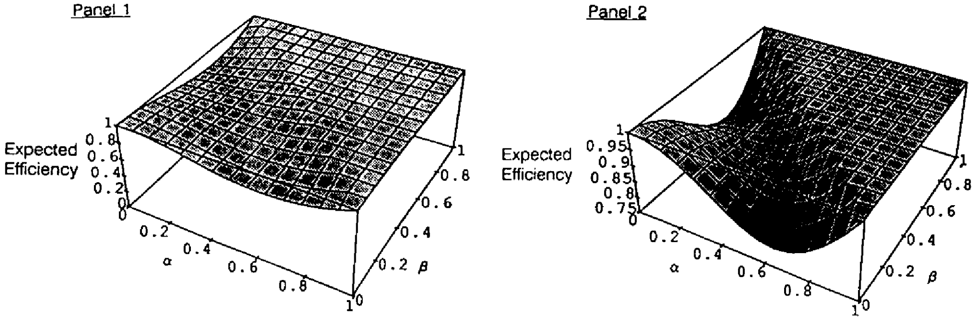


FIGURE 2 Expected efficiency of a synchronized double auction with zero-intelligence traders. Expected efficiency $= 1-(1-\alpha-\beta)(\alpha^2/(1-\alpha(1-\alpha))+\beta^2/(1-\beta(1-\beta)))$. Cost of one intra-marginal unit $= 0$. Cost of infinite extra-marginal units $= 1-\alpha$. Value of one intra-marginal unit $= 1$. Value of infinite extra-marginal units $= \beta$.

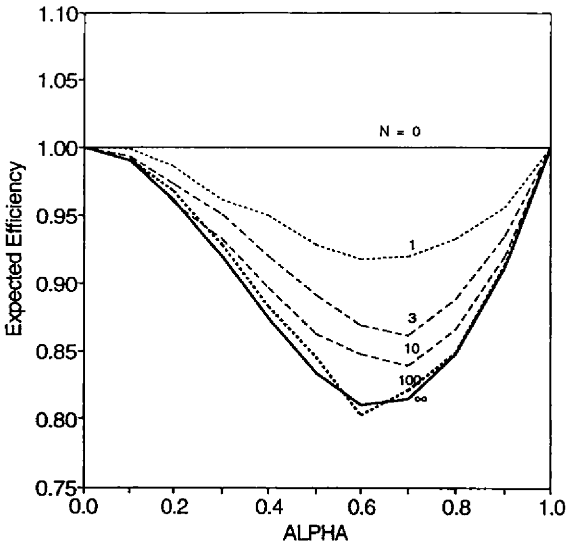


FIGURE 3 Effect of the number of extra-marginal traders on expected efficiency of synchronized double auctions (simulations with zero-intelligence traders). Value of extra-marginal buyers (β) is kept fixed at zero throughout, while the cost of extra-marginal sellers $(1-\alpha)$ is varied between 0 and 1. Expected efficiencies for $N=0$ and $N \to \infty$ have been computed; all others have been simulated.

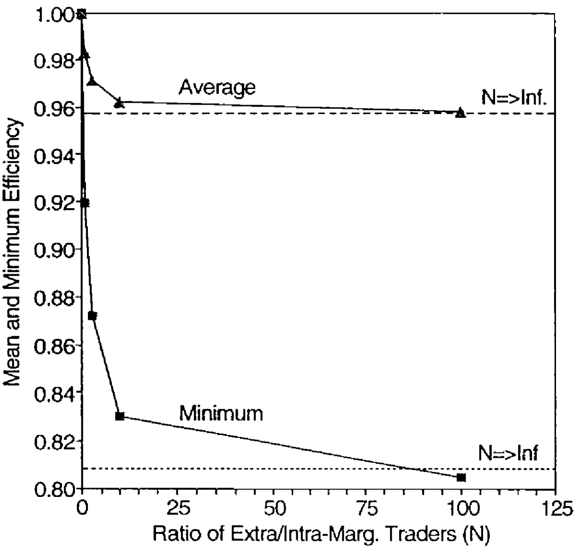


FIGURE 4 Effect of number of extra-marginal traders on minimum and mean efficiency over α, β-plane. Assuming $\alpha(=1-$ cost of extra-marginal sellers) and β (value to extra-marginal buyers) are distributed uniformly and independently between 0 and 1. Simulations with zero-intelligence traders. Efficiencies for $N = 0$ and $N \to \infty$ have been computed; all others have been simulated.

the intersection of the efficiency surface with $\beta = 0$ plane in Figure 2. For finite values of N, we don't have closed form expressions; the broken lines in Figure 3 represent average efficiencies computed from 1,000 iterations of synchronized double auctions populated by budget-constrained zero-intelligence traders for $\beta = 0$ and $\alpha = 0, 0.1, 0.2, \ldots, 1$. When there are no extra-marginal traders, it is not possible for an efficiency-reducing transaction to take place. Since intra-marginal units necessarily get traded, efficiency of a market with $N = 0$ is necessarily 100%. As the number of extra-marginal trader increases without bound, expected efficiency converges to the lower bound specified by Expression (1) and shown in Figure 2. With fewer extra-marginal traders, the chances of efficiency-reducing transactions that may involve such traders also decrease, raising the expected efficiency of the market.

Simulated synchronized double auctions show that the general shape of the expected efficiency surface with respect to α and β given in Figure 2 for $N \to \infty$ remains unchanged for smaller values of N. Minimum expected efficiency is attained when either α or β is zero, and the other parameter is between 0.6 and 0.7. The minimum level attained drops sharply when N is increased from 0 to 1; further increases in N bring further reductions of diminishing magnitudes in the minimum level of expected efficiency attained. Virtually all the drop has been attained by the time N reaches 100. This behavior of minimum expected efficiency is confirmed in the lower curve in Figure 4.[4]

Figure 4 shows the expected efficiencies of synchronized double auction averaged over the α, β-plane for various values of N. The asymptotic value of this average for $N \to \infty$ is computed by integrating Expression (1) and is shown by a dashed line. Averages for finite values of N are obtained from simulations over an α, β-grid of fineness 0.1 and are shown in solid triangles. Average expected efficiency drops at a decreasing rate from 1 (for $N = 0$) to 0.958 (for $N = 100$) with the lower bound of 0.957 as $N \to \infty$. The total loss of this average expected efficiency is no more than 4.3% in the worst possible case. The dotted line and solid squares plot the corresponding minimum expected efficiency levels attained in the α, β-plane as a function of N.

Since there is only one intra-marginal buyer and seller each in these markets, N can be interpreted as the *ratio* of extra- to intra-marginal traders.[5] For most experimental markets, this ratio is rarely greater than two or three, corresponding to mean expected efficiency (over the α, β-plane) of 97% or higher. In naturally occurring markets, larger values of this ratio could be observed. Even then, the mean expected efficiency over the parameter space cannot drop much below 96%.

CONTINUOUSLY CLEARING DOUBLE AUCTIONS

Most double auctions in the field and laboratory differ from the synchronized double auction examined above in an important respect: the timing of their bids/asks is determined by the free will of individual traders, and a transaction is completed as soon as a bid and an ask cross, without waiting for the remaining traders in the market to submit their bids/asks. How sensitive are the findings of the preceding section to this variation in the rules of a double auction market?

There are several possible ways of formally modeling a continuously clearing double auction. A precise specification of the market rules is necessary for unambiguous analysis. We use a simple specification: all traders are randomly sampled (without replacement) to submit a bid or ask. Sampling without replacement in this case means that every trader gets a chance to submit one bid/ask before any trader is able to submit a second bid/ask, and so on. A bid (ask) becomes the current or market bid (ask) if it is higher (lower) than the standing market bid (ask). As soon as the inside quotes match or cross, a transaction is executed without waiting for

[4] Note that at one point in Figure 3 ($\alpha = 0.6, \beta = 0$), the expected efficiency for $N = 100$ is less than the lower bound of expected efficiency with these parameters when $N \to \infty$. This is possible because the lower bound is a computed value from Expression (1) while the expected efficiency for $N = 100$ is a sample mean from 1000 iterations, with a positive sampling error.

[5] At this point, it is not clear if the ratio of extra- to intra-marginal traders is an appropriate parameterization of the problem. We intend to conduct further studies to find out.

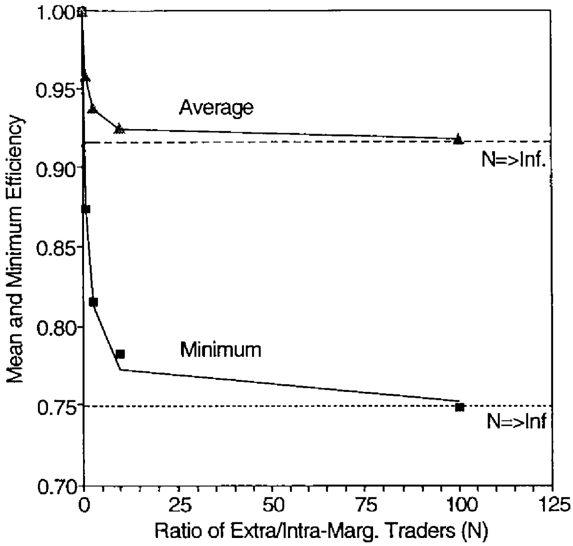


FIGURE 5 Effect of number of extra-marginal traders on minimum and mean efficiency over α, β-plane. Assuming α ($= 1-$ cost of extra-marginal sellers) and β (value to extra-marginal buyers) are distributed uniformly and independently between 0 and 1. Simulations of continuous DA with zero-intelligence traders. Efficiencies for $N = 0$ and $N \to \infty$ have been computed; all others have been simulated.

the remaining buyers(sellers) to submit their bids (asks). If a transaction does not take place in the first round of bidding, further rounds are continued until it does. As soon as a transaction is executed, all unaccepted bids and asks are cancelled (i.e., the market bid is set to zero and the market ask is set to its maximum permissible value). In this model, market outcomes depend only on the order in which a given set of bids and asks are received; the time distribution of arrivals does not matter. Since the simulations presented in this section involve identical machine traders, the order in which each trader submits its bid/ask is randomized within each round by the above procedure.

Figure 5 shows the sensitivity of minimum and mean expected efficiency of continuous auctions to the value of N from 1,000 simulations. The solid squares plot the minimum observed value of mean expected efficiency over the α, β-plane as a function of N. The solid triangles plot the observed value of mean expected efficiency averaged over the α, β-plane from 1000 iterations of continuous double auctions.

These simulation results are closely approximated by the following expression:

$$1 - \frac{N}{1+N}(\alpha + \beta)(1 - \alpha - \beta) \tag{2}$$

where N is the number of extra-marginal buyers *and* the number of extra-marginal sellers. In order to understand this approximation, it is helpful to consider the four terms of this expression in turn. If there were no surplus-reducing transactions (i.e., transactions involving the extra-marginal traders), the expected efficiency would be 1, the first term of the expression. Whenever one extra-marginal buyer buys from the intra-marginal seller (and a surplus of β is realized), the intra-marginal buyer is necessarily forced to trade with one extra-marginal seller (to realize a surplus of α). Thus a surplus-reducing transaction results in a lost surplus of $(1 - \alpha - \beta)$, the last term of the expression. $N/(N+1)$ is the probability that an intra-marginal seller will face an extra-marginal buyer before it faces an intra-marginal buyer, and β is the approximate probability that the ask submitted by these zero-intelligence traders will cross each other.[6] Similarly, $N/(N+1)$ is the probability that the intra-marginal buyer will face an extra-marginal seller before facing an intra-marginal seller, and α *is* the approximate probability that they will consummate a transaction. Subtracting the expected loss of surplus from 1 yields expression (2) for approximate expected efficiency of the continuously clearing double auction. The shape of this approximate expected efficiency surface as a function of α and β as $N \to \infty$ is shown in Figure 6. The minimum value and mean value (over the α, β-plane) of this approximate expected efficiency as functions of N are shown in Figure 5 by the respective firm lines. The approximation is sufficiently precise that most of the simulated points (dark squares and triangles) appear, in spite of their sampling errors, to have been joined together by these lines.

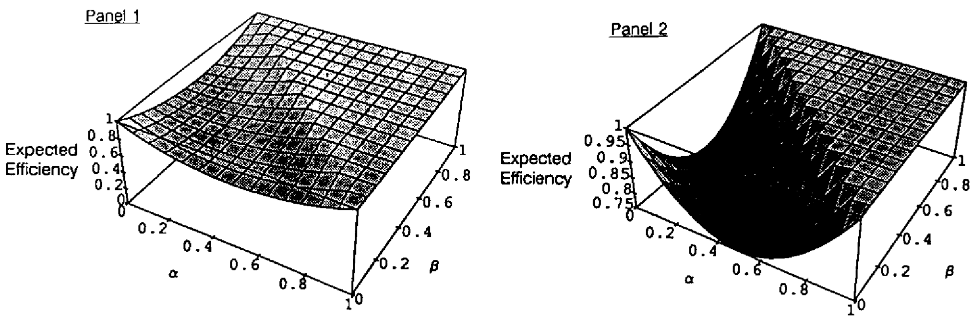


FIGURE 6 Effect of the number of continuous double auctions with zero-intelligence traders. Expected efficiency $= 1 - (N/(N+1)(\alpha + \beta)(1 - \alpha - \beta)$. Cost of one intra-marginal unit $= 0$. Cost of infinite extra-marginal units $= 1 - \alpha$. Value of one intra-marginal unit $= 1$. Value of infinite extra-marginal units $= \beta$.

[6]The reason this probability is approximate is that it ignores the consequences of no trade in the first round of bidding.

A comparison of Figure 6 with Figure 2 reveals some similarities and differences. The surface in Figure 6 is below the surface in Figure 2 everywhere. This means that the expected efficiency of synchronized double auction dominates the expected efficiency of continuous double auction at every point and on average. The maximum efficiency of 100% is retained for $\alpha = \beta = 0$ and for $\alpha + \beta \geq 1$. The minimum expected efficiency of continuous double auction is reached at all points of a line defined by $\alpha + \beta = 0.5$ (instead of just two points in Figure 2). The minimum level drops from 0.808 in Figure 2 to 0.75 in Figure 6. In interpreting this finding, we should be careful about several factors. First, these differences may narrow or disappear in markets populated by profit-motivated intelligent traders. Second, continuous double auctions may have the advantage of faster price discovery in dynamically changing markets and our present analysis does not include consideration of such factors. While the New York Stock Exchange could be thought of as a continuously clearing market, some of its new challengers are being designed as call markets somewhat similar to the synchronized double auction. Many stock markets around the world, especially those with less liquidity, operate as call markets. In a dynamic environment with new information, the price discovery role of markets may be important. It is possible that the higher static surplus extraction efficiency of the synchronized double auctions documented in this paper may be traded off against the possibly higher dynamic informational efficiency of continuous double auctions. For markets with sufficient liquidity, informational advantages of the continuous auctions may become sufficient to overcome their static surplus extraction disadvantages. The results obtained here point to these and other conjectures and directions of further investigation.

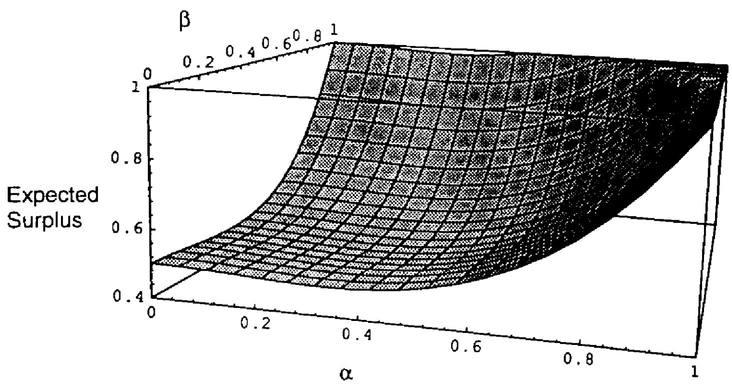


FIGURE 7 Expected surplus extraction in the first round of bidding in synchronized DA with zero-intelligence traders. Expected surplus $= 0.5 + \alpha^2(\alpha - 0.5) + \beta^2(\beta - 0.5)$. Cost of one intra-marginal unit $= 0$. Cost of infinite extra-marginal units $= 1 - \alpha$. Value of one intra-marginal unit $= 1$. Value of infinite extra-marginal units $= \beta$.

TIME PROFILE OF SURPLUS EXTRACTION

For a synchronized double auction, probabilities given in Table 1 (see the Appendix) can be used to derive the amount of surplus expected to be extracted in the first round of bidding when the number of extra-marginal traders $N \to \infty$:

$$0.5 + \alpha^2(\alpha - 0,5) + \beta^2(\beta - 0.5) \text{ for } \alpha + \beta \leq 1\,. \tag{3}$$

The shape of this function is shown in Figure 7 (only the part of the surface in the lower left side, $\alpha + \beta < 1$, is valid). Surplus extracted in the first round attains its maximum of 1 at the two corner points ($\alpha = 1, \beta = 0$) and ($\alpha = 0, \beta = 1$). The first-round expected efficiency is at its minimum of 0.463 at ($\alpha = 1/3, \beta = 1/3$) which is a little over half of the total expected efficiency of 0.905 at this point given

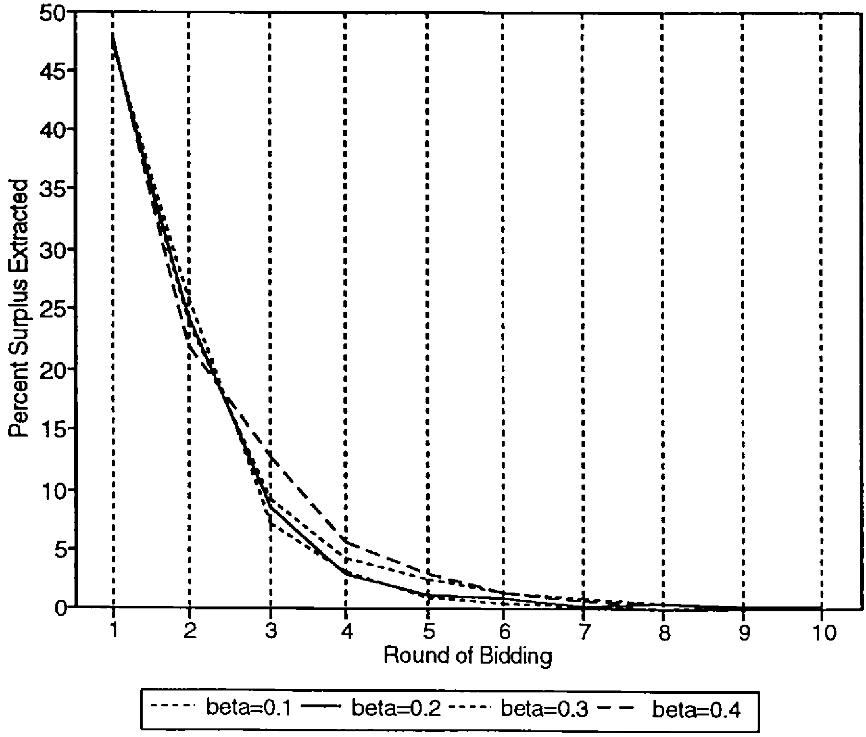


FIGURE 8 Time profile of surplus extraction in synchronized double auctions (by round).

by Expression (1). Thus, anywhere from 50–100% of the surplus extracted in a synchronized double auction is extracted within the first round itself.[7]

While we do not have the expressions for the amount of surplus extracted in the subsequent rounds, it must decline at a rapid rate because such a high proportion is extracted in the first round itself. Figure 8 shows the time profile of surplus extraction in the first ten rounds of such auctions for $\alpha = 0.5, \beta = 0.1, 0.2, 0.3$, and 0.4 and $N \to \infty$ from 1000 iterations of a computer simulation.[8] As expected, the surplus extraction declines sharply within a few rounds, little surplus being extracted after the first four or five rounds of bidding. Repeated bidding is an essential feature of double auctions. The first round efficiency of the double auction could be compared to the efficiency of similar market institutions that are constrained to a single round of bidding.

CONCLUDING REMARKS

The assumption of utility-maximizing traders in economic theory is often criticized because it ignores the well-documented cognitive limitations of human beings. However, the implications of this discrepancy between facts and assumptions about human behavior for the predictions of economic theory have remained controversial. Zero-intelligence (ZI) traders serve as a lower benchmark of intelligence. The results obtained here demonstrate the robustness of certain theoretical predictions when the assumption of individual rationality is relaxed in the extreme.

The design of exchanges populated by the identical, high-speed ZI traders requires a precise specification of market rules. This poses interesting problems for modeling and implementation of these markets. In the computerized markets of the laboratory, the speed of the underlying hardware and software exceeds the speed of human response by several orders of magnitude, giving rise to a wide range of micro-level issues in modeling of trader behavior and implementation of market institutions. We do not address these important micro-level design issues.

The demand and supply configuration used for the analysis can serve as a useful framework for studying the impact of extra-marginal traders on market efficiency. We have used a simple model of double auction, populated by zero-intelligence traders, to arrive at the following conclusions. Whether, or to what extent, these conclusions will hold in more complex double auction settings is an open issue.

[7] Recall that we are considering an environment in which there is only one intra-marginal unit. Consequently, the first round efficiency of synchronized double auction given by expression (3) is the same as the efficiency of buyer's bid double auction (see Satterthwaite and Williams,[11] in this volume). For double auctions with multiple intra-marginal units, the meaning of the "first round" would have to be appropriately redefined.

[8] In these simulations of synchronized double auctions it is surprisingly easy to let $N \to \infty$ simply by setting the highest extra-marginal bid to β and the lowest extra-marginal ask to $1 - \alpha$ in each round of bidding.

1. The expected efficiency of double auctions has a lower bound. Even when they are populated by zero-intelligence traders (no ability to maximize or even seek profits, or observe or remember market events), these markets are guaranteed to yield, on average, a high proportion of their surplus to the traders. A large part of the efficiency of these markets is the result of their structural properties, independent of the motivations or abilities of the traders who participate in them.
2. The apparently minor differences in the rules of continuously clearing and synchronized double auctions have important consequences for their surplus extraction properties. When compared to synchronized double auctions, the minimum expected efficiency of continuous double auctions is lower by about 6% while their mean efficiency over the feasible parameter range is lower by about 4%. It would be premature, however, to conclude on the basis of this result that the continuously clearing form is necessarily less desirable than the synchronized form of double auction in all situations. Continuous auctions may have superior price discovery properties and may therefore dominate synchronized markets in environments where the price discovery function is an important issue. The results and conjectures presented here only point to interesting directions for future work.
3. There are two possible causes of reduction in efficiency of auctions: (a) money left on the table by traders and (b) displacement of intra-marginal traders by extra-marginal traders. Money is left on the table when potentially profitable trades are unexploited at the end of the auction. Double auctions with inexperienced human traders often exhibit such behavior in early periods, but it tends to disappear with even a small amount of experience. The zero-intelligence machine traders of our markets repeatedly submit bids and asks from beginning of a period to the end; therefore, these markets do not suffer from this source of inefficiency. Displacement of intra-marginal traders is the only source of inefficiency in the markets examined here.
4. Efficiency of double auctions is influenced not only by the shape of demand and supply to the left of the equilibrium point but also by their shape to the right. However, the magnitude of this influence of extra-marginal units on expected efficiency is constrained by two countervailing forces. As the value of either α or β is increased from zero, chances that an efficiency-reducing transaction will occur increase; at the same time, the magnitude of efficiency reduction from such a transaction declines, yielding the efficiency-minimizing combination of parameters shown in Figure 2.
5. The ratio of extra- to intra-marginal traders in the market determines the magnitude of shortfall in expected efficiency of double auctions. As this ratio increases, expected efficiency declines as the extra-marginal traders increase their chances of displacing the intra-marginal traders. This effect is especially pronounced in continuously clearing double auctions.
6. Of the total surplus exploited in a synchronized double auction, a significant proportion (from 50–100%) is extracted within the first round itself, leaving only a small amount for the next few rounds, and little for the rest. This time

profile of surplus extraction measures the efficiency gains arising from the repeat bidding feature of double auction.

ZI traders can serve as a powerful tool for analysis and comparison of market institutions because they help isolate the consequences of market structure from the behavior of market participants. The behavior of human traders may change in response to changes in economic institutions, making it difficult to isolate the impact of changes in market structure alone on the basis of data from the field. The use of artificial traders is a convenient tool for holding behavior constant, while the structure of the economic institutions is varied to examine their consequences.

ACKNOWLEDGMENTS

We have benefitted from comments of Dan Friedman, Charles Plott, John Rust, Vernon L. Smith, Robert Wilson, and from the participants in presentations given at Carnegie-Mellon University, Economic Science Association, University of Kobe, and Yale University. Detailed comments of an anonymous referee, and from various participants at Santa Fe Institute and Bonn workshops were especially valuable. Financial support for this research was provided by Richard and Margaret Cyert Family Fund, Deloitte and Touche Foundation, National Science Foundation under grant SES-8915225, Santa Fe Institute, and Deutsche Forschungsgemeinschaft, Gottfried-Wilhelm-Leibniz-Forderpreis is gratefully acknowledged. We alone are responsible for any errors.

APPENDIX

DERIVATION OF EXPECTED EFFICIENCY OF SYNCHRONIZED DOUBLE AUCTION

Expected efficiency of a synchronized double auction populated by budget-constrained zero-intelligence traders (with one intra-marginal buyer and seller each, an infinity of extra-marginal buyers and sellers) is given by:

$$\begin{array}{ll} 1-(1-\alpha-\beta)\Big(\dfrac{\alpha^2}{1-\alpha(1-\alpha)}+\dfrac{\beta^2}{1-\beta(1-\beta)}\Big) & \text{for } \alpha+\beta\le 1\,, \\ 1 & \text{for } \alpha+\beta>1\,. \end{array} \tag{4}$$

To derive this expression, let a_1 and b_1 be the ask and bid submitted by the intra-marginal seller and buyer respectively in the first round. Both a_1 and b_1 submitted by "zero intelligence" traders are distributed independently and uniformly over [0, 1]. In addition, the asks submitted by "zero intelligence" extra-marginal sellers are distributed uniformly over $[(1-\alpha), 1]$. As the number of extra-marginal sellers increases without bound, the lowest ask submitted by these sellers converges to $1-\alpha$. Bids submitted by "zero intelligence" extra-marginal buyers are distributed uniformly over $[0, \beta]$; as the number of extra-marginal buyers increases without bound, the highest bid submitted by these buyers converges to β.

In order to calculate the expected surplus extracted in this market when $(\alpha+\beta)<1$, divide the square in (a_1, b_1) plane into nine cells as shown in Table 1.

CELL 1:. Both intra-marginal bid and ask are less than β. Probability of this event is β^2. The intra-marginal bid $b_1 < \beta$ is outbid by the maximum of the extra-marginal bids (at β). This highest extra-marginal bid of β is crossed with the intra-marginal ask $a_1 < \beta$ to effect a transaction and realize a surplus of β. (At this point we are concerned only with the total surplus, and not with how this surplus is split between buyers and sellers.) With $(\alpha+\beta)<1$, extra-marginal buyers and sellers cannot trade with each other. Therefore the only remaining trade possible is between the intra-marginal buyer and an extra-marginal seller. Since the current ask submitted by these sellers is $(1-\alpha)$, it is only a matter of time when, in subsequent rounds of bidding, the intra-marginal buyer submits a bid higher than $(1-\alpha)$ and a transaction takes place, realizing a surplus of α. Thus total surplus of $(\alpha+\beta)$ is extracted in Cell 1, yielding an expected surplus of $\beta^2(\alpha+\beta)$. Since the maximum expected surplus that could have been extracted is β^2, it represents a loss of $g_1 = \beta^2(1-\alpha-\beta)$ in expected surplus.

TABLE 1 Intra-Marginal Asks and Bids

	Bid b_1 from intra-marginal buyer		
	Cell 7 Prob: $\alpha\beta$ $(1-\alpha) < a_1 < 1$ $0 < b_1 < \beta$	**Cell 8** Prob: $\alpha(1-\alpha-\beta)$ $(1-\alpha) < a_1 < 1$ $\beta < b_1 < (1-\alpha)$	**Cell 9** Prob: α^2 $(1-\alpha) < a_1 < 1$ $(1-\alpha) < b_1 < 1$
Ask a_1 from intra-marg. seller	**Cell 4** Prob: $\beta(1-\alpha-\beta)$ $\beta < a_1 < (1-\alpha)$ $0 < b_1 < \beta$	**Cell 5** Prob: $(1-\alpha-\beta)^2$ $\beta < a_1 < (1-\alpha)$ $\beta < b_1 < (1-\alpha)$	**Cell 6** Prob: $\alpha(1-\alpha-\beta)$ $\beta < a_1 < (1-\alpha)$ $(1-\alpha) < b_1 < 1$
	Cell 1 Prob: β^2 $0 < a_1 < \beta$ $0 < b_1 < \beta$	**Cell 2** Prob: $\beta(1-\alpha-\beta)$ $0 < a_1 < \beta$ $\beta < b_1 < (1-\alpha)$	**Cell 3** Prob: $\alpha\beta$ $0 < a_1 < \beta$ $(1-\alpha) < b_1 < 1$

CELL 9: This case is analogous to Cell 1; the expected surplus associated with this cell is $\alpha^2(\alpha+\beta)$, or a loss of $g_9 = \alpha^2(1-\alpha-\beta)$ in expected surplus.

CELL 2: In this cell, the intra-marginal buyer outbids the extra-marginal buyers ($b_1 > \beta$) and the former crosses with the intra-marginal ask ($\alpha_1 < \beta$), yielding a surplus of 1 with probability $\beta(1-\alpha-\beta)$. Expected surplus associated with this cell is therefore the maximum possible $\beta(1-\alpha-\beta)$ and $g_2 = 0$.

CELL 6: By argument analogous to Cell 2, the expected surplus associated with this cell is the maximum possible $\alpha(1-\alpha-\beta)$ and $g_6 = 0$.

CELL 3: Again, intra-marginal units transact with each other, yielding a surplus of 1 with probability $\alpha\beta$ or expected surplus of $\alpha\beta$ equal to its maximum possible value and $g_3 = 0$.

CELL 5: Since the intra-marginal buyer outbids the extra-marginal buyers in this cell, and intra-marginal sellers have outbid the extra-marginal sellers in the first round, the extra-marginals have no chance of entering trading. The bid/ask improvement rule of double auction makes it impossible for these traders to hold the current bid/ask in a subsequent round after the first round bids fall in this cell. Intra-marginal units will necessarily trade with each other (either in the first round if the intra-marginal bid exceeds the intra-marginal ask, or in a later round of bidding). In any case, there can be no loss of surplus once the intra-marginal bid and ask occupy this cell. Therefore, $g_5 = 0$.

When the first round bids and asks fall in Cells 4, 7, or 8, no transaction can take place until subsequent rounds of bidding.

CELL 7: This cell probability in the first round is $\alpha\beta$. Probabilities in the second and subsequent rounds in this cell are exactly the same as in the first. If g is the expected loss of surplus for this double auction, $\alpha\beta g$ is the expected loss of surplus associated with this cell.

CELL 4: The first round probability of this cell is $\beta(1-\alpha-\beta)$, and it forces a second round of bidding. If intra-marginal bids/asks submitted in the second round fall in Cells 2, 3, 5, or 6, no loss of surplus takes place. If the second-round submissions are in Cell 8 and 9, the maximum bid/minimum ask over the two rounds falls in Cell 5 and 6 respectively, again leading to transactions with no loss of surplus. If second-round submissions are in Cell 1 (probability β^2), surplus-reducing transactions cause an expected loss of $\beta^2(1-\alpha-\beta)$. If second-round submissions are in Cell 7 (probability $\alpha\beta$), the minimum ask/maximum bid fall in Cell 4. Thus, with probability $\alpha\beta+\beta(1-\alpha-\beta) = \beta(1-\beta)$, bidding is forced into a third round, and so on. Thus the probability of loss, from the first round bids falling in this cell, is given by:

$$\beta(1-\alpha-\beta)\beta^2(1+\beta(1-\beta)+\beta^2(1-\beta)^2+\beta^3(1-\beta)^3+\ldots) = \frac{(1-\alpha-\beta)\beta^3}{1-\beta(1-\beta)}.$$

The corresponding expected loss from first round bids falling in this cell is

$$g_4 = \frac{(1-\alpha-\beta)^2\beta^3}{1-\beta(1-\beta)}.$$

CELL 8: By argument analogous to Cell 4, the expected loss of surplus from Cell 8, g_8, is given by

$$g_8 = \frac{(1-\alpha-\beta)^2\alpha^3}{1-\alpha(1-\alpha)}\,.$$

Let g be the expected loss of surplus from all nine cells.

$$\begin{aligned}
g &= \sum_{i=1}^{9} g_i \\
&= \alpha^2(1-\alpha-\beta) + \frac{\alpha^3(1-\alpha-\beta)^2}{1-\alpha(1-\alpha)} + \beta^2(1-\alpha-\beta) + \frac{\beta^3(1-\alpha-\beta)^2}{1-\beta(1-\beta)} + \alpha\beta \times g\,\cdot \\
&= (1-\alpha-\beta)\left(\frac{\alpha^2}{1-\alpha(1-\alpha)} + \frac{\beta^2}{1-\beta(1-\beta)}\right)
\end{aligned}$$

Expected efficiency is

$$1 - g = 1 - (1-\alpha-\beta)\left(\frac{\alpha^2}{1-\alpha(1-\alpha)} + \frac{\beta^2}{1-\beta(1-\beta)}\right).$$

REFERENCES

1. Becker, Gary S. "Irrational Behavior and Economic Theory." *J. Pol. Econ.* **70** (1962): 1–13.
2. Easley, D., and John Ledyard. "Theories of Price Formation and Exchange in Double Oral Auctions." This volume.
3. Friedman Daniel. "On the Efficiency of Experimental Double Auction Markets." *Am. Econ. Rev.* **74(1)** (1984).
4. Friedman, Daniel. "The Double Auction Institution: A Survey." This volume.
5. Gode, Dhananjay, and S. Sunder. "Allocative Efficiency of Markets with Zero Intelligence (ZI) Traders: Market as a Partial Substitute for Individual Rationality." Carnegie-Mellon University Working Paper, 1991.
6. Gode, Dhananjay, and S. Sunder. "Human and Artificially Intelligent Traders in Computer Double Auctions." Presented at the meetings of Economic Science Association, Tucson, 1989.
7. Palmer, R., J. Rust, and J. Miller. "Participant's Manual for a Double Auction Tournament." Manuscript, Santa Fe Institute, 1989.
8. Plott, Charles R. "Industrial Organization Theory and Experimental Economics." *J. Econ. Lit.* **20** (1982): 1485–1527.
9. Rust, J., R. Palmer, and J. Miller. "A Double Auction Market for Computerized Traders." Working Paper 92-02-008, Santa Fe Institute, 1992.
10. Rust, J., J. Miller, and R. Palmer. "Behavior of Trading Automata in a Computerized Double Auction Market." This volume.
11. Satterthwaite, Mark A., and Steven R. Williams. "The Bayesian Theory of the k-Double Auction." This volume.
12. Smith, V. L. "An Experimental Study of Competitive Market Behavior." *J. Pol. Econ.* **70** (1962): 111–137.
13. Smith, V. L. "Experimental Economics: Induced Value Theory." *Am. Econ. Rev.* **66** (1962): 274–279.
14. Smith V. L. "Microeconomic Systems as an Experimental Science." *Am. Econ. Rev.* **72(5)** (1982).
15. Wilson, Robert. "On Equilibria of Bid-Ask Markets." Technical Report 452, Institute for Mathematical Studies in Social Sciences, Stanford University, 1984.

Tim Bollerslev† and Ian Domowitz‡
†Department of Finance, Kellogg Graduate School of Management, Northwestern University, and ‡Economics Department, Northwestern University

EIGHT
Some Effects of Restricting the Electronic Order Book in an Automated Trade Execution System

We use computer simulations to study the effects of varying the length of an electronic order book within a particular automated trade execution system. The length of the book is defined as the number of bids and offers allowed to rest on the book at any instant in time for eventual execution in accordance with the price, time, and quantity priority rules governing the trade matching algorithm of the automated system. Efficiency defined in terms of surplus extraction is unaffected by variations in the length of the book, but price variability around the equilibrium value increases with book length. Speed of convergence to equilibrium is slower as book length grows for the case of the trading of single units of the asset, but improves with book length if multiple units are traded. We show that certain properties of transaction prices and bid-ask spreads, credited to behavioral considerations in the financial literature, can be a function of this particular aspect of system design alone. Two different models of the market environment are used to document the speed of decline in price volatility, bid-ask spread, and spread volatility as the length of the order book increases. The appearance and increasing persistence of serial correlation in the variance of transactions price returns is traced to the existence and length of the electronic book, as is the degree of non-normality in transactions returns. Increases

The Double Auction Market, Eds. D. Friedman and J. Rust,
SFI Studies in the Sciences of Complexity,
Proc. Vol. XIV, Addison-Wesley, 1993

in the serial correlation of the market bid-ask spread as the book lengthens is isolated as one possible transmission mechanism of serial dependence in the variance of transactions prices.

1. INTRODUCTION

Automated systems that enable the price discovery and quantity determination process in financial markets include: price information dissemination services, automatic order routing, clearance and settlement procedures, and computerized trade execution mechanisms. The importance of the latter class of automated procedures has grown enormously in recent history. On an international scale, there are over twenty automated exchanges for the trading of futures and options contracts alone, including exchanges in England, Japan, Germany, Switzerland, Australia, Ireland, New Zealand, and the United States. A variety of automated trade execution systems for stock trading are surveyed in Domowitz.[12]

The essential function of an automated trade execution mechanism in a continuous market is simply stated.[1] Bids and offers for a security, consisting of price and quantity pairs, are submitted by traders via computer terminals to a central processor. The execution mechanism is an algorithm that performs a matching function according to a set of rules governing the priority of submitted orders to buy and sell. The sequential stream of bids and offers then is transformed into transactions prices and quantities. For every transaction, an individual trader receives some fraction of the trader's desired quantity at a price that is as least as good as desired. The fraction of quantity received can be zero; i.e., execution is not guaranteed but depends on market conditions as reflected in the order flow.

This computerized matching algorithm may be contrasted with the traditional floor- or pit-trading process, where individual traders meet face to face. In such a setting, open outcry auction rules prevail. A contract is traded when an outstanding verbal bid or offer is accepted. A new auction starts with each completed transaction and the priority of bids and offers does not carry over from auction to auction.

Efforts to compare the pricing properties of floor trading with a particular automated futures market in Domowitz[10,11] and Bollerslev and Domowitz[4] reveal that an electronic order book which saves incoming orders over time for eventual execution provides the main difference in performance between the two trading mechanisms. The existence of such an electronic order book is part of most automated

[1] Automated execution also exists for call auctions. In such a setting, bids and offers are accumulated and an algorithm finds a clearing price at discrete intervals according to a criterion such as the maximization of volume traded (see, for example, Cohen and Schwartz[9]). We restrict attention to continuous auction systems in this paper.

trade execution systems.[2] The purpose of this paper is to provide a preliminary characterization of the effect of such an order book on properties of transactions prices and market spreads within an automated system. In particular, we focus on the length of the book. By length, we mean the number of bids or offers allowed to rest on the book at any instant in time in accordance with the priority rules governing the matching algorithm. This emphasis is in the spirit of the experimental work of Smith and Williams[27] on alternative queuing systems.

Our interest in this aspect of system design is motivated in part by the nature of the results obtained in Bollerslev and Domowitz[4] and Domowitz and Wang.[13] Floor trading was essentially modeled as a set of price and time priority rules with a book of length one. That is, orders were not saved or "remembered" if they failed to better the existing best price. Similarly, if a standing best price was superseded by a better bid or offer, it was discarded as being ineligible for a trade in the future.[3] On the other hand, the automated system contained a book of infinite length. If an offer to sell was submitted at a price above current quotes in the market, that offer was saved for eventual execution according to the system's priority rules. The system with a book of infinite length was characterized by lower transactions price volatility, smaller bid-ask spreads and spread volatility, higher first-order serial correlation in returns, and better order execution rates. Most interesting, perhaps, was the finding that returns based on the continuously compounded price changes from transaction to transaction in a system with a book of length one exhibit little predictable variation in the volatility of the returns. The volatility of the transactions price returns from the infinite-length book system varies over time with a great deal of persistence and predictability. Although this result says little about efficiency in a market studied in isolation, it is relevant in the consideration of efficiency in markets for derivative securities, such as options.[4] In this paper we investigate the extent to which such differences depend on the length of the book; e.g., how fast does the price volatility, defined as the variance of logarithmic price changes conditional on past history, decrease as the book length increases?

We also show that certain properties of transactions prices and bid-ask spreads credited to behavioral considerations in the financial literature can be a function of this particular aspect of system design alone. Lower price volatility and smaller bid-ask spreads in a specialist market such as the New York Stock Exchange relative to a dealer market have been ascribed to the activities of the specialist (e.g., Beja

[2] Exceptions include the London Futures Exchange system and the Aurora system initially proposed by the Chicago Board of Trade. Both systems were designed to simulate pit trading directly. It seems likely that the latter system will not go into operation, however.

[3] This is a restrictive abstraction of floor-trading behavior, however. For example, the best price may be represented by several traders in a trading pit, in which it is difficult to maintain strict price priority. Order resubmission after a transaction may be virtually simultaneous amongst traders. Under such a scenario, an incoming order may be executed against any one of such traders, all of whom are offering the best price in the market. Analysis of this representation of the floor relative to automated execution is contained in Domowitz.[11] The qualitative conclusions cited in the text with respect to relative pricing and bid-ask spread properties still hold, however.

[4] See, for example, the experimental study of options markets by O'Brien and Srivastava.[25]

and Hakansson[2]). Stoll[26] describes key variables lowering spreads to include volume of trading, lower unsystematic risk, possibly related to the amount of adverse information, and number of competing dealers or markets in a security. First-order serial correlation in returns has been attributed to the interaction between market orders and limit orders (see e.g., Cohen, Maier, Schwartz and Whitcomb[5] (henceforth CMSW) and Niederhoffer and Osborne[24]). Volatility clustering has been modeled primarily in terms of rational reactions to news events (see e.g., Baillie and Bollerslev[1] and Engle, Ito, and Lin[14]). We do not quarrel with any of these arguments. The work presented here does show, however, that all such effects can be accounted for by differences in the memory of an order execution system modeled in terms of the length of a limit order book. In this regard, our work is most closely related to that of CMSW,[5] who study the effect of what they term the "stickiness" of a limit order book on serial correlation in returns in a simulation model of a market such as the New York Stock Exchange.

We provide additional evidence on market efficiency in the spirit of the Smith and Williams[27] study of order queuing systems in an experimental setting. In our terminology, they analyze a book of length one and a book of length infinity. Subjects were not shown the content of the book in the latter case. As in their experiments, surplus extraction is unaffected in any systematic way by changing the length of the book. The Smith-Williams volatility measure includes two additive components, the price variance and the squared deviation of price from the equilibrium value. This measure declines in their experiments as subjects move from an institution characterized by book length one to an auction with infinite book length. We find that the variance of prices decreases in a smooth fashion as the book length increases. Variability around the equilibrium price tends to increase, however, and the speed of convergence to equilibrium declines in experiments restricted to trades of a single unit of the security. The latter result is reversed, however, in the trading of multiple units.

We also study alternative measures characterizing prices and market efficiency. Across two very different models of the market environment, we find that the book length is inversely related to the size of the bid-ask spread, defined as the difference between the best bid and offer outstanding in the market, and the volatility of the spread. The degree of serial correlation in the transactions returns and the bid-ask spread process also are linked directly to the length of the book. The latter provides a possible explanation for the appearance of serial correlation in the conditional second moments of transactions returns once the length of the book exceeds unity. This serial persistence, in variance, also increases with the length of the book, and is accompanied by a greater degree of kurtosis in transaction-price returns. A common feature of all such results is the fact that most of the relevant effect (e.g., the decrease in overall market volatility) is obtained with a book of very short length. The automated trade execution algorithm examined is that of the Chicago Mercantile Exchange/Reuters Globex futures trading system, which is described along with the modification allowing varying book lengths in Section 2. The analysis of a set of computerized auction experiments is the topic of Section 3. In Section 4, we modify the market simulation used in Bollerslev and Domowitz[4] to allow the study of conditional volatility and market spreads as the length of the book is varied. Section 5 contains some caveats and concluding remarks oriented towards further work on this problem.

2. AN AUTOMATED EXECUTION SYSTEM WITH A VARIABLE BOOK

The Globex trading system has been under development by the Chicago Mercantile Exchange and Reuters since at least 1988.[5] Globex is an automated system for trading CME futures and options outside regular floor-trading hours in Chicago. Our focus on Globex trading rules is motivated by the desire to study a concrete example of an automated trade execution system, rather than to attempt a stylized characterization which may hide aspects of the functioning of such systems. At the same time, preliminary examination of a variety of automated markets for continuous trading indicates broad similarities, and the rules of the Globex system are both simple and representative of the class of algorithms of interest.[6]

The Globex system is strictly a limit-order system, not permitting market orders, which is ideal for the purpose of this study. Good-til-canceled limit orders are written into the electronic book of bids and offers, and are executed in accordance with rules described below. In principle, any number of orders may be carried on this book.[7] Our only modification of Globex trading rules will be to limit the number of orders carried on the book at any given time. The trading rules ensure that bids and offers are matched based on criteria of price, time, and quantity. A complete algorithmic description is contained in Domowitz.[10] A simpler outline and example should suffice for our purpose here.[8] In the discussion below, we refer to the trading of a single financial instrument.[9]

[5]The CME and Reuters first put forward a written proposal for the project in May, 1988, in the form of a series of letters to the CFTC detailing the basic organization, rules, and amendments to existing CME trading rules.

[6]Stoll[28] also notes that the functions of automated markets appear to be remarkably similar across different financial instruments, including stocks, bonds, futures, and options. See Domowitz[12] for a survey of systems.

[7]The visual terminal display shows only the ten best bids and offers, but the information content for strategic trading is not a topic of this study. In any case, the results reported here suggests that increasing the size of the book beyond length ten is not required with respect to improvements in market performance.

[8]We ignore trading rules associated with so-called primary versus secondary (or "more" in Globex terminology) orders. The protocol is a bit complicated and not directly relevant. There also are special rules governing the setting of an opening price and pertaining to the changing or placing of an order "on hold." The latter concerns time priorities and also is not relevant. Analysis of the opening is a study in call market mechanisms; see, for example, Cohen and Schwartz[9] and Stoll[28] for discussion of many of the issues. We are concerned here only with continuous trading.

[9]Globex has provision for trading spreads between the prices of different securities, but this is done by specifying the spread as a separate financial instrument. This feature simplifies trading

TABLE 1 Order Arrival Sequence

Time	Bid	Quantity	Offer	Quantity
10:00	48.96	2		
10:01			49.50	4
10:02			49.50	4
10:03			49.48	4
10:04	49.49	3		
10:05			49.51	2
10:06	49.50	2		

Globex is a strict price and time priority system, in that order. For buy orders, higher price is higher priority; for sell orders, lower price is higher priority. If two orders are at the same price, the older order has priority for execution. A buy and sell order may be matched when the buy order price is greater than or equal to the price of the sell order. The price at which the transaction occurs is the price of the standing order; i.e., the price of the order already on the book when the contraside order arrives in the system. If there are multiple standing orders eligible for matching against a new order, then matching will be considered in price and time priority sequence until either the new order is completely filled or all eligible standing orders have been considered. If some quantity remains of the incoming order after this matching process, that quantity is written into the book as a standing order. Cancellation and revisions of orders can be carried out at any time. The cancellation and resubmission of an order sacrifices the time priority of the original quote.

In order to illustrate the process and introduce the idea of an abbreviated book, consider the sequence of bids and offers in Table 1 for a single financial instrument.

Given no previous orders on record in the book at 10:00, in the Globex system with an unrestricted book, the book at 10:04 would look precisely as above, with three offers and two bids. Upon the arrival of the bid of 49.49 at 10:04, three contracts would be traded at a price of 49.48 (the price of the standing order). Three offers would remain on the book, with the offer at a price of 49.48 for the remaining one contract. The offer arriving at 10:05 is recorded in the book. The bid of 2 contracts at 49.50 first is executed against the standing offer of 49.48 for a trade of one contract at the latter price, followed by execution against the 10:01 offer at 49.50 for a trade of one contract. The book at 10:06, therefore, looks like Table 2.

rules, but is not common to all automated systems. The Swiss SOFFEX system, for example, allows contingency orders linking different futures and options contracts.

TABLE 2 Order Book at 10:06. Infinite Book Length

Time	Bid	Quantity	Offer	Quantity
10:00	48.96	2		
10:01			49.50	3
10:02			49.50	4
10:05			49.51	2

TABLE 3 Order Book at 10:06. Book Length One

Time	Bid	Quantity	Offer	Quantity
10:06	49.50	1		

TABLE 4 Order Book at 10:04. Book Length Two

Time	Bid	Quantity	Offer	Quantity
10:00	48.96	2		
10:01			49.50	4
10:03			49.48	1

TABLE 5 Order Book at 10:06. Book Length Two

Time	Bid	Quantity	Offer	Quantity
10:00	48.96	2		
10:01			49.50	3

The polar case is a book of length one. Consider the same sequence of bids and offers as before. The offer of four contracts at 10:02 is not recorded; the previous offer was at the same price and has time priority. The trader submitting at the price of 49.50 at 10:02 must resubmit at a later time or better price to achieve execution. The arrival of an offer of 49.48 at 10:03 replaces the offer at 49.50, based on price priority. The offer of 49.50 no longer is carried on the book. The arrival of a bid at 10:04 triggers a trade of three contracts at the price of the standing order, 49.48, leaving one contract at 49.48 still offered on the book. The incoming offer of 49.51 is not recorded. At 10:06, one contract is traded at a price of 49.48, leaving one contract bid at 49.50 on the book, with the offer side of the book now empty (Table 3).

It is obvious that this sequence of trades and implied market spreads is quite different from that achieved with a book of unrestricted length. Now consider the same sequence of bids and offers with an order book of length two. After matching the eligible orders, at 10:04 the book has the form shown in Table 4.

Note that the 10:02 offer of 49.50 is missing from the book. It was supplanted by the offer at a price of 49.48. With a book of length two, the 10:01 order at 49.50 remains on the book, however. Again, three contracts are traded at 10:04 at a price of 49.48. The incoming 10:05 offer of 49.51 is not recorded, however, because the offer book still contains two orders at better prices. As in the first example, the incoming bid of 49.50 is executed first against the standing order of 49.48 for one contract, then against the 10:01 offer for one contract. After this transaction, the book has the form of Table 5.

Although the pattern of executed trades looks the same for a book of length two as for an unrestricted book, the remaining orders on the book for future transactions now look very different. The sequence of future trades will change relative to that experienced through the unrestricted book. Of course, the pattern of market spreads has changed as well.

3. A COMPUTERIZED AUCTION EXPERIMENT

The design of the first market environment used in this study is common to many auction experiments. Each market consists of 24 computerized trading strategies and a set of demand and supply schedules.[10] Six sets of demand and supply schedules are used and illustrated in Figure 1. Prices range from 0 to 100, and the equilibrium price in each market is 50. A "market environment" consists of one of the six sets of demand and supply curves and a fixed length of book. Books of length 1, 2, 3, 4, 6, 8, 10, and 12 are investigated here, yielding results for 48 environments.

[10] The effect of varying the number of traders in the setup used here is explored in Domowitz and Wang.[13]

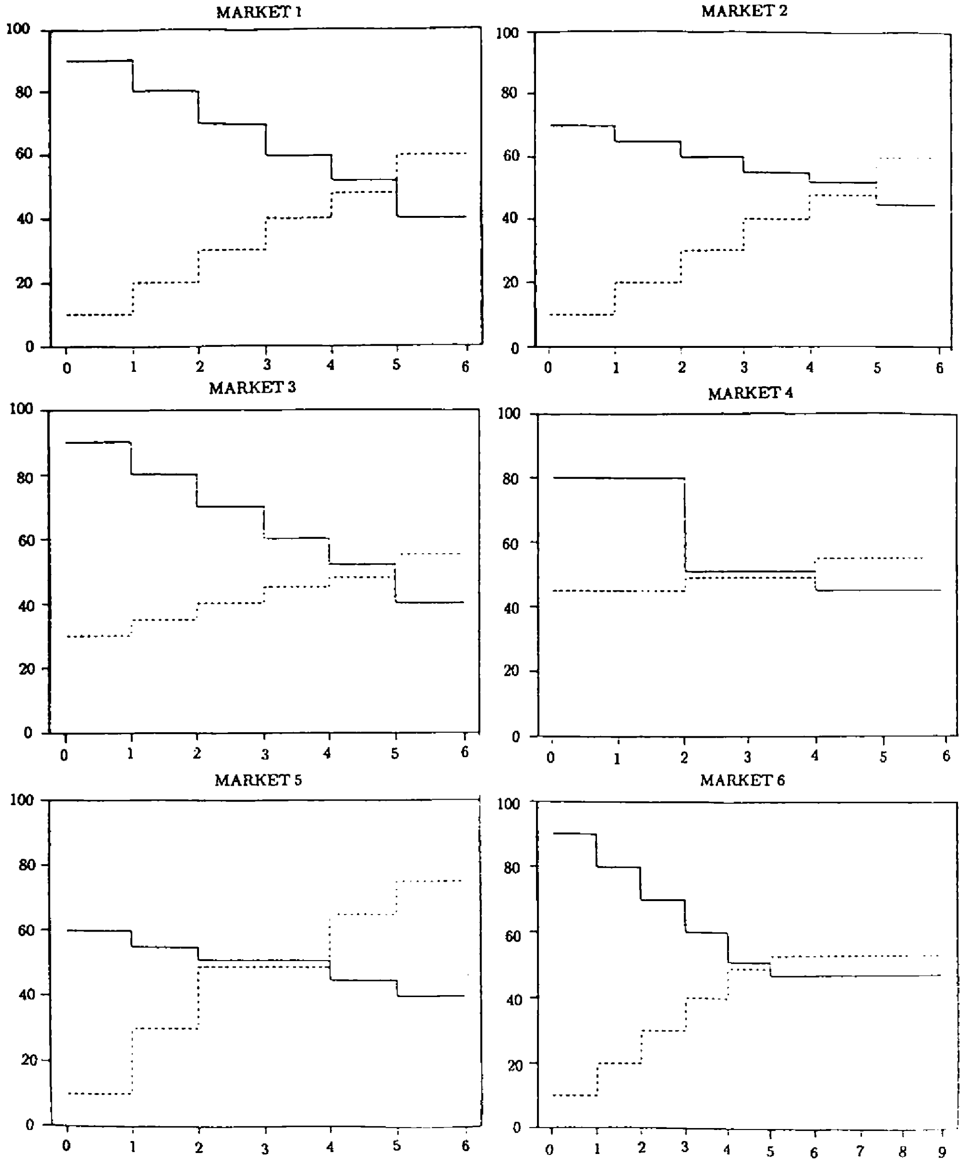


FIGURE 1 Supply and Demand Schedules.

Each step on the demand schedule represents a limit-buy order received by a trader. The steps of the supply schedules represent sell orders. The prices of these limit orders constitute reservation values for the traders. Traders are classified as either buyers or sellers. A buyer receives one order for a single contract at each price (step) of the demand curve, and a seller similarly receives a single sell order per period at each price of the supply curve.[11]

Traders with unfilled orders are randomly chosen in continuous time to make bids or offers. A buyer first will fill the order with the highest reservation value and "walk down the demand curve," and vice versa for a seller. If a trader has an order active on the book at the time chosen, the trader cancels the existing bid or offer and resubmits. This cancellation rule ensures that orders do not become stale. Resubmission can be at a random price, subject to the no-loss strategy discussed below, or at a better price than offered or bid previously by the same trader. Both cancellation rules were investigated, and the qualitative nature of the results reported is invariant to the rule chosen. A period ends when the highest reservation value for buyers is lower than the lowest reservation value for sellers, i.e., when no further trade is possible.

Trading strategy deliberately is kept as simple as possible in order to focus on the effect of the length of the book. The results of Gode and Sunder[17] show that only a mild form of individual rationality is required to avoid significant reductions in the total exploited surplus in double auctions.[12] Their characterization of a "weakly rational" (ZI) trader is used here, augmented by the order cancellation and resubmission strategy. The "weakly rational" strategy is simply a no-loss constraint on a randomized bid/offer process. A buyer cannot bid above the underlying value of the order; a trader who sells cannot offer a contract at a price below the order value. Bids and offers are generated from uniform distributions subject to this constraint. This implies varying bounds on the support of the uniform distributions generating bids and offers. The maximum support is 0 to 90, for the bid of a trader attempting to fill an order with a limit price of 90, in market 1. The analogue for a seller is a support of 10 to 100. Supports narrow as trading progresses. If a trader has succeeded in transacting all buy orders down to a price of 60, the support for a randomized bid is now 0 to 60. The minimum range of the support for buyers and sellers is 55.

It is certainly true that for any given environment, an increase in the support will degrade market efficiency measures, including price volatilities and the magnitude of the average bid-ask spread. Our interest here is in comparisons across environments differentiated by length of book, however. Variations in the size of the supports do not affect comparisons reported in terms of relative percentage increases or decreases in the measures reported here. The supports used here are quite large by financial trading market standards. Differences in prices across a

[11]The results reported here are robust to variations in design to include Poisson arrivals of orders at each step of the supply/demand curves, as well as to designs in which the order arrival rates shift over time. See Domowitz and Wang[13] for details. The only exception to this statement concerns speed of convergence to equilibrium, discussed in the text.

[12]Explicit bounds for surplus extraction efficiency in the case of budget-constrained machine traders are provided in Gode and Sunder.[18]

limit order book at any instant in time typically will not be more than a few units defined in terms of some minimum price variation or "tick." Limiting the analysis to such a case of discrete pricing over a more narrow range also does not change qualitative comparisons.[13]

A trader's profit or surplus is the difference between the price of the order and the price at which the trade is consummated according to Globex rules. The trading process continues until no further trades are possible. Each market is replicated 100 times, providing 100 observations on the pattern of transactions prices, average bids and offers, surplus extraction, and the number of bids, offers, and trades required to clear the market. These figures are used to construct the performance measures described below. The results reported here are based on independent trials.[14]

The discussion in this paper concentrates on volatility, the bid-ask spread, and liquidity as performance measures. This emphasis is more in line with the literature on financial market microstructure, which tends to focus on observable characteristics of performance in working markets. Some analysis of trading efficiency in the tradition of experimental auction markets is possible and appropriate, however. Since the differences in supply and demand configuration did not substantially affect comparisons across book lengths, the discussion is limited to results obtained from market 1.

Surplus extraction is virtually 100% for all book lengths, averaged across periods. This echoes the findings of Gode and Sunder[17] with respect to weakly rational trading strategies. Although some inefficiency, by this measure, is reported by Gode and Sunder[18] for continuous-time trading activity, our machine traders are endowed with a bit more "intelligence" by means of the order cancellation and resubmission strategies. The cancellation procedure is not unrealistic, and the price improvement rule upon resubmission is common practice in the case of real-world traders attempting to achieve execution of trades. Once again, the issue is not the absolute percentage of surplus extracted in any given environment, but the differences across book lengths. No systematic differences were observed. The same result is found in Smith and Williams[27] in experimental comparisons between auctions with book lengths of one and infinity.

Average transactions prices are quite close to the equilibrium value, regardless of book length. On the other hand, the average variability of prices around the equilibrium value increases as book length increases. The standard deviation of prices centered around the equilibrium price increases from 1.75 to 3.07, moving from length one to length two, with a further increase to 3.60 at a book length of 12. It is not possible to compare this result directly to Smith and Williams,[27]

[13]See Domowitz.[11]

[14]It is possible to design the experiments so as to mimic 100 periods within a trading day, in which case anything remaining on the Globex book would be carried over to the next period. The qualitative results reported here are simply accentuated under such a design. Clearing the book at the end of a period corresponds to the practice of clearing the Globex book at the end of a trading day.

because the figures reported therein compound deviations around equilibrium with the variance of price.

This finding also says little about the speed of convergence to equilibrium. We document such convergence by examining the percentage surplus extracted as the number of bids and offers, relative to the total number required to clear the market, grows. For example, in market 1 with an order book of length one, 51.7% of potential surplus is extracted by the first 10% of total bids and offers needed to clear the market. This percentage of total surplus declines to 35% when book length is equal to 12, a significant difference (at any reasonable level of statistical significance) of 32%. The discrepancies across book length decrease as the percentage of surplus extracted relative to total bids and offers increases. For book length of one, 75.5% of surplus is extracted by the first 20% of total bids and offers, declining to 58.7% for books of length 12, a decrease of 22%. Any differences in surplus for higher percentages of bids and offers relative to the total are not statistically significantly different from zero. These observed declines in speed of convergence to equilibrium are not robust to changing the trading environment to include bids and offers for multiple units of the security, however. In market 1 with trading in multiple units, Domowitz and Wang[13] find that as the book increases in length, the percentage of surplus extracted from the first 10% of bids and offers increases from 26.7% to 35.1%, a factor of about 31%. No further statistically significant increases are observed for percentages of bids and offers relative to the total higher than 10%. The difference between the multiple-unit and single-unit cases is traceable to the handling of partially unfilled orders by the limit order book. If a five-unit bid can be matched with a ten-unit offer, five units are transacted and the remainder of the offer is written to the book for execution at a later time. In a book of length one, those five units must be resubmitted, constituting an additional offer and slowing convergence as measured by amount of surplus extracted for given percentages of bids and offers relative to the total. This phenomenon is not observed with single units, where any transaction absorbs all quantities bid and offered at an eligible price.

Thus, the evidence concerning the effect of book length on standard measures of market efficiency is mixed. There are no systematic differences across book structures with respect to surplus extraction. On the other hand, price variability around the competitive value appears to increase with book length. Differences in the speed of convergence to competitive equilibrium, as measured here, depend on whether the market is for multiple or single units of the security. In the more realistic case of multiple-unit trades, speed of convergence to equilibrium increases with book length.

We now turn to volatility, the bid-ask spread, and measures of liquidity as performance measures. Qualitatively, the results are quite similar across all markets investigated. The discussion below singles out a symmetric market with respect to supply and demand (market 1) and an asymmetric market (market 5).

TABLE 6 Standard Deviation of Transactions Price[1]

Book Length	Market 1	2	3	4	5	6
1	14.4	11.0	11.2	8.4	9.3	13.8
	(1.3)	(1.1)	(1.1)	(1.1)	(1.4)	(1.3)
2	12.9	10.0	9.9	7.6	8.0	13.0
	(1.6)	(1.4)	(1.7)	(1.6)	(1.8)	(1.7)
3	12.0	8.9	8.7	6.4	7.2	11.3
	(1.6)	(1.6)	(1.6)	(1.8)	(2.1)	(1.8)
4	11.6	9.0	9.0	7.0	7.7	11.4
	(1.6)	(1.8)	(1.6)	(2.1)	(2.2)	(1.5)
6	11.2	9.0	9.1	6.9	7.6	11.1
	(1.7)	(1.6)	(1.5)	(1.9)	(2.3)	(1.6)
8	11.1	8.7	8.9	6.6	7.4	11.1
	(1.7)	(1.5)	(1.5)	(2.0)	(2.5)	(1.6)
10	11.4	8.7	8.6	7.0	7.6	11.4
	(1.6)	(1.3)	(1.6)	(2.2)	(2.3)	(1.5)
12	11.3	8.9	8.8	7.0	6.9	11.3
	(1.5)	(1.7)	(1.5)	(2.1)	(2.1)	(1.6)

[1] Standard errors in parentheses.

Price volatility theoretically is linked to profits and transactions costs in such a way as to suggest the desirability of lower volatility in system design, all other things equal.[15] In the Kyle[21] trading model, volatility declines as information is absorbed into prices. The only difference in information modeled here is through the length of the electronic order book. As the book lengthens, the system implicitly carries more information about the market at any given point in time.

The standard deviation of transactions prices for all markets and all book lengths is reported in Table 6, and illustrated graphically in Figures 2 and 3 for markets 1 and 5.[16] In market 1, a decline in the standard deviation of 22% is observed as the book length increases from one to six. Roughly half of this decrease is achieved as we go from the polar case of length one to length two. Such results

[15]In Kyle's[21] model, profit and volatility are inversely related. CMSW[6] connect increases in transactions costs to increases in volatility.

[16]The standard deviation of prices is calculated as an average over the 100 trading periods of the standard deviation of transactions prices estimated using all prices in each period.

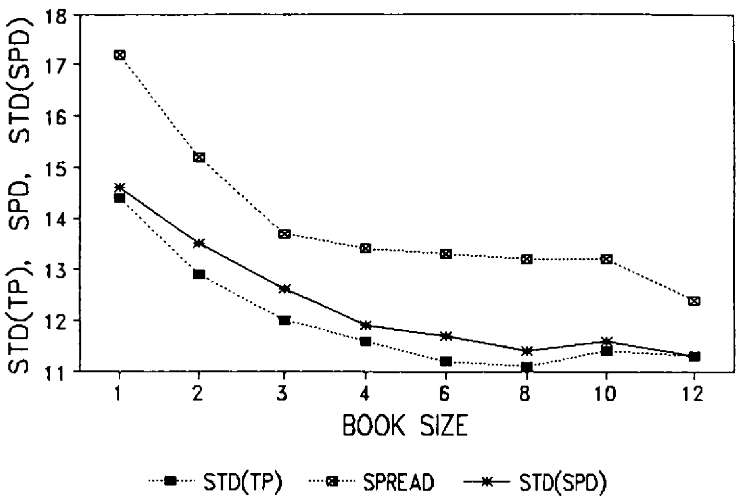


FIGURE 2 The effect of book size, 24 identical traders in market 1.

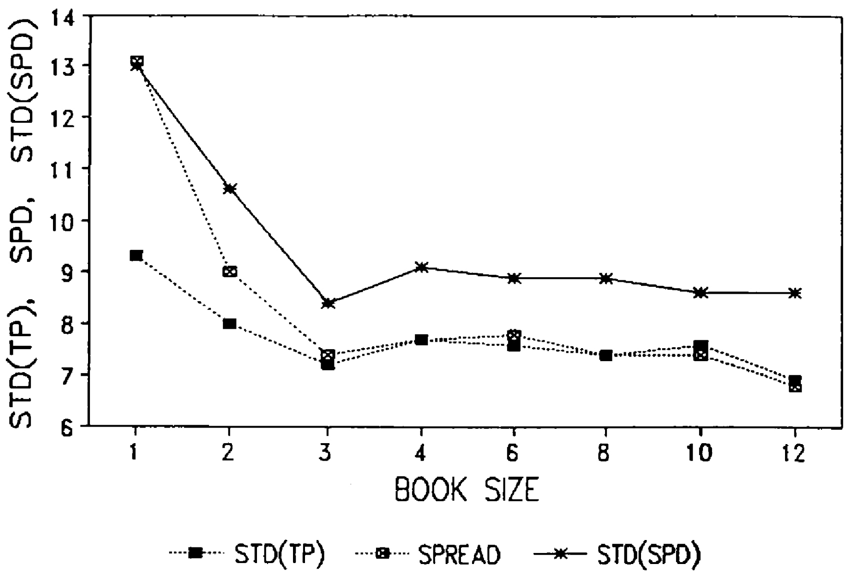


FIGURE 3 The effect of book size, 24 identical traders in market 5.

TABLE 7 Market Spread[1]

Book Length	Market 1	2	3	4	5	6
1	17.2	15.1	15.4	10.3	13.1	10.0
	(1.8)	(1.7)	(1.7)	(0.8)	(1.3)	(1.9)
2	15.2	12.7	12.9	8.4	9.0	11.3
	(1.7)	(2.2)	(1.8)	(1.2)	(1.7)	(2.1)
3	13.7	10.6	10.8	6.8	7.4	10.2
	(1.8)	(1.6)	(1.7)	(1.8)	(1.3)	(1.6)
4	13.4	10.9	10.7	7.0	7.7	10.4
	(1.6)	(1.6)	(1.6)	(1.0)	(1.4)	(1.6)
6	13.3	10.6	10.7	7.0	7.8	10.6
	(1.7)	(1.5)	(1.9)	(1.0)	(1.5)	(1.7)
8	13.2	10.7	10.6	6.7	7.4	10.0
	(1.8)	(1.5)	(1.8)	(1.0)	(1.3)	(1.7)
10	13.2	10.7	10.5	7.0	7.4	10.1
	(1.6)	(1.7)	(1.5)	(1.0)	(1.3)	(1.8)
12	12.4	9.9	10.4	6.4	6.8	10.4
	(1.7)	(1.4)	(1.5)	(1.0)	(1.5)	(1.8)

[1] Standard errors in parentheses.

are echoed in the market 5 experiments, in which a decline of 18% in the standard deviation is recorded as the book grows to length six. In the asymmetric market, however, 78% of this decline is attributable to an increase to length two. Smith and Williams[27] also report declines in their measure of volatility as the market institution changes from a book of length one to one of infinite length. For inexperienced traders, the appropriate comparison group for the present study, the decrease in the variance of prices plus the average squared deviation around equilibrium is 37.5%. The results reported here suggest that this effect is monotone in the length of the book.

The difference between the best bid and offer outstanding in the market, the size of the market spread, has proven to be a measure of market efficiency in different market contexts.[17] Tables 7 and 8 contain estimates of the magnitude and standard

[17] Black,[3] for example, relates the size of the spread to market efficiency through market liquidity. Madhaven[23] models the precision of the estimate of fair value of a security as proportional to the spread.

TABLE 8 Standard Deviation of Market Spread[1]

Book Length	Market 1	2	3	4	5	6
1	14.6	13.8	14.1	10.9	13.0	12.2
	(1.4)	(1.4)	(1.5)	(1.3)	(1.1)	(2.1)
2	13.5	12.2	12.6	10.0	10.6	13.0
	(1.6)	(1.8)	(1.7)	(1.9)	(1.8)	(1.8)
3	12.6	10.6	9.8	7.7	8.4	11.3
	(1.8)	(1.6)	(1.7)	(1.7)	(1.7)	(1.8)
4	11.9	10.2	10.2	8.2	9.1	11.8
	(1.8)	(1.6)	(1.9)	(1.7)	(1.8)	(1.9)
6	11.7	9.8	10.2	8.0	8.9	11.8
	(1.9)	(1.6)	(1.9)	(1.7)	(2.0)	(1.7)
8	11.4	10.0	10.0	7.8	8.9	11.3
	(1.9)	(1.7)	(1.8)	(1.6)	(2.0)	(1.9)
10	11.6	10.2	9.8	8.0	8.6	11.5
	(1.8)	(2.1)	(1.7)	(1.9)	(1.8)	(1.9)
12	11.3	9.8	9.9	8.2	8.6	11.7
	(2.1)	(1.8)	(1.6)	(2.0)	(1.9)	(1.9)

[1] Standard errors in parentheses.

deviation of the spread, respectively, for all markets and book lengths.[18] These measures are graphed against the standard deviation of price in Figures 2 and 3 for markets 1 and 5. It is clear that declines in spread size and volatility follow very similar patterns.

In market 1, moving from a book length of one to length six brings about a 23% decline in spread, accompanied by a 20% decline in spread standard deviation. Once again, about half of this effect is attributable to a move to book length two alone. More dramatic declines are observed in an asymmetric market such as 5, in which the spread and spread standard deviation are reduced by 40% and 32%, respectively, as book length increases to six.

It also is clear from Figures 2 and 3 that significant declines in all three statistics cease once the book reaches a length of six. In fact, in many cases these measures

[18] Reported spreads are calculated as the group average of all differences between the best bid and offer outstanding in the market averaged within the period and across periods. Spread standard deviation is calculated in the same fashion as price standard deviation.

basically stop changing, except for sampling variability, once the length of the book reaches four.

The same phenomenon is apparent in our examination of liquidity in these experimental markets. Kyle[21] discusses liquidity in terms of market tightness and depth. The measure reported as "liquidity" in Table 9 is related both to Kyle's parametric index of the effect of order flow on price changes and to Black's[3] characterization of liquidity as the market's ability to absorb quantity without an appreciable effect on price. The price disturbance from each trade should be small if the market is liquid. One such measure is provided by the ratio of the number of trades to the standard deviation at price. From Table 9 and Figure 4, liquidity by this measure increases by more than 30% in both markets 1 and 5 as the book length grows to six from one. From half to two thirds of this adjustment is attributable to a book length of only two, and any further improvements in liquidity cease once the book reaches a length of four to six.

Market tightness is defined by Kyle[21] as the cost of turning over a position in a short period of time. High market liquidity embodies the idea that such a cost

TABLE 9 Liquidity[1]

Book	Market					
Length	1	2	3	4	5	6
1	4.2	5.5	5.4	5.8	5.3	4.5
	(0.4)	(0.6)	(0.6)	(0.8)	(0.9)	(0.4)
2	4.8	6.1	6.3	6.6	6.4	4.8
	(0.6)	(0.9)	(1.1)	(1.6)	(1.6)	(0.6)
3	5.1	6.9	7.1	8.1	7.3	5.6
	(0 7)	(0.7)	(1.3)	(2.5)	(1.3)	(0.9)
4	5.3	6.9	6.9	7.7	6.8	5.4
	(0.7)	(1.4)	(1.3)	(3.1)	(2.3)	(0.8)
6	5.5	6.9	6.9	7.6	7.0	5.6
	(0.8)	(1.3)	(1.3)	(2.4)	(2.4)	(0.9)
8	5.5	7.1	6.8	8.1	7.4	5.6
	(0.9)	(1.2)	(1.1)	(3.0)	(2.8)	(0.8)
10	5.4	7.1	7.2	7.6	7.0	5.4
	(0.8)	(1.1)	(1.3)	(3.0)	(2.8)	(0.8)
12	5.4	7.0	7.0	7.6	7.6	5.5
	(0.7)	(1.3)	(1.2)	(2.6)	(2.4)	(0.8)

[1] Standard errors in parentheses.

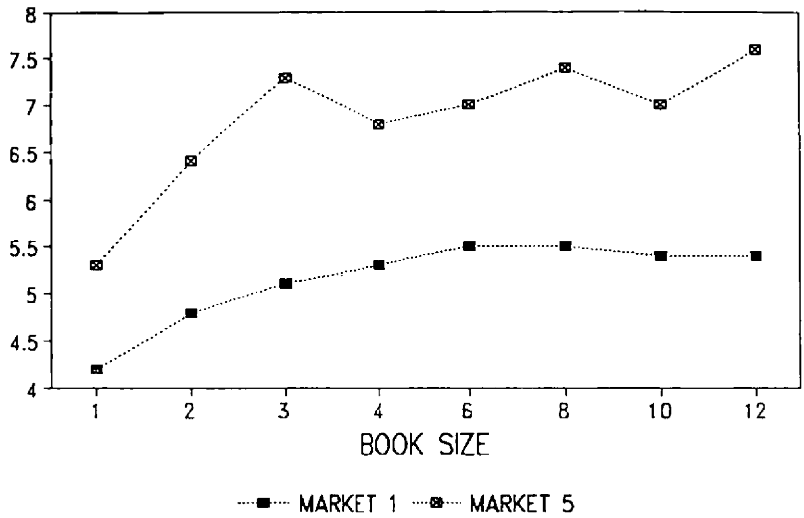


FIGURE 4 Liquidity, 24 identical traders.

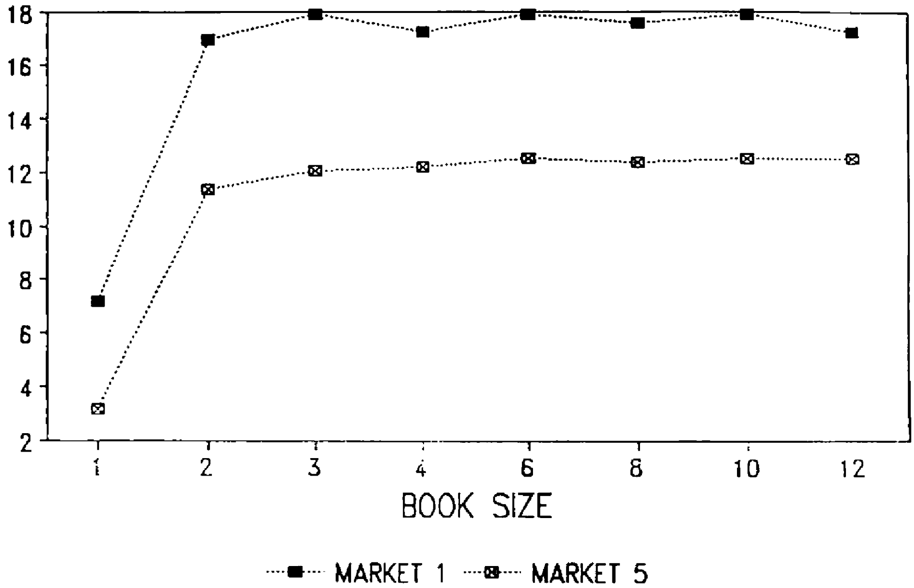


FIGURE 5 Order execution rate, 24 identical traders.

TABLE 10 Order Execution Rate[1]

Book Length	Market 1	2	3	4	5	6
1	14.0	15.7	14.8	41.3	31.6	25.5
	(2.9)	(3.4)	(3.2)	(11.9)	(7.6)	(12.8)
2	5.9	6.5	6.4	8.9	8.8	6.8
	(0.7)	(0.9)	(0.9)	(1.3)	(1.5)	(1.1)
3	5.6	6.2	6.3	8.6	8.3	6.6
	(0.8)	(0.7)	(0.9)	(1.3)	(1.5)	(1.0)
4	5.8	6.3	6.5	8.7	8.2	6.5
	(0.8)	(1.0)	(1.0)	(1.3)	(1.3)	(1.0)
6	5.6	6.2	6.2	8.6	8.0	6.4
	(0.8)	(0.9)	(0.8)	(1.4)	(1.3)	(0.9)
8	5.7	6.3	6.2	8.6	8.1	6.6
	(0.9)	(0.8)	(0.8)	(1.4)	(1.2)	(0.9)
10	5.6	6.2	6.3	8.7	8.0	6.6
	(0.8)	(0.9)	(0.9)	(1.6)	(1.1)	(1.1)
12	5.8	6.3	6.1	8.4	8.0	6.4
	(1.0)	(1.0)	(0.8)	(1.2)	(1.1)	(1.0)

[1] Standard errors in parentheses.

should be small. Treynor[29] models this cost as a function of the order execution rate, which is a measure that we can record from these experiments. We give this rate in Table 10 as the number of orders (buy and sell) that result in a trade, on average. The order execution rate improves by 58% and 72% in markets 1 and 5, respectively, as the book length goes from the polar case of one to a length of two. As seems obvious from an examination of Figure 5, which graphs the inverse of the execution rate reported in Table 10, there is no detectable additional improvement in the order execution rate as the book lengthens further.

4. CONTINUOUS TRADING IN A SIMULATED MARKET

We now turn to a model better suited for the investigation of the time-series properties of transactions prices and the bid-ask spread. Generation of bids and offers as inputs to the Globex algorithm with varying book length is based on the simulation model of CMSW.[5,8][19] We describe here the offer generation process; the generation of bids is completely analogous.

Offers arrive according to a Poisson process with mean arrival rate of 100 orders per period. The use of a Poisson arrival process follows the theoretical models of Garman[15] and CMSW.[6] Offer quantities are generated from a standardized Gamma distribution with the mean set equal to ten contracts.

Offer prices follow the process:

$$\log P_t^a = \log P_t^{mb} + e_t^a , \tag{1}$$

where P_t^{mb} denotes the current market's best bid, and e_t^a is an innovation drawn from the Yawl distribution of CMSW.[5,8] The density for e_t^a is given by:

$$f^a(e_t^a) = \begin{bmatrix} 0 & e_t^a < -3 \\ Y_t + (\frac{Y_t}{3})e_t^a & -3 \le e_t^a < 0 \\ (\frac{Y_t}{S_t})e_t^a & 0 \le e_t^a < S_t \\ Y_t - (\frac{Y_t}{2})(e_t^a - S_t) & S_t \le e_t^a < 2 + S_t \\ 0 & 2 + S_t \le e_t^a \end{bmatrix} \tag{2}$$

where $S_t \equiv \log P_t^{ma} - \log P_t^{mb}$ now defines the market spread, and $Y_t \equiv 2/(5 + S_t)$ in order to guarantee that the density integrates to unity. Variation in the numerical values for the parameters characterizing the shape of the offer innovation distribution makes little difference in the results reported below.[20] The shape and the moments of this density depend nontrivally on the logarithmic spread of the current market's best bid and ask quotations at the time of offer generation. This feature is a major reason for the choice of an innovations density of this form, and relates both to models of trading activity and to our interest in identifying a mechanism for the transmission of differential price volatility effects via forms of market institutions defined by the length of the order book. Feedback from best market quotes to the generation of bids and offers is consistent with the models of Grossman and Stiglitz,[19] Glosten,[16] and Madhaven.[23] This feedback is modeled here by a density of orders linked to current market prices through the spread via Eq. (2) and directly to current quotes via Eq. (1). In the Glosten and Madhaven models, in particular, traders use information contained in the trading-price history, including best bids and asks, to set current quotes. In equilibrium, new quotes are related linearly to the previous quote, similar to Eq. (1). Madhaven summarizes this conditioning by stating that the size of the market spread forms a sufficient statistic for the entire history of trading. In the model used here, bids and offers are

[19]CMSW[8] modeled the agency/auction structure of the specialist market, and the work in CMSW[5] allowed for the generation of market, as well as limit, orders. We are concerned with the specific structure of Globex and its book, and rely only on the general setup of CMSW with respect to order generation.

[20]See Bollerslev and Domowitz[4] for investigation of parameter variation in the polar cases of book length one and infinity.

conditioned on the spread by simply making the bid and offer distributions directly dependent on the spread. Finally, the shape of the innovations density is consistent with the order placement strategy results of CMSW.[7]

Our interest in transmission mechanisms is reflected in the fact that the variance of the offer distribution conditioned on current period information is a function of the spread; i.e.,

$$\begin{aligned}\operatorname{var}(e_t^a) =& \frac{1}{12} Y_t[35 + 14S_t + 9S_t^2 + 4S_t^3 + S_t^4] \\ &- \frac{1}{36} Y_t^2[169 + 156S_t + 88S_t^2 + 24S_t^3 + 4S_t^4]\,.\end{aligned} \tag{3}$$

If the spread between the best bid and offer is constant or time invariant, the offer innovations process is independent and identically distributed. Conditional on the current market spread, however, the offer process is heteroskedastic, but its second moments are not serially correlated so long as the market spread is independently distributed through time. The market mechanism that processes bids and offers into transaction prices may, however, transmit some degree of serial correlation to the spread process. The amount of such serial correlation in turn may depend on the length of the electronic book. We investigate the extent to which this phenomenon does indeed occur, and the degree to which such serial correlation in the second

TABLE 11 Transactions Returns Summary Statistics

	Book Length							
	1	2	3	4	6	8	10	12
ρ_1	-.231	-.090	-.067	-.052	-.052	-.050	-.057	-.049
ρ_2	.035	-.058	-.066	-.067	-.088	-.099	-.084	-.100
ρ_3	-.016	-.012	-.010	-.036	-.027	-.025	-.020	-.021
ρ_4	.006	-.004	-.021	.031	-.013	-.009	-.036	-.030
Q_{20}	578	141	111	120	130	152	153	153
$\rho_1^{(2)}$	-.060	.100	.116	.095	.134	.134	.122	.117
$\rho_2^{(2)}$	.021	.051	.100	.077	.126	.177	.105	.130
$\rho_3^{(2)}$	-.005	.019	.054	.069	.117	.100	.113	.086
$\rho_4^{(2)}$	.008	.034	.039	.030	.066	.088	.101	.082
$Q_{20}^{(2)}$	60	167	299	226	709	904	701	662
b_3	-.016	-.032	.004	-.005	-.010	.092	.145	.104
b_4	2.804	4.157	5.099	5.848	6.531	7.517	7.380	7.577
μ	-.009	-.001	-.001	-.009	.007	.007	-.004	-.002
σ^2	1.788	1.130	.933	.807	.764	.750	.656	.649

moments of the market spread translates into serial correlation in the conditional variance of transactions prices. The simulation proceeds by generating bids and offers according to the model above for transmission to the Globex algorithm with various book lengths corresponding to those investigated in the previous section. Separate series are generated for entry into each market mechanism as differentiated by book length, because of the feedback from the order generation process through the market spread. Series of the best bids, best offers, the bid-ask spread, and transaction prices are recorded and matched in terms of transaction-price time. A total of 10,000 transactions is analyzed for each book length.

Testing of the price series, as reported in Bollerslev and Domowitz,[4] indicates the presence of a unit root in the logarithmic transaction price series. We therefore limit our attention to the percentage transactions returns in the statistical analysis of trading prices. Some pertinent statistics are contained in Table 11.

We first reaffirm the results of the last section with respect to volatility. Average transactions price returns volatility, measured by the unconditional sample variance of the 10,000 returns, declines by 57% as the length of the book is extended to six from one. About 65% of the decline is attributable to extension of the length of book from one to just two. Using the same bid/offer generation process, CMSW[5] recorded a 72% decline in returns volatility as the average life of a limit order increased from one to two days. Their simulation described the average age of an order according to an exponential distribution, but did not restrict the length of the limit order book.[21]

The likelihood of large transactions price changes also grows with the length of the book. Figure 6 contains a graph of the sample kurtosis, b_4, plotted together with the variance of transactions returns. While the sample kurtosis and skewness coefficients, b_3, indicate that returns from a book of length one are approximately normal, the kurtosis of returns grows by 232% by the time the book is of length six and becomes highly statistically significant, indicating a strong departure from normality.[22] There is little further increase in kurtosis once the book reaches a length of eight, however, while decreases in the unconditional returns variance seem exhausted as the book reaches a length of three. These results obviously suggest that the unconditional distribution of returns is non-normal with fat tails once two bids or offers at a time are allowed to rest on the book. One possible implication of this result is that returns now are characterized by a normal distribution with a time-varying variance. The extent of such time variation is investigated below.

[21] Also similar in spirit to the results here is the fact that in the CMSW[5] model, the decline in volatility going from an average order life of two days to three days was much less, approximately 15%; see Table 1 of their paper.

[22] Under the null hypothesis of independent and identically normally distributed returns, the sample skewness is distributed $N(0, 6/10,000)$ and the sample kurtosis is distributed $N(3, 24/10,000)$. See Jarque and Bera.[20]

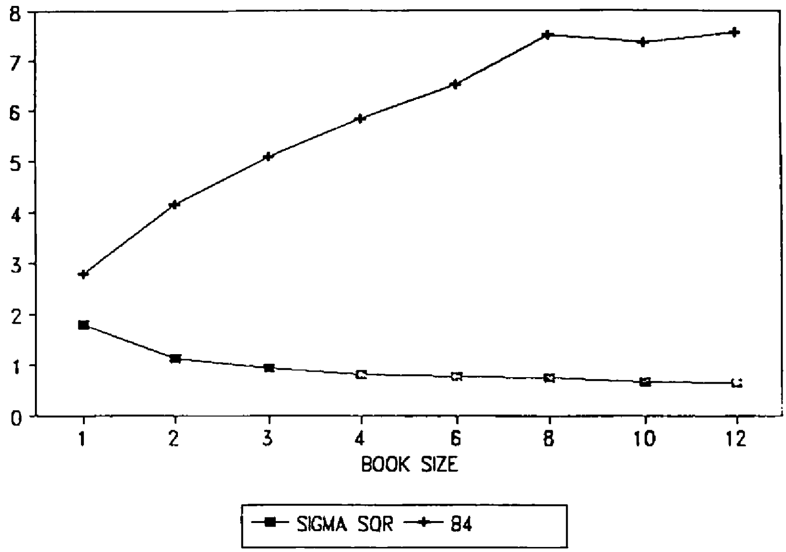


FIGURE 6 Sigma square and B4.

In a simulation allowing market as well as limit orders, CMSW[5] find an increasing degree of negative first-order serial correlation in returns as the average life of an order on the book is increased, which is similar in spirit to the lengthening of our book here. They explain this effect through the interplay of market and limit orders.[23] Upriced market orders executable at the current best bid or offer are not allowed on the Globex system, and once they are excluded from consideration, we find the opposite effect. A plot of the first-order serial-correlation coefficient of returns, ρ_1, against the length of book is presented in Figure 7.[24] In absolute value, the correlation coefficient decreases by 77% by the time the book length increases to six, while remaining slightly negative and statistically significantly different from zero. Almost 80% of this decline occurs as we pass from a mechanism with book length one to one generating prices with a book length of two. A similar pattern is observed for the Ljung-Box portmanteau statistic, Q_{20}, which registers serial correlation over twenty lags of transactions returns.[25]

[23] For example, they say that when the spread is unusually narrow, it is more likely to widen, because market orders become relatively more attractive (Cohen et al.,[5] p.733).

[24] Under the null hypothesis of serially uncorrelated returns, $\rho_i \sim N(0, 1/10{,}000)$.

[25] Following Ljung and Box,[22] $Q_{20} = (10{,}000)(10{,}002)\sum_{i=1}^{20} \rho_i^2/(10{,}000 - i)$.

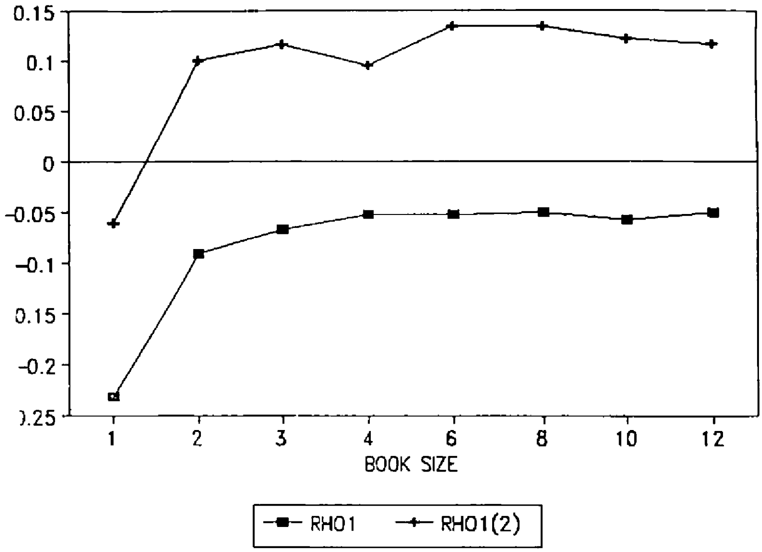


FIGURE 7 RHO1 and RHO1(2).

In sharp contrast, the Ljung-Box statistic, $Q_{20}^{(2)}$, for up to 20th-order serial correlation in the squared-returns process climbs sharply as the length of the book grows.[26] This indicates a very pronounced degree of serially dependent conditional heteroskedasticity in the returns process for mechanisms with a book with length greater than or equal to two. A graph of the first-order serial correlation coefficient of squared returns, $\rho_1^{(2)}$, against book length is given in Figure 7. The coefficient is negative but close to zero at book length one, rising to a highly statistically significant 0.134 by the time the book reaches length six. This serial correlation remains very persistent for higher order lags of the returns process. From Eqs. (1) and (3), the variance of the bid and offer processes is heteroskedastic conditional on the contemporaneous market spread. It is clear from (1) and (3), therefore, that higher degrees of serial correlation in the spread and its higher moments should result in correspondingly higher degrees of serial correlation in the conditional variances of bids and offers. This suggests that the serial correlation in the conditional variance of the transactions price returns may be traced to serial correlation in the spread.

[26]The $Q_{20}^{(2)}$ statistic is calculated as $Q_{20} = (10,000)(10,002)\sum_{i=1}^{20} \rho_i^{(2)}/(10,000 - i)$ where $\rho_i^{(2)}$ denotes the ith-order autocorrelation for the squared returns. Under the null of no-serial correlation in the squared returns, $Q_{20}^{(2)}$ is distributed chi-square with 20 degrees of freedom.

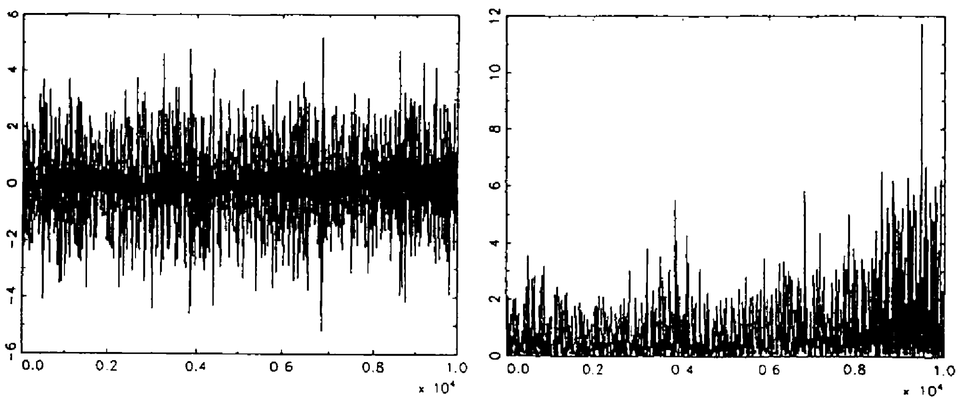


FIGURE 8 Book size = 1. Percentage returns (left graph) and bid-ask spread (right graph).

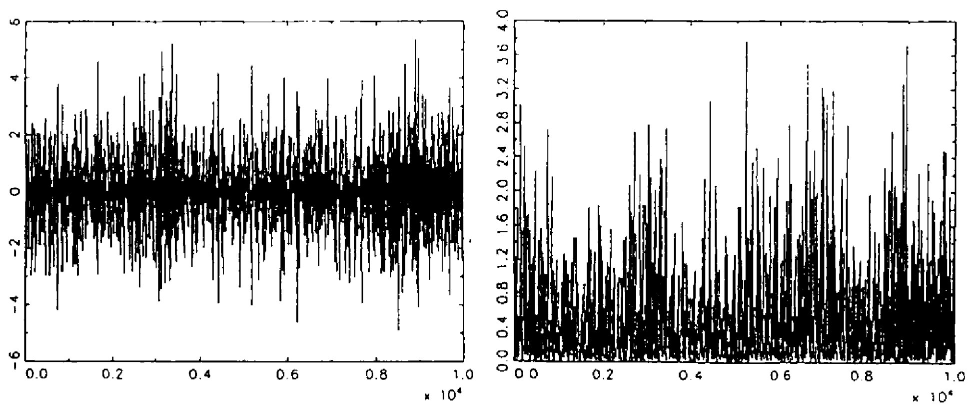


FIGURE 9 Book size = 2. Percentage returns (left graph) and bid-ask spread (right graph).

To illustrate this, we graph the bid-ask spread with transactions returns in Figures 8 through 11 for book lengths of 1, 2, 6, and 12.[27] In spite of the contemporaneous heteroskedasticity conditional on the current market spread, the lack of

[27]The downward drift in the spread for book length one as the simulation reported here progressed is a product of sampling variation. In other runs, including those reported in Bollerslev and Domowitz,[4] no drift was observed.

time-varying volatility in returns is evident in the graph for book length equal to one. A very low degree of serial correlation in the spread is documented in Bollerslev and Domowitz[4] for this case. The remaining figures illustrate that periods of high returns volatility are characterized by high bid-ask spreads. Increasing degrees of serial correlation in the market spread are evident as the book lengthens, and the relationship between the spread and returns variance is more pronounced as more orders are allowed onto the book.

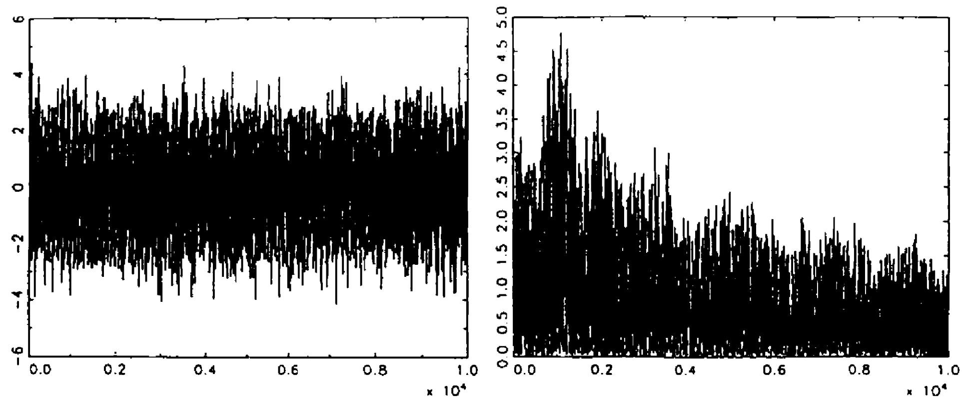


FIGURE 10 Book size = 6. Percentage returns (left graph) and bid-ask spread (right graph).

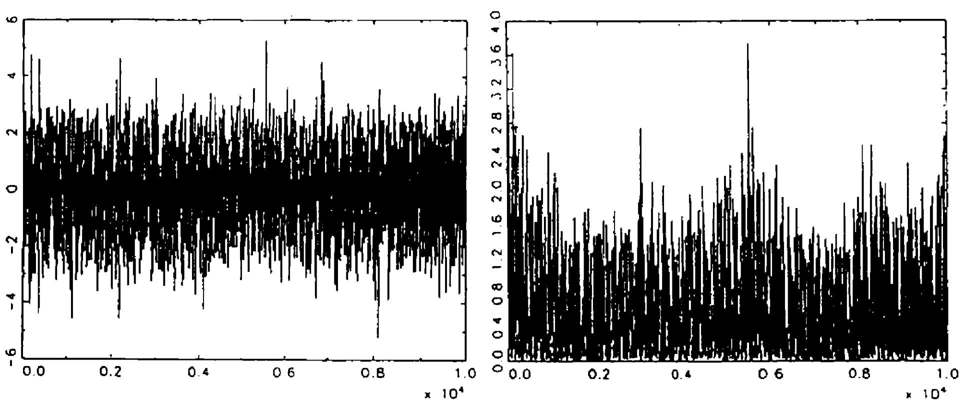


FIGURE 11 Book size = 12. Percentage returns (left graph) and bid-ask spread (right graph).

TABLE 12 Returns and Spread Correlations

	Book Length							
	1	2	3	4	6	8	10	12
$Corr(\Delta \log P_t, SP_t)$	-.013	-.019	.021	-.017	.027	.001	.017	.017
$Corr(\Delta \log P_t, SP_{t-1})$	.008	.001	-.010	-.017	.004	.002	-.012	-.003
$Corr(\Delta \log P_t, SP_{t-2})$	-.029	-.010	.007	-.016	.000	.002	-.011	-.001
$Corr(\Delta \log P_t, SP_{t-3})$	.003	.000	.014	-.003	-.001	-.003	-.017	-.004
$Corr(\Delta \log P_t^2, SP_t)$	.367	.366	.350	.301	.302	.325	.330	.305
$Corr(\Delta \log P_t^2, SP_{t_1})$	.022	.137	.164	.159	.221	.225	.220	.241
$Corr(\Delta \log P_t^2, SP_{t_2})$	.006	.108	.142	.118	.215	.258	.220	.246
$Corr(\Delta \log P_t^2, SP_{t_3})$	-.001	.045	.089	.094	.172	.168	.187	.174

The correlations between the spread and the levels of returns are basically zero, both contemporaneously and at lags 1 through 3. This is true for all book lengths as illustrated in Table 12, where the correlation between the transactions price returns and the corresponding market spread is reported for lags 0, 1, 2, and 3. We also report in Table 12 the sample correlations between the spread at various lags and squared contemporaneous returns. The correlation between the contemporaneous spread and squared returns is quite large and statistically significantly different from zero for all book lengths, and as expected, the magnitude does not vary much across the different books. The average of the correlations is 0.33, ranging from 0.37 to 0.30. Correlations between squared returns and lagged values of the spread are virtually zero for a book of length one, while such correlations increase very notably with the length of the book, with values stabilizing at a book length of about six.

We tentatively conclude that increasing the order book length induces an increasing amount of serial correlation into the market spread process. This serial correlation is translated into a serially dependent form of conditional heteroskedasticity in the bid and offer generation process through the model of order response to market conditions. The volatility movements of the order process are transmitted to transaction returns through the nonlinear filter of the automated trade execution mechanism. Although correlation properties do not always survive nonlinear filtering, we do find successively stronger correlations between squared returns and lagged spreads as the length of the order book increases. The transactions prices, determined through the design of the execution system and modified by the length of the order book, in turn determine the current best bid and offer, which condition the next round of bids and offers through the spread. Time-varying variance conditional on past information is thus endogenously determined as a function of the memory in the system provided by the limit order book.

CONCLUDING REMARKS

Using two very different models of trading activity, we find that the existence of an electronic order book in a limit order-based automated trade execution system has a stabilizing effect on the market in terms of price volatility, the magnitude of the bid-ask spread, and spread variability. This result is independent of behavioral and strategic considerations, and is consistent with work by CMSW[5] using a mixture of market and limit orders. We find, however, that the increases in negative serial correlation in transactions returns observed by CMSW as the average life of an order increases must be a function of the interaction between market and limit orders, as they suggest. The Globex system studied here does not permit market orders, and our results indicate a constant and small amount of negative serial correlation once an electronic order book is in place. Empirical work on intra-day trading in functioning financial markets also indicates a low degree of negative serial correlation in transactions price returns, and suggests that this correlation is not arbitraged away by trading activity.

The key variable in our analysis is the length of the limit order book, i.e., the maximum number of bids or offers allowed on the book for eventual execution. As the length of the book increases, price and spread volatility decrease, as well as the magnitude of the market spread, while certain measures of liquidity increase. A book of some nontrivial length improves market performance along these dimensions. These results do not depend on orders becoming "stale," in the sense that they rest on the book for long periods. Much of the change in market performance is observed with a book of length two, in which case orders in general only rest on the book for very short periods. In fact, market performance shows its full improvement by the time the length of the order book reaches four to six in virtually all cases examined. The more conventional measure of market efficiency, percent of total surplus extracted from the market, shows no systematic variation as the length of the book changes. This result coincides with that of Smith and Williams[27] in their experimental investigation of books of length one and infinity. Variability of transactions prices around the competitive clearing price does increase with book length, however. Finally, a nontrivial order book slows speed of convergence to equilibrium in environments allowing only transactions of a single unit at a time, but speeds convergence when multiple units are traded. This difference is traceable to the way an automated order book handles partially filled orders.

A novel finding of this paper is the link forged between the length of the order book and degree of serial correlation in the variance of transactions returns. Following Bollerslev and Domowitz[4] we model the possibility that temporal dependence in the variance of bids and offers conditional on past trading history can be induced through changes in the market bid-ask spread. Serially dependent forms of conditional heteroskedasticity can be transmitted through the automated trade execution mechanism to transactions prices. The length of the order book appears to modify the distribution of the best bid and offer outstanding in the market in such a way as to foster greater serial correlation in the variance of transactions returns as the book lengthens. We also confirm a higher degree of association between the spread and squared transactions returns as the book gets longer. This association is negligible when the book is only of length one, and in that situation we observe no serial correlation in the returns variance process. On the other hand, the degree of such serial correlation seems to reach its peak by the time the book is of length six. It is roughly constant, as measured by serial correlation coefficients, for book lengths greater than that.

We suspect, however, that in addition to the effects documented here, the book length may contribute also to serial correlation in the second moments of transactions returns in a fashion independent of the bid-ask spread. The spread is a measure of the market only in terms of the best bid and offer outstanding. Book length, however, helps determine the shape of the joint distribution of bids and offers on the book. It is this joint distribution that must determine the degree of serial correlation in the conditional volatility process. Further characterization of this link is left to future research.

Variations across market institutions in surplus extraction, speed of convergence to equilibrium, price volatility, liquidity, and the market spread all have direct implications for market efficiency. The results of Section 4 concerning serial correlation in variance have no such implications for a market in a single asset. Our conclusions suggest that particular variations in auction market structure may improve the predictability of transactions price volatility. Such predictability does not mean that one forecasts the direction of market movements with any certainty. On the other hand, these findings have implications for market efficiency in markets for more than a single security under several circumstances. A leading example is the case of an option on an underlying security, such as a stock or futures contract. Options valuation at any given time depends on the forecast of volatility in the price of the underlying security over the remainder of the life of the option. Improved predictability of this volatility suggests more efficient pricing in the options market. Similar arguments can be made for the market for swap instruments. The theory of portfolio risk management also suggests that predictions of increased volatility in the returns of a subset of assets will lead to portfolio adjustment. It follows that the price of securities in a multiple asset framework will respond to better predictability of returns variance. In this case, it may be possible to imagine arbitrage as a stabilizing force for the variance of transactions returns, although we know of no theory to that effect. In the case of options, however, there is certainly no theoretical or empirical evidence that suggests that price arbitrage across the security and options markets eliminates movements in the conditional variance of the underlying security.

This last comment brings up a caveat to the results presented in this paper that must be addressed. Arbitrage of the type discussed above is a form of strategy, and there are no strategic dimensions in this paper. Although the pricing in Section 4 depends explicitly on market information, our machine traders in Section 3 do not respond to institutional changes in their environment. The only signal they act upon is the status of their previous bid or offer on the book at the time they trade. In that case, they operate according to the maxim "the trend is your friend," and offer price improvement in some instances. It is certainly true that human traders may respond more intelligently to market signals, possibly with a corresponding improvement in some dimension of overall market efficiency in a given environment. Gode and Sunder[18] show that price volatility for machine traders is higher than that observed with human subjects, for example. Our interest is in comparisons of efficiency across environments differentiated only by different market institutions, however. There is evidence that such qualitative comparisons are somewhat robust to intelligence differences and strategic reactions to variations in environments. For the polar cases of book lengths of one and infinity, Smith and Williams[27] demonstrate efficiency and price volatility differences across institutions similar to the results in this paper, using human subjects. These subjects were carefully informed of the form of these institutions and had the opportunity to react. Domowitz and Wang[13] analyze market performance measures of the type used here in a setting that incorporates specific reactions both to the length of the book and to market information. In that context, surplus extraction does appear to be better in systems with shorter book lengths, but the speed of convergence to equilibrium improves as the length of the book increases. Although declines in the magnitude of the market spread as the book lengthens are not as striking when traders react to the size of the book, decreases in the spread still are observed. The results of this paper with respect to diminishing price volatility are completely preserved in the more strategic environment.

ACKNOWLEDGMENTS

We thank the National Science Foundation for financial support, and Jianxin Wang for excellent research assistance. John Rust and an anonymous referee offered a variety of comments helpful in the writing of this paper.

REFERENCES

1. Baillie, R. T., and T. Bollerslev. "Intra-Day and Inter-Market Volatility in Foreign Exchange Rates." *Rev. Econ. Stud.* **58** (1991): 565–585.
2. Beja, A., and N. H. Hakansson. "From Orders to Trades: Some Alternative Market Mechanisms." In *Impending Changes for Securities Markets: What Role for the Exchange?*, edited by E. Block and R. A. Schwartz. Greenwich: JAF Press, 1979.
3. Black, F. "Toward a Fully Automated Exchange." *Fin. Ana. J.* **July-August** (1971): 27, 29–35 and **November-December** 25–28,86–87.
4. Bollerslev, T., and I. Domowitz. "Price Volatility, Spread Variability, and the Role of Alternative Market Mechanisms." *Rev. Futures Markets* (1992): forthcoming.
5. Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb. "Limit Orders, Market Structure, and the Returns Generation Process." *J. Fin.* **33** (1978): 723–736.
6. Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb. "The Returns Generation Process, Returns Variance, and the Effect of Thinness in Securities Markets." *J. Fin.* **33** (1978): 149–167.
7. Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb. "Transactions Costs, Order Placement Strategy, and Existence of the Bid-Ask Spread." *J. Pol. Econ.* **89** (1981): 287–305.
8. Cohen, K. J., S. F. Maier, R. A. Schwartz, and D. K. Whitcomb. "A Simulation Model of Stock Exchange Trading." *Simulation* (1983): 181–191.
9. Cohen, K. J., and R. A. Schwartz. "An Electronic Call Market: Its Design and Desirability." In *The Challenge of Information Technology for the Securities Markets*, edited by H. C. Lucas and R. A. Schwartz. Homewood: Dow Jones Irwin, 1989.
10. Domowitz, I. "The Mechanics of Automated Trade Execution Systems." *J. Fin. Intermediation* **1** (1990): 167–194.
11. Domowitz, I. "Equally Open and Competitive: Regulatory Approval of Automated Trade Execution in the Futures Markets." *J. Futures Markets* (1992): forthcoming.
12. Domowitz, I. "Automating the Continuous Double Auction in Practice: Automated Trade Execution Systems in Financial Markets." This volume.
13. Domowitz, I., and J. Wang. "The Electronic Order Book and Automated Trade Execution in Futures Markets." Working Paper No. 108, Department of Finance, Kellogg Graduate School of Management, Northwestern University, 1991.
14. Engle, R. F., T. Ito, and W. L. Lin. "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market." *Econometrica* **58** (1990): 525–542.
15. Garman, M. B. "Market Microstructure." *J. Fin. Econ.* **3** (1976): 257–275.

16. Glosten, L. "Insider Trading, Liquidity, and the Role of the Monopolist Specialist." *J. Bus.* **62** (1989): 211–235.
17. Gode, D. K., and S. Sunder. "Efficiency of Double Auctions with Weakly Rational Traders: Market as a Partial Substitute for Individual Rationality." Working Paper No. 1990–15, Graduate School of Industrial Administration, Carnegie-Mellon University, 1990.
18. Gode, D. K., and S. Sunder. "Lower Bounds for Efficiency of Surplus Extraction in Double Auctions." This volume.
19. Grossman, S. J., and J. E. Stiglitz. "Information and Competitive Price Systems." *Amer. Econ. Rev.* **66** (1976): 246–253.
20. Jarque, C. M., and A. K. Bera. "Efficient Tests for Normality, Heteroskedasticity, and Serial Independence of Regression Residuals." *Econ. Lett.* **6** (1980): 255–252.
21. Kyle, A. S. "Continuous Auctions and Insider Trading." *Econometrica* **53** (1985): 1315–1335.
22. Ljung, G. M., and G. E. P. Box. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* **65** (1978): 297–303.
23. Madhavan, A. N. "Trading Mechanisms in Securities Markets." Working Paper No. 16–90, Rodney L. White Center for Financial Research, The Wharton School, University of Pennsylvania, 1990.
24. Niederhoffer, V., and M. Osborne. "Market Making and Reversal on the Stock Exchange." *J. Amer. Stat. Assoc.* **61** (1966): 897–916.
25. O'Brien, J. R., and S. Srivastava. "Liquidity and Persistence of Arbitrage in Experimental Options Markets." This volume.
26. Stoll, H. R. *The Stock Exchange Specialist System: An Economic Analysis.* Monograph 1985-2, Monograph Series in Finance and Economics, Salomon Brothers Center for the Study of Financial Institutions, New York University, 1985.
27. Smith, V. L., and A. Williams. "An Experimental Comparison of Alternative Rules for Competitive Market Exchange." In *Auctions, Bidding, and Contracting: Uses and Theory*, edited by R. Engelbrecht-Wiggins, M. Shubik, and R. M. Start. New York: New York University Press, 1983.
28. Stoll, H. "Principles of Trading Market Structure," Working Paper No. 90–31, Financial Markets Research Center, Vanderbilt University, 1990.
29. Treynor, J. "What Does It Take to Win the Trading Game?" *Fin. Anal. J.* (1981): 55–60.

Timothy N. Cason† and Daniel Friedman‡
†Department of Economics, University of Southern California, Los Angeles, CA 90089-0253 and ‡Economics Department, University of California at Santa Cruz, Santa Cruz, CA 95064

NINE
An Empirical Analysis of Price Formation in Double Auction Markets

This paper uses data from 30 previously conducted laboratory markets to investigate the price-formation process in continuous double auctions. Our analysis makes use of all market information—bids, asks, and transactions—and is based primarily on the theoretical models of Wilson[18] and Friedman[7] and the zero-intelligence algorithm from Gode and Sunder.[9] The models make essentially the same predictions regarding efficiency and transactions order, and the data generally support these predictions, though not always strongly. The models differ in their predictions regarding bid-ask sequences and regarding serial correlation in price changes. Our analysis finds greater support for the bid-ask sequences predicted by the Friedman model and the negative autocorrelation in price changes predicted by the Gode and Sunder algorithm.

INTRODUCTION

When market conditions change—for any reason, the introduction of new technology, the removal of trade barriers, or even the emergence of new information about pre-existing conditions—any previously established equilibrium is upset. Standard economic analysis presumes that a new equilibrium will be achieved in due course, perhaps more rapidly in some markets than in others, at which point the body of economic theory again becomes relevant. The process of price formation—*how* the market finds its way to a new equilibrium—therefore is literally fundamental to the application of most economic theory.

Price formation has direct practical as well as theoretical importance since many policy interventions, ranging from job retraining programs to the SEC reporting requirements in securities markets, are intended to influence the adjustment to equilibrium. For example, the New York Stock Exchange recently changed its rules to forbid program trading after a 50-point price change, with the announced intention of reducing volatility and increasing market efficiency. Since price formation is not well understood, it is debatable whether the new rule will achieve its worthy goals.

Indeed, economists have remarkably little to say about what factors will speed or impair the price-formation process. The standard theoretical tools of optimization and equilibrium are difficult to deploy in the absence of market-clearing and well-defined budget constraints. Standard empirical techniques are frustrated by the absence of field data on individual preferences, costs, and information; typically equilibrium must be assumed to interpret the data.

Laboratory data are more promising than field data as a source of evidence on price formation. Continuous double auction markets in the laboratory possess strong convergence properties[12,14] and provide very detailed performance data with controlled (induced) preferences, costs, and information. Such markets have been the focus of some recent theoretical work on price formation, surveyed in the next section. However, most previous empirical analyses of laboratory markets have focused on tests of market-equilibrium theories or across-period market-convergence properties, and generally have ignored the within-period price dynamics that are directly relevant in field environments.

We fill this gap by analyzing individual behavior—bids, asks, and transactions—during the price-adjustment process in continuous double auction laboratory markets. Most existing laboratory markets are repetitive stationary. That is, traders have the same endowments and marginal cost and valuation schedules in successive trading periods. Since double auction markets converge very rapidly, most transactions in stationary markets after the first period are at, or near, the equilibrium price. In order to have a significant number of observations of disequilibrium actions, we need nonstationary markets, in which traders must seek a new equilibrium price in each period. As explained in Section 3, we located two series of previously conducted experiments satisfying this requirement.

The data analysis relies primarily on the theoretical models of Wilson[18] and Friedman[7] and uses the zero-intelligence algorithm discussed in Gode and Sunder[9] as a "non-strategic" benchmark. Our initial analysis results suggested the zero-intelligence algorithm because it can explain the persistent negative price change autocorrelation observed in the data. Section 2 provides an informal summary of the models as well as a brief discussion of related literature on price formation. Section 3 describes the data.

Section 4 collects the results. The theoretical models make the same predictions regarding efficiency and transactions order, and the data generally support these predictions, though not always strongly. The models differ in their predictions regarding bid-ask sequences, and our analysis finds greater support for the bid-ask sequences predicted by the Friedman model. The models also differ in their predictions regarding serial correlation in price changes, and the negative autocorrelation in price changes observed in the data provides support for the Gode and Sunder algorithm. The final section interprets the findings and proposes new experiments designed specifically to investigate price formation.

THEORETICAL LITERATURE

Price formation, broadly construed as processes that lead to market equilibrium, has been the subject of innumerable theoretical discussions from the time of Walras and Marshall. The dominant tradition, at least since the second world war, has been to assume that markets are organized by a fictitious non-self-interested agent called an "auctioneer" who raises (lowers) prices on goods in excess demand (supply) until equilibrium is achieved. Only recently have theorists begun to consider price formation in the context of a viable trading institution. We are concerned here with the continuous double auction (DA), the trading institution most widely used in laboratory markets and field markets for homogeneous goods. In the DA, self-interested traders themselves announce buying prices (bids) and selling prices (asks) and transact by accepting other traders' bids or asks at any moment during a trading period.

Except for a little-known precursor, Garcia,[8] Easley and Ledyard[2] offer the first theoretical model designed to explain price formation in simple laboratory double auction markets. Their model postulates a reservation price for each participant which is not linked to induced value/cost parameters. They make plausible but *ad hoc* assumptions as to how participants adjust their reservation prices and offers within and across trading periods. The primary result is that after sufficiently many trading periods in which costs and values are held constant, transaction prices will always lie in a (usually narrow) interval that brackets the competitive equilibrium price. They also offer three testable predictions regarding (a) the range of transaction prices in successive trading periods, (b) the trading sequence, and (c) a lower bound on the number of transactions. The data do not strongly disconfirm any of these predictions, but only (b) is sharp enough to be useful for present purposes. It states that in any given trading period a subset of buyers (sellers) with the largest potential gains from trade will all transact before the remaining buyers (sellers); however, the transaction order within these sets is left unspecified, so the rank-order correlation of buyer valuation (seller cost) and transaction order should be negative (positive) but can be greater than -1.0 (less than 1.0).

MODELS TO BE TESTED

Our main concern in this paper is with price formation within a continuous double auction trading period, and we are aware of only three directly relevant models. In order of decreasing rationality, these models are the waiting game/Dutch auction (WGDA) model of Wilson,[18] the Bayesian game against nature (BGAN) model of Friedman,[7] and the zero-intelligence (ZI) algorithm discussed in Gode and Sunder.[9]

Wilson[18] regards the price-formation process as a sequential equilibrium of an extensive-form game in which the private values of n single-unit buyers and m single-unit sellers are drawn from a commonly known joint distribution. The basic idea is that agents play a waiting game, with each buyer's (seller's) impatience arising from the possible preemption of gains by the other buyers (sellers). Eventually some buyer (or seller) finally makes a "serious" bid (ask)—one which has a positive probability in sequential equilibrium of being accepted. If her offer is not immediately accepted, the bidder (asker) will steadily improve the offer (while other traders remain passive) until it is accepted, as in a Dutch auction.[1] The transactors will be the highest value buyer and the lowest cost seller remaining in the market. The gains from a transaction are split between buyer and seller according to relative hazard rates ("impatience"), so the ratio of remaining buyers to remaining sellers determines the split. With high probability enough transactions will occur to exhaust most potential gains from trade.

Appropriate tests of Wilson's waiting game/Dutch auction (henceforth WGDA) model are not immediately obvious. One might first consider comparing the entire set of event predictions (e.g., the model predicts that trader 3 bids \$1.87 at $t = 32.6$ seconds) to the events in an actual experiment. In principle the model is sufficiently precise, but there is a practical difficulty. Solutions to the WGDA model are defined implicitly by a nested set of partial differential equations (PDE's) whose boundary conditions ("continuation values") at each stage are derived recursively from the solutions to the subsequent stage PDE's, with some arbitrariness as to the final stage specification.[2]

No numerical algorithms presently are available to solve the equations, even for simple value distributions and auxillary assumptions. Hence explicit predictions regarding bid, ask, and acceptance behavior are not available at present for the WGDA model.

On the other hand, it would be inappropriate to begin with such specific tests even if the tests were feasible. The potential value of the WGDA model, or any specific model of price formation, should be apparent in its qualitative (or more aggregate quantitative) implications. Fortunately the WGDA model offers several striking general implications that are readily tested:

[1] In an unpublished alternative version, Wilson replaces the Dutch auction by a waiting game by the sellers (buyers). The alternative version seems less able to account for observed bid and ask behavior, but otherwise gives the same predictions as the original version. Hence, we confine our analysis to the original version.

[2] Continuation value is defined as the expected payoff to a player of participating in the remainder of the game if he or she does not transact in the current stage.

1W. Prices and profit distribution: The waiting game aspect implies that in each transaction the gains from trade in excess of continuation values will be split between buyer and seller in the ratio of the number of other remaining sellers to other remaining buyers. In particular, the buyer/seller ratio of realized profits will increase (decrease) in successive transactions when initially more sellers (buyers) are present. Furthermore, to preclude intertemporal arbitrage, the best predictor of future transaction prices is the current transaction price; i.e., prices follow a martingale. This martingale property implies that price changes are serially uncorrelated.

2W. Bid and ask behavior: The Dutch auction aspect typically produces successive improvements on a bid (ask) by a given buyer (seller) culminating in an acceptance by a seller (buyer). On the other hand, the waiting game aspect implies that successive improvements by *different* buyers (sellers) will be rare. These implications apply to "serious" bids and asks, i.e., those with *a priori* positive probability of acceptance.

3W. Transactions partners: Early (later) transactions will be between high-value (low value) bidders and low-cost (high cost) sellers. Assuming risk neutrality and symmetric expectations, the rank-order correlation of buyer valuation and transaction order should be -1.0, and the rank-order correlation of seller cost and transaction order should be 1.0. This is a stronger version of the Easley-Ledyard prediction (b).

4W. Efficiency: Most potential gains from trade will be exhausted; any unrealized profitable transactions will be those offering the smallest gains. Easley and Ledyard make the same prediction, but only for later trading periods under stationary repetition.

The second model draws on the Bertrand perspective of Friedman,[5] which assumes traders carry reservation prices for buying or for selling. A buyer seizes the market bid if she can do so at a price better than her reservation price, and she accepts the market ask whenever it exceeds her reservation price; sellers are analogous. To complete the model, the dependence of reservation prices on time and history must be specified. In Friedman,[7] this is done by means of a drastic simplification: buyers and sellers are assumed to ignore the impact that their own current bids and asks will have on subsequent offers by others. This "game against nature" assumption, together with Bayesian updating and some auxillary assumptions similar to Wilson's (e.g., risk neutrality and proportional time parameterization) gives reservation prices as solutions of the optimal stopping problem associated with current parameter estimates for "Nature's" bid- and ask-generating process. The reservation prices in turn permit numerical calculation of bids, asks, and acceptances for any set of induced value parameters, random or otherwise. More generally, this Bayesian game against nature (BGAN) model agrees with implications (3W) and (4W) of Wilson's WGDA model but differs sharply from (1W) and (2W):

1B. Successive price changes in the BGAN model will be positively correlated.[3] This effect will lessen as traders beliefs become more firmly established, i.e., later in a trading period.

2B. The Bertrand aspect of the BGAN model implies that successive improvements to the market bid (ask) by a given buyer (seller) will be rare, while successive improvements by different buyers (sellers) will be common.

Gode and Sunder,[9] and chapter 7 of this volume, convert an oral tradition regarding brainless trader behavior into a well-specified algorithm called zero-intelligence (ZI) traders. In the ZI model, each bid is uniformly distributed between zero and the buyer's redemption value, and seller's asks are similarly distributed between cost and an *a priori* upper bound on buyers' values. Under the usual "NYSE convention" for the DA, bids (or asks) which do not improve the current market bid (or ask) will not be observed. Buyers bid and sellers ask independently at random times. A transaction occurs at the current market ask (or market bid) whenever a new bid exceeds (or a new ask falls below) the current market ask (or market bid).

These ZI assumptions imply that transaction prices are independent draws from a distribution which changes over time as successful transactors exit from the market. If prices are independent draws from a *fixed* distribution, then successive price changes have a correlation coefficient τ of precisely -0.5.[4]

The changes in the distribution have only a small effect on τ. Hence, we have the following general implications for the ZI model:

1Z. Successive price changes will have correlation of about -0.5.

2Z. Successive improvements to the market bid (ask) by a given buyer (seller) will occur with a probability which depends on the level of the market bid (ask) and all buyers' values (sellers' costs). The probability can be readily estimated by Monte Carlo methods.

3Z. There will be a slightly greater transaction probability for higher valued buyers and lower cost sellers.

[3]The martingale property fails in the BGAN model because traders incorrectly assume that other traders' bids and asks arise as if from "natural" processes whose parameters are unknown but unchanging. To the extent that traders' beliefs change in response to a new bid or ask observation, their reservation prices will shift, and as a result, the actual data-generating process will also shift. It can be shown that such an unanticipated shift in the data-generating processes leads to positively correlated transaction price changes. This effect dies out fairly rapidly as traders accumulate observations and therefore change their beliefs less in response to new observations. A second possible reason for positive serial correlation (probably negligible in practice) is an asymmetry in variance estimates for bid versus ask distributions. We know of no arguments within the BGAN model for *negative* correlation in transaction price changes.

[4]**Proof:** Suppose transaction prices P_t are IID with (finite, positive) variance V. Normalize P_t so $EP_t = 0$. Then $V = E_t(P_{k+t})^2$ and $E_0 P_t P_{t+1} = E_0 P_t E_t P_{t+1} = 0$. The correlation coefficient is $\tau = E_0(P_{t+1} - P_t)(P_t - P_{t-1})/E_0(P_{t+1} - P_t)^2$. The numerator of τ is $E_0(-P_t)^2 = -V$ and the denominator is $E_0(P_{t+1})^2 + E_0(P_t)^2 = 2V$. Therefore $\tau = -0.5$.

Hypotheses:	WGDA	BGAN	ZI
1. Price Change Autocorrelation	Zero	Positive	Negative (near -0.5)
2. Bid/Ask Improvements	By same buyer/seller	By different buyers/sellers	Intermediate (estimated)
3. Transaction Order	In order of valuations	In order of valuations	Weakly in order of valuations
4. Market Efficiency	High	High	High

FIGURE 1 Summary of model implications.

4Z. Most potential gains from trade will be exhausted, but unrealized gains need not be the least profitable. The expected efficiency can be readily calculated from the arrival rate of bids and asks and from the value and cost parameters.

The four implications of each of the three models are summarized by Figure 1. The WGDA model implies serially uncorrelated price changes, while the BGAN and ZI models imply positive and negative serial correlation in price changes, respectively. Each model also has distinct predictions regarding the bid and ask behavior. The WGDA model implies that successive improvements on a bid (ask) will be typically made by a given buyer (seller), and the BGAN model implies that successive improvements on a bid (ask) will be typically made by different buyers (sellers). The ZI model predicts both kinds of bid and ask improvements. All three models predict that high-value buyers and low-cost sellers will transact earlier in the period than low-value buyers and high-cost sellers; however, the WGDA and BGAN models both imply a stronger form of this hypothesis than the ZI model. Finally, all three models predict that most potential gains from trade will be exhausted.[5]

THE DATA

Vast numbers of double auction market experiments have been conducted in many laboratories. Unfortunately for us, most of the experiments involve stationary repetition or fixed (deterministic) shifts: all traders receive the same endowments and induced preferences in most successive trading periods. Thus, the equilibrium price (or equilibrium price interval) also remains the same across trading periods. In this case the experimenter has only a single observation of price formation, often spread across many trading periods. To focus cleanly on the within-period process, we need data from experiments in which equilibrium prices shift unpredictably between trading periods. We found two series of experiments meeting this requirement, both conducted using the PLATO computerized double auction. All experiments employed inexperienced subjects.

In PLATO double auctions both buyers and sellers are free at any moment during the trading period to initiate price quotes (bids to buy and offers to sell) for a single commodity unit by typing in a number and then touching a box shaped area on their display screen. The market (highest) bid and market (lowest) ask are displayed continuously on each buyer and seller's screen. Any seller (buyer) is free to accept the market bid (ask) by first touching a box labelled "ACCEPT" and then touching another box labelled "CONFIRM" within five seconds. After the transaction is confirmed, it is recorded in both the buyer's and seller's private record sheet and the price is publicly displayed. All transactions and certain bids and asks are public information (i.e., appear on all subject's display screens). The double auction rules used for the experiments reported here included both the New York Stock Exchange "improvement rule" and a computerized "specialist book." The improvement rule requires each ask (bid) to be lower (higher) than the current market ask (bid) for it to be announced. An ask (bid) that is higher (lower) than the market ask (bid) is placed in a queue ("specialist book"), ordered with lower asks (higher bids) having priority. After acceptances of the market bid (ask), the highest priority bid (ask) in the queue automatically becomes the new market bid (ask). Additional details of the PLATO trading procedures are provided elsewhere; see, for example, Smith and Williams.[16]

[5] To be more precise, the WGDA model predicts that market inefficiencies should only be realized if some transactions offering the smallest gains do not occur, while the ZI model predicts that most market inefficiencies should occur when extra-marginal units trade.

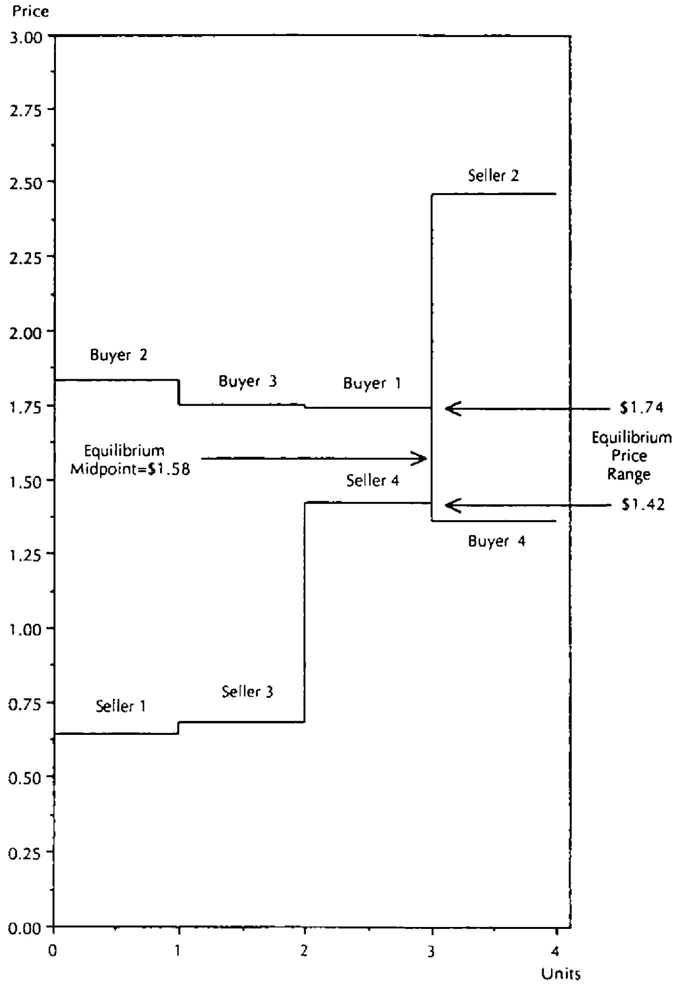


FIGURE 2 Example of experiment series 1 buyer and seller valuations: Period 1.

SERIES 1

The first series of experiments conforms closely to the assumptions of the WGDA model. In each period, each agent is endowed with a single unit whose valuation is random but whose distribution is common knowledge. To date, three such experiments have been conducted, all at the University of Arizona. Eight subjects (four buyers and four sellers) participated in each 15-period experiment, with 120 seconds per period. Sellers' unit costs were drawn independently from a uniform distribution with support [\$0.00, \$2.50], and buyers' redemption values were drawn independently from a uniform distribution with support [\$1.00, \$3.50].[6] Because valuations were drawn randomly, the equilibrium price range shifted each period; the support of the equilibrium distribution is clearly [\$1.00, \$2.50]. The midpoint of this range varied between \$1.21 and \$2.41 for the 15 periods. Figure 2 plots the induced supply and demand arrays for Period 1 of this series. The equilibrium price range is \$1.42 to \$1.74, with midpoint $P_m = \$1.58$.

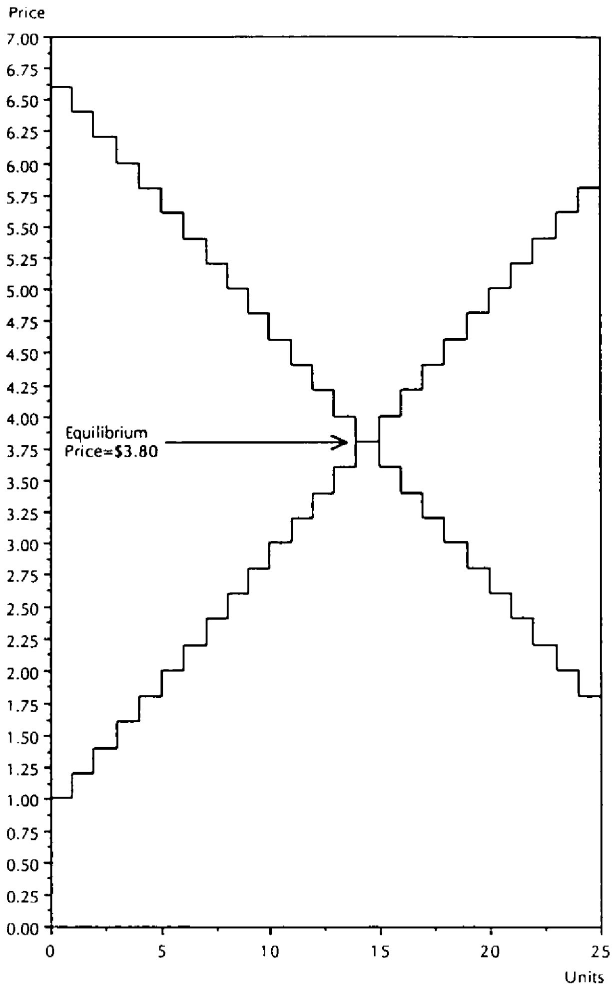


FIGURE 3 Example of experiment series 2 buyer and seller valuations: Period 1.

[6] All three experiments used the same random sequences of costs and values. For the third experiment (3pd206), all costs and values were shifted up by \$1.00.

SERIES 2

The second series of experiments has characteristics more typical of laboratory double auctions, but some characteristics violate assumptions of the WGDA and BGAN models. For example, each agent has a trading capacity of five units per period (although each typically traded only two or three units per period), and no information is provided regarding other traders' unit valuations. The unique feature of these experiments that make them useful for this research is that traders' valuations (and therefore the equilibrium price) shifted every period.

The 27 experiments were conducted by Jim Cox and Ron Oaxaca at the University of Arizona to train subjects for their research evaluating econometric estimators (see Cox and Oaxaca).[1] These trainer experiments were used to expose subjects to the randomly shifting induced supply and demand functions employed in their experimental design, and the results are not reported by Cox and Oaxaca.

Ten subjects (five buyers and five sellers) participated in each six-period experiment, with 360 seconds per period. Sellers' unit costs fit along the following induced inverse supply curve:

$$P_s = -0.2 + 0.2Q + 1.0X_s + U_s\,, \tag{1}$$

and buyers' redemption values fit along the following induced inverse demand curve:

$$P_d = 0.2 - 0.2Q + 1.0X_d + U_d\,, \tag{2}$$

where the P's and Q's are prices and quantities, respectively, X_s and X_d are supply and demand stocks, respectively, and U_s and U_d are IID random terms drawn from a uniform distribution with support $[-0.4, 0.4]$. Because of the random supply and demand shocks, the equilibrium price range shifted each period.[7] The midpoint of this range varied between \$3.20 and \$6.60 for the six periods. Figure 3 plots the induced supply and demand arrays for Period 1 of this series, with a equilibrium midpoint $P_m = \$3.80$. The "steps" on the supply and demand curves occupied by each trader were also changed each period. Each seller (buyer) had exactly one unit among the five lowest cost (highest-valuation) units on the supply (demand) curve, exactly one unit among the five cost (valuation) units ranking six through ten, and so on; see Figure 3. All 27 experiments used the same set of valuation parameters.

RESULTS

The results are presented in four subsections, each corresponding to the four implications described above in Section 2.

[7]The valuations were not shifted between periods five and six.

TRANSACTION PRICE CHANGES

WGDA HYPOTHESIS (1W): Transaction price changes will not be serially correlated because prices follow a martingale to preclude intertemporal arbitrage.

BGAN HYPOTHESIS (1B): Transaction price changes will be positively correlated.

ZI HYPOTHESIS (1Z): Transaction price changes will have an autocorrelation, coefficient of -0.5.[8]

The null hypothesis from the WGDA model (1W) is that transaction prices P_t follow a martingale:[9]

$$E[P_t | P_{t-j}, j > 0] = P_{t-1}. \tag{3}$$

If we reject this hypothesis in favor of positive price change autocorrelation, we have support for the alternative Hypothesis (1B), and if we reject this hypothesis in favor of negative price change autocorrelation, we have support for the alternative Hypothesis (1Z). In particular, we shall test the hypothesis that prices follow a second-order martingale:

$$P_t = P_{t-1} + u_t, \text{ where } E[u_t] = 0 \text{ and Cov } (u_t, u_{t-s}) = 0 \text{ for all } s \neq 0. \tag{4}$$

Rewrite Eq. (4) in terms of price changes as follows:

$$u_t(T) = P_t - P_{t-1}, t = 1, 2, ... \tag{5}$$

where $u_t(T)$ is the price change over the interval T. For the following results, an interval T is given by one transaction.[10] The null hypothesis given above in Eq. (4) implies that[11]

$$\text{Cov } (u_j(T), u_{j-s}(T)) = 0 \text{ for all } s \neq 0. \tag{6}$$

[8] The model in Roll[13] also predicts that transaction price changes in an efficient market in equilibrium will be negatively correlated. This result occurs because, assuming no new information arrives, prices will oscillate in the bid-ask spread from accepted bid to accepted ask. However, the time series of price changes from each accepted ask to the next accepted ask (and accepted bid to accepted bid) should be serially uncorrelated.

[9] For similar tests using the time series of securities prices, see the classic paper by Fama,[3] or for surveys of related research on stock returns, see Granger and Morgenstern[11] or Fama.[4]

[10] Results are essentially unchanged, and thus not reported here, when examining price changes over nonoverlapping intervals of more than one transaction or fixed time intervals such as one minute.

[11] In practice, prices P_t are often transformed to $\hat{P}_t = \log(P_t)$ because (among other reasons) the series P_t is unbounded from above but is bounded below by zero, while $\log(P_t)$ is symmetrically unbounded. All of the tests reported below were also conducted for $\log(P_t)$ with no significant impact on the results.

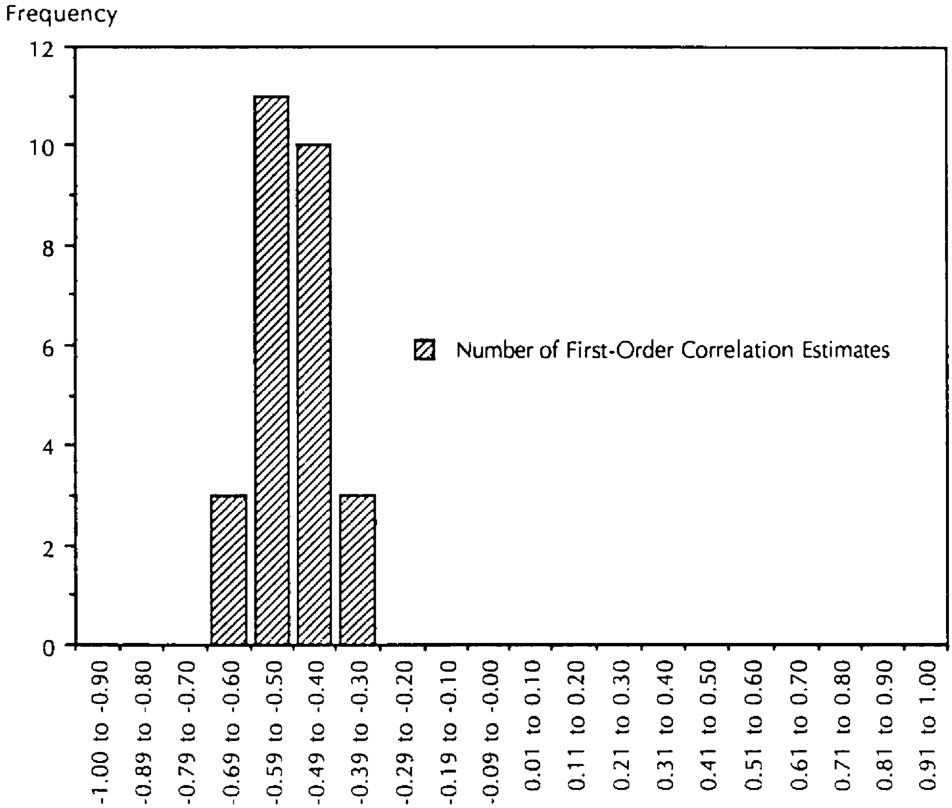
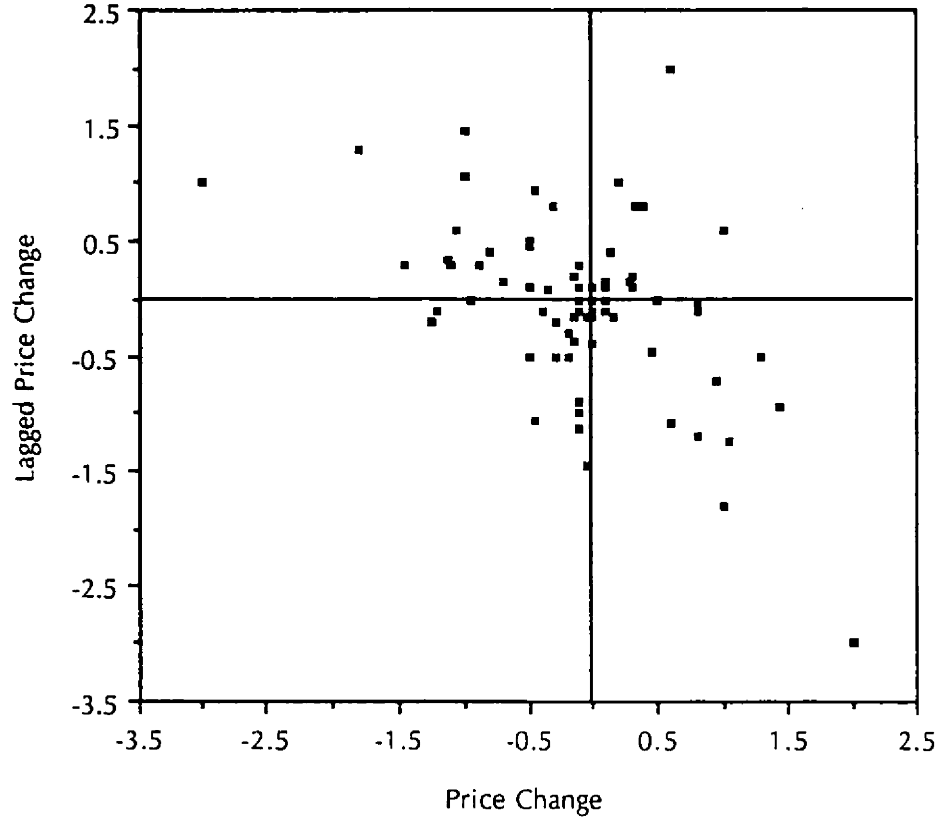


FIGURE 4 Frequency distribution of first-order correlation coefficients for price changes: Experiment series 2.

EXPERIMENT SERIES 1: Note that this hypothesis applies to *within*-period price changes. The models make no predictions for between-period price changes, so for the estimates below we exclude the price change that occurs at the first transaction each period. Because of this restriction, a period with n transactions contributes no more than $n-2$ observations for the correlation estimates. The Series 1 experiments typically had only two or three transactions per period, so they contributed only about one observation per period for this test. The estimated first-order correlation coefficients for price changes are 0.43 ($n = 11$) for experiment 3pd204, -0.14 ($n = 16$) for experiment 3pd205 and -0.43 ($n = 13$) for experiment 3pd206. None of



Note: Estimated Correlation = -0.5

FIGURE 5 Scatter plot of price change against lagged price change: Experiment AZTRO8.

these estimates are significantly different from zero.[12] Because the small number of observations in the Series 1 experiments severely limits the power of the statistical tests, the remainder of this section reports results for the high-volume Series 2 experiments.

[12]This first-order correlation statistic was also calculated based on a simulation of 1000 markets populated with only ZI traders using the Series 1 parameters. Because the transaction prices are drawn from a distribution that changes slowly over time, the mean estimate of this correlation is greater than −0.5 (specifically, −0.34); 90% of the 1000 estimates are within the range −0.68 and 0.09. One of the three Series 1 experiments has an estimated correlation outside of this range.

EXPERIMENT SERIES 2: When estimated separately for each Series 2 experiment, the first-order correlation coefficient of price changes for successive transactions (Eq. (6) with $T = 1$) is always negative, significantly different from zero and near -0.5. Figure 4 presents the distribution of the first-order correlation coefficients for the 27 experiments. The estimated correlations range between -0.30 and -0.69, and nearly all fall within the range -0.4 and -0.6 and are not statistically different from -0.5. Figure 5 illustrates this negative correlation with a scatter plot of each price change against each *lagged* price change for experiment aztr08. This experiment has an estimated correlation coefficient of -0.50, and the correlation coefficient is also -0.50 when pooled across all 27 experiments ($n = 1980$). We estimated the first-order serial correlation by period to determine if it decreased later in the experiments due to learning or some other factor. The estimated coefficients for the pooled experiments are $-0.50, -0.51, -0.49, -0.50, -0.44$, and -0.55 for periods 1 through 6. All are *extremely* different from zero but not statistically different from -0.5.

As mentioned previously in footnote 8, the model in Roll[13] predicts negative serial correlation in price changes but zero serial correlation in price changes *from accepted ask to accepted ask* and *from accepted bid to accepted bid.* We test this prediction by estimating the first-order serial correlation coefficients for price changes between accepted asks and between accepted bids. Pooling across all Series 2 experiments, the estimated coefficients are *both* -0.44 for ask- and bid-acceptance price changes, and both are significantly different from zero at the one-percent level. According to the Roll model, these price changes should follow a martingale (i.e., have zero serial correlation), so the Roll model does not explain the price change negative autocorrelation result.

This strong negative correlation result for price changes provides support for the zero-intelligence model (Hypothesis (1Z)) and implies that price changes are not unpredictable; if the price just increased by \$0.50, a forecast that the next price will be about \$0.25 below the current price is more accurate than the (martingale) forecast that the next price will be equal to the current price (see Eq. (3)). Traders identifying this pattern can increase their expected trading profit. Because of the existence of these intertemporal arbitrage opportunities, we estimated the first-order serial correlation coefficient in price changes for some other computerized double auction markets to determine if this result is specific to these randomly shifting supply and demand experiments with inexperienced subjects. In particular, we examined the first period of seven experiments from Smith and Williams,[15] all using inexperienced subjects, the first two periods[13] of eight experiments from Williams and Smith,[17] all using experienced subjects, and all twelve periods of one asset market experiment with experienced subjects from Friedman.[6]

With the inexperienced subjects in the 1982 study, the negative and significant serial correlation persists (estimate equal to -0.43, $n = 53$, significantly different from zero at the one-percent level). With experienced subjects in the 1984 study

[13]The first two periods of these experiments provide useful data because the supply and demand shifted between periods.

and no "speculators" that can buy *and* sell units, the estimated coefficient is still negative but is no longer significantly different from zero (estimate equal to −0.26, $n = 32$). When adding speculators that can buy and sell units and carry them between periods, the estimated coefficient increases to −0.24 ($n = 57$), also not significantly different from zero. In the Friedman asset market, the first-order serial correlation of price changes was significantly negative at −0.26 ($n = 172$). This cursory examination of other double auction data suggests that subject inexperience may be one source of the negative serial correlation in price changes, and that adding speculators that can buy and sell units may further decrease the negative correlation.

CONCLUSION. Experiment Series 2 provides strong evidence that price changes are negatively serially correlated, so that both WGDA Hypothesis (1W) and BGAN Hypothesis (1B) can be rejected in favor of ZI Hypothesis (1Z).

BID AND ASK BEHAVIOR

WGDA HYPOTHESIS (2W) Successive improvements on a bid (ask) by a given buyer (seller) culminating in an acceptance by a seller (buyer) will be common, and successive improvements by different buyers (sellers) will be rare. This prediction applies to "serious" bids and asks.

BGAN HYPOTHESIS (2B) Successive improvements on a bid (ask) by a given buyer (seller) will be rare, and successive improvements by different buyers (sellers) will be common.

ZI HYPOTHESIS (2Z) Successive improvements on a bid (ask) by a given buyer (seller) will occur with a probability which depends on the level of the bid (ask) and all buyers' values (sellers' costs).

In order to test these alternative hypotheses, we first specify the criteria for regarding a transaction observation as inconsistent with the BGAN and WGDA models:

DEFINITION 1 ("NOT BGAN") A transaction is regarded as inconsistent with BGAN if any outstanding bid or offer since the previous transaction is updated by the same agent currently holding the bid or offer.

DEFINITION 2 ("NOT WGDA") A transaction is regarded as inconsistent with WGDA if it is an accepted bid (ask) and any bid (ask) revisions since the previous transaction are made by different buyers (sellers).

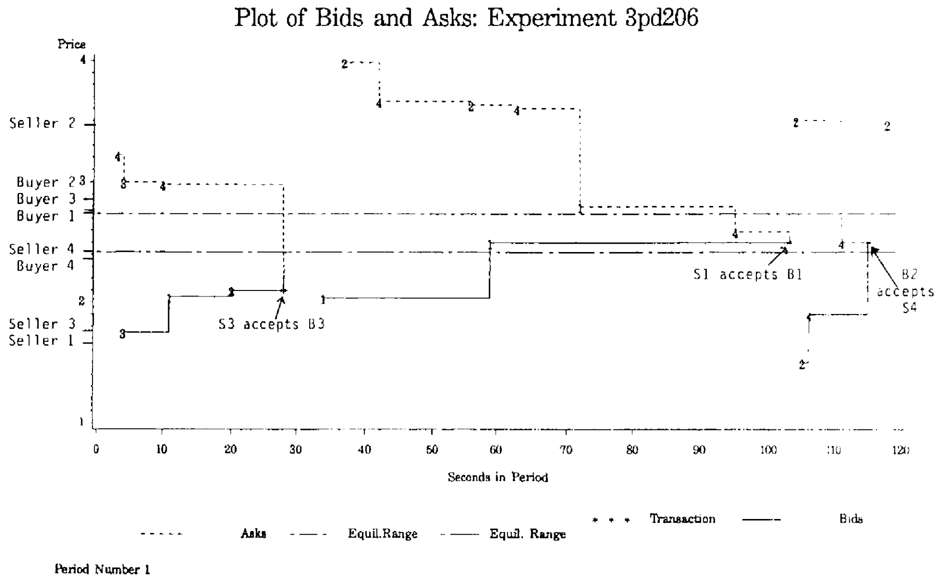


FIGURE 6 Experiment 3pd206 (Series 1), Period 1.

Figures 6 through 9 illustrate these definitions. These figures present the market state for each second of some typical periods. Figures 6, 7, and 8 present data from the Series 1 experiments, and Figure 9 presents data from a Series 2 experiment. The dotted lines represent the market asks leading up to each transaction, and the solid lines represent the market bids. Each action that improves the market bid or ask is flagged with the identification number of the buyer or seller posting the quote. The buyer and seller unit valuations are provided on the vertical axis. The dot-dash lines bracket the range of equilibrium prices.

Consider the second transaction in Figure 6. Since Buyer 1 revises her (standing market) bid at 59 seconds, this transaction is "Not BGAN" according to Definition 1. (In fact, the BGAN model does not rule out this behavior, but does suggest that it is relatively rare.) Since this transaction is an accepted bid and the only active bidder is Buyer 1, it is consistent with WGDA according to Definition 2. (Actually, the formal WGDA model rules out the frequent ask revisions prior to this transaction.) The final transaction of Figure 7 provides a better example of WGDA behavior: Seller 2 gradually improves her ask until Buyer 1 accepts. Next, consider the first transaction in Figure 8. The flurry of 8 bids and 11 asks prior to this transaction is entirely consistent with BGAN (because all revisions are made by different agents) and is "Not WGDA" (because the transaction is an accepted bid and the market bid is raised by different buyers). One observation indicated

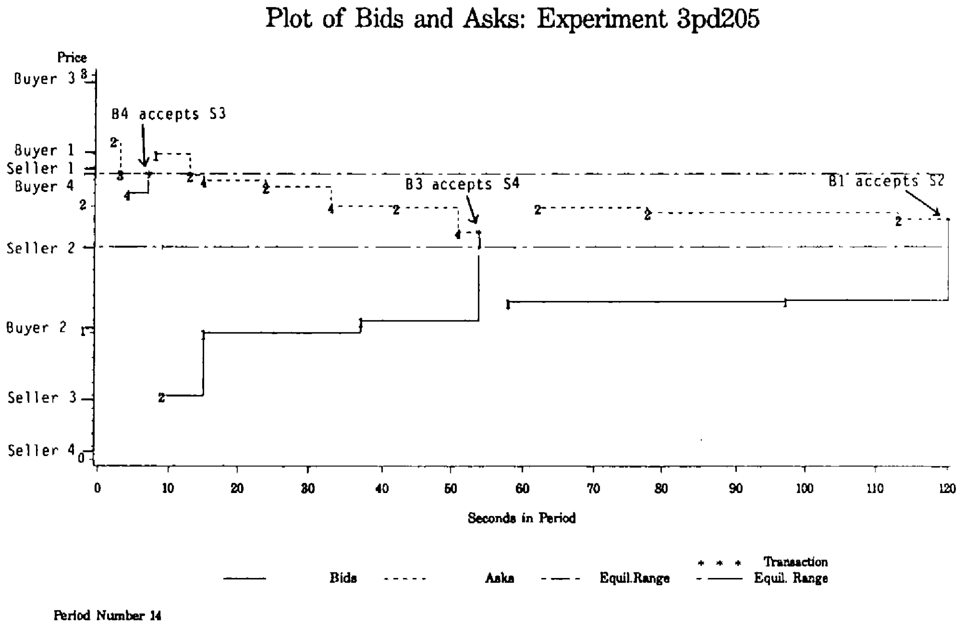


FIGURE 7 Experiment 3pd205 (Series 1), Period 14.

by these figures (and confirmed by the formal analysis below) is that quotes are more frequently revised by different agents (i.e., are "Not WGDA") rather than by the same agent that currently holds the quote (i.e., are "Not BGAN"). For that reason, we are comfortable with these simplistic definitions, even though Definition 1 is more strict than the formal BGAN model and Definition 2 is more lax than the formal WGDA model.

The null hypothesis for the statistical tests below is the zero-intelligence model. The "Not BGAN" and "Not WGDA" definitions are too complicated for an analytical derivation of their statistical properties; fortunately, the ZI model's simple rules and non-strategic behavior permit straightforward computer simulation of bids, asks, and transactions for any set of induced value and cost parameters. We conducted a Monte Carlo simulation of 1000 ZI markets using the Series 1 parameters to determine how many ZI algorithm transactions are Not BGAN and Not WGDA. Based on the empirical distributions from the simulation, we can determine the probability that a given "human" experiment's Not BGAN and Not WGDA frequencies come from the ZI model. Finding too few Not BGAN observations and too many Not WGDA observations would lead us to reject Hypothesis (2Z) in favor of the BGAN Hypothesis (2B), and finding too many Not BGAN observations and too few Not WGDA observations would lead us to reject Hypothesis (2Z) in favor of the WGDA Hypothesis (2W).

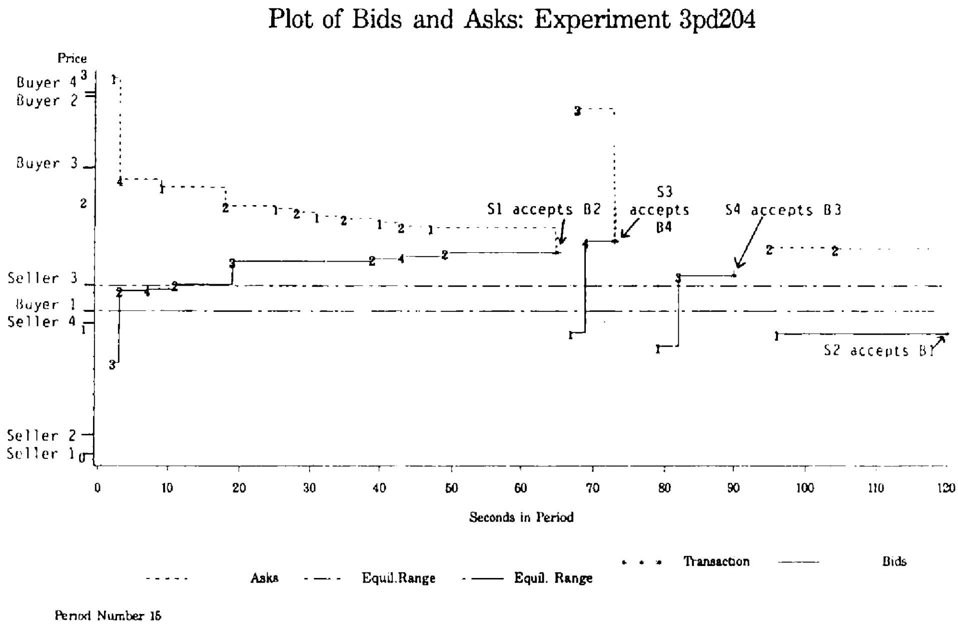


FIGURE 8 Experiment 3pd204 (Series 1), Period 15.

DEFINITION OF "SERIOUS" BIDS AND ASKS: Recall that Hypothesis (2W) applies only to "serious" bids and asks, i.e., those bids and asks that (in a sequential equilibrium) have a positive probability of acceptance. An empirical implementation could be cast in terms of either actual beliefs or beliefs in sequential equilibrium. Unfortunately, the former are unobservable and the latter are at present not computable. We examine two crude approximations for the Series 1 experiments. In the first alternative, we exclude no bids and asks as non-serious, and in the second alternative we exclude quotes outside the support of equilibrium prices, [$1.00, $2.50].[14] It turns out that neither alternative affects the conclusions.

EXPERIMENT SERIES 1: Results classifying the transactions from the three experiments of Series 1 are presented in Table 1. When all market quotes are considered (Table 1(a)), about 40% of the transactions are inconsistent with BGAN

[14]This criterion excludes 13.7% of the 1027 market actions in these three experiments. It also excludes some transaction prices (for example, the final transaction on Figure 8 is at $0.98). We are left uncertain as to whether this criterion is too strict or too lax.

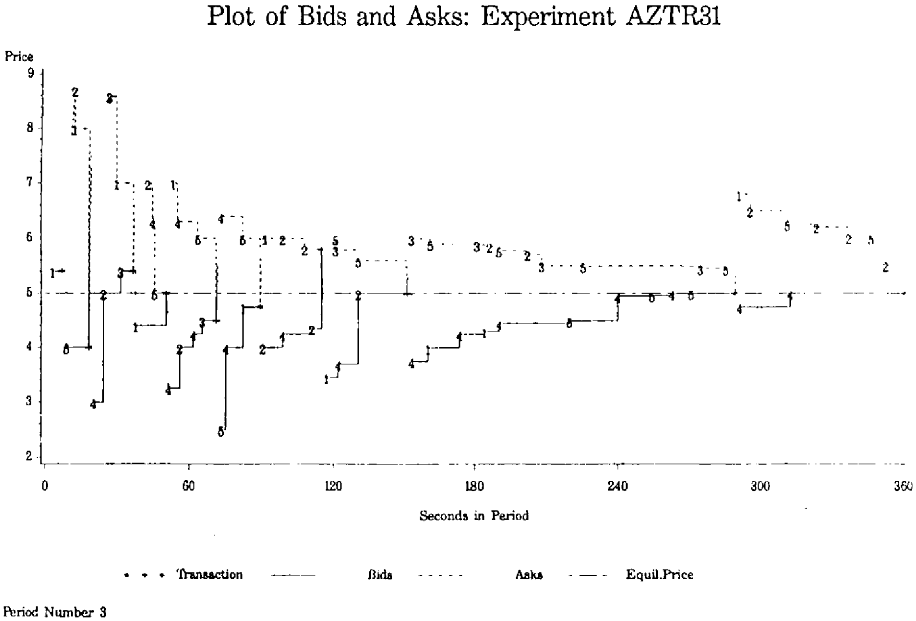


FIGURE 9 Experiment aztr31 (Series 2), Period 3.

and 75% of the transactions are inconsistent with WGDA. The simulation results indicate that Not BGAN and Not WGDA observations are equally likely in the ZI model. In all three experiments, the frequency of Not WGDA transactions exceeds the ZI critical value of 60% and the frequency of Not BGAN transactions is near the lower critical value of 38%. Furthermore, there are never enough "WGDA but Not BGAN" observations and the "BGAN but not WGDA" observations are always too frequent to be consistent with the ZI model.

When we exclude the quotes outside the support of equilibrium prices as suggested by the WGDA model (Table 1(b)), the percentage of transactions inconsistent with Hypotheses (2B) and (2W) decline slightly. However, the ZI simulation indicates that the distributions of these statistics should decrease more substantially. In general, the ZI model can be rejected in favor of the BGAN model even when excluding non-serious bids and asks to give WGDA a "better shot" at explaining the data. Furthermore, only 24% (12 of 51) of transactions inconsistent with BGAN can be classified as consistent with WGDA, while 59% (57 of 96) of transactions inconsistent with WGDA are consistent with BGAN. Put another way, nearly five times as many transactions are [BGAN but Not WGDA] than are [WGDA but Not BGAN]. When excluding quotes outside the support of equilibrium prices, over three times as many transactions are [BGAN but Not WGDA] than are [WGDA but Not BGAN].[15]

EXPERIMENT SERIES 2: For the Series 2 experiments, we have not excluded any bids and asks as non-serious. Figure 10 presents the distribution of the percentage of transactions inconsistent with BGAN and WGDA across the 27 experiments in Series 2. The percentage inconsistent with BGAN clusters around 10%, while the

TABLE 1 (a) Classification of observations as inconsistent with BGAN and/or WGDA: Exp. series 1—All observed quotes.

Exper. Number	Number of Observ.	Inconsist. with BGAN	Inconsist. with WGDA	Inconsistent with BGAN but consist. w/WGDA	Inconsistent with WGDA but consist. w/BGAN
3pd204	41	16 (39%)	34 (83%)[1]	3 (7%)[1]	21 (51%)[1]
3pd205	45	19 (42%)	30 (67%)[1]	5 (11%)[1]	16 (36%)[1]
3pd206	42	16 (38%)[1]	32 (76%)[1]	4 (10%)[1]	20 (48%)[1]
Pooled	128	51 (40%)	96 (75%)[1]	12 (9%)[1]	57 (45%)[1]
		ZI Simulation		1000 Markets	
Mean		50%	50%	24%	24%
5th percentile		38%	39%	15%	13%
95th percentile		61%	60%	33%	34%

[1] Indicates rejection of ZI Hypothesis (2Z) in favor of BGAN Hypothesis (2B) at the 5% level.

[15] A slightly different version of the ZI algorithm was also implemented in another Monte Carlo simulation. In the unreported version, each bid is uniformly distributed between the current market bid and the buyer's redemption value, and sellers' asks are similarly distributed between cost and the current market ask. Because the DA rules filter out the quotes that do not improve the current bid and ask, the results described above are basically unchanged when using this alternative algorithm.

TABLE 1 (b) Classification of observations as inconsistent with BGAN and/or WGDA: Exp. series 1—Excluding quotes from outside the equilibrium price range.

Exper. Number	Number of Observ.	Inconsist. with BGAN	Inconsist. with WGDA	Inconsistent with BGAN but consist. w/WGDA	Inconsistent with WGDA but consist. w/BGAN
3pd204	41	15 (37%)	25 (61%)[1]	5 (12%)[1]	15 (37%)[1]
3pd205	45	15 (33%)	26 (58%)[1]	7 (16%)	18 (40%)[1]
3pd206	42	15 (36%)	28 (67%)[1]	4 (10%)[1]	17 (41%)[1]
Pooled	128	45 (35%)	79 (62%)[1]	16 (13%)[1]	50 (39%)[1]
		ZI Simulation		1000 Markets	
Mean		32%	22%	23%	14%
5th percentile		23%	13%	14%	5%
95th percentile		43%	32%	33%	22%

[1] Indicates rejection of ZI Hypothesis (2Z) in favor of BGAN Hypothesis (2B) at the 5% level.

percentage inconsistent with WGDA clusters around 50–60%. Table 2 presents results numerically when all experiments of this series are pooled (a total of 2304 transactions). While we should reiterate that this series of experiments differs in several important ways from assumptions employed in the WGDA model—in particular, multiple units per trader and no public information regarding the distribution of unit valuations—it does indicate that behavior consistent with BGAN is much more common than behavior consistent with WGDA. Nearly twenty times as many transactions are [BGAN but Not WGDA] than are [WGDA but Not BGAN].

An additional feature of this series is that bid and ask behavior preceding each transaction changes for the later transactions in each period. This is illustrated by Figure 11, which plots the percentage of transactions inconsistent with each hypothesis for each transaction of the period. All 27 Series 2 experiments are combined for this figure. For the first five or six transactions in each period, almost no observations are inconsistent with BGAN (i.e., no traders are revising their own outstanding quote), and less than one-half of all transactions are inconsistent with WGDA. This occurs because early in each period bids and asks are accepted quickly, with little effort expended "bargaining" over prices with bid and ask improvements. The bottom line of this figure illustrates that on average only two or three market quotes precede these early transactions. Later in the period, however, the percentage of transactions inconsistent with both hypotheses increases substantially; there appears to be more "haggling" over prices. The bottom line in the figure shows that the average number of quotes preceding the later transactions increases to over eight. One feature of the Not WGDA and Not BGAN definitions is that increased bid and ask activity can only increase the likelihood of making a transaction inconsistent with one or both models.

In the Series 1 experiments, one buyer and one seller must exit the market after each transaction because each trader can transact only one unit. Consequently, only one or two traders are on each side of the market for the final transaction in each period, making Not BGAN observations much more likely than Not WGDA observations. This may partially explain the higher percentage of Not BGAN transactions observed in Series 1.

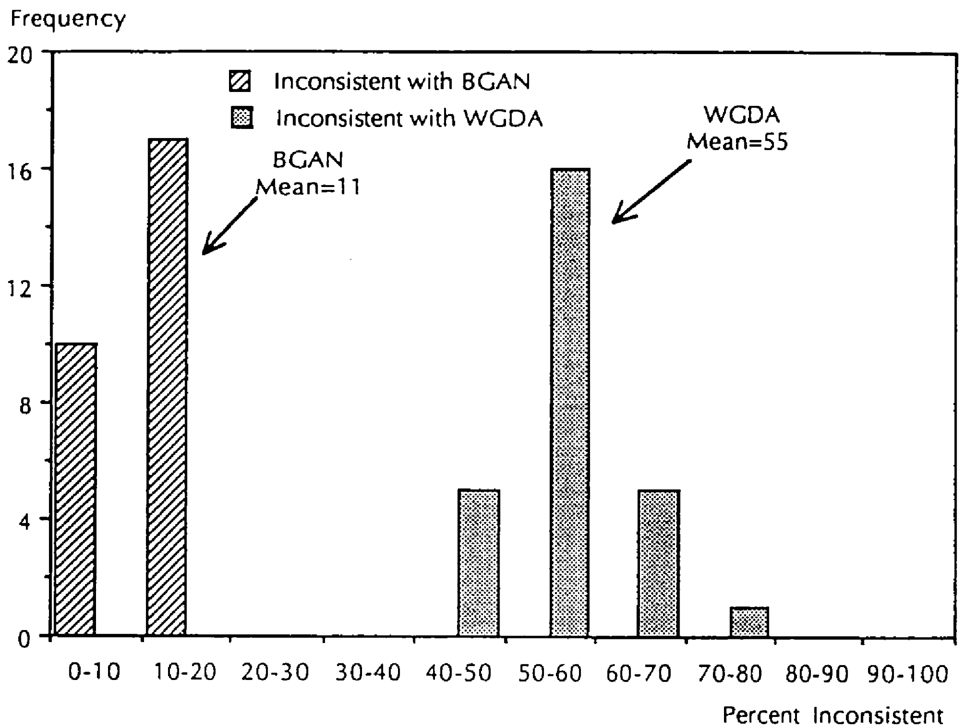


FIGURE 10 Distribution of percentages inconsistent with hypothesis 2 WGDA and BGAN models across experiment series 2 (27 observations).

TABLE 2 Classification of observations as inconsistent with BGAN and/or WGDA: experiment series 2.

Exper. Number	Number of Observ.	Inconsist. with BGAN	Inconsist. with WGDA	Inconsistent with BGAN but consist. w/WGDA	Inconsistent with WGDA but consist. w/BGAN
Pooled	2304	264 (11%)	1269 (55%)	52 (2%)	1057 (46%)

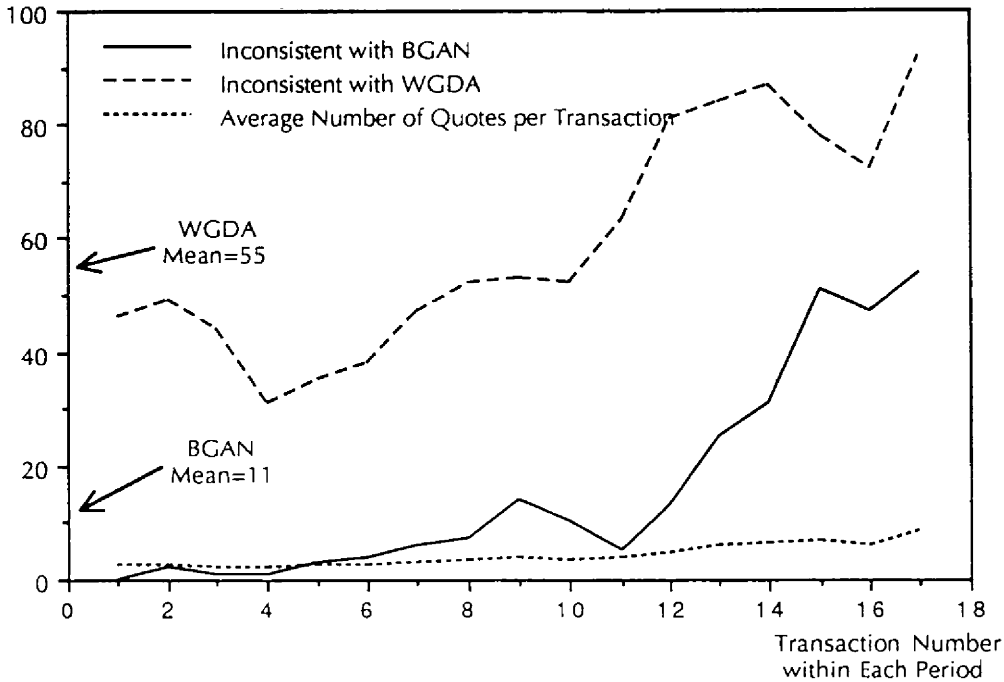


FIGURE 11 Percent of transactions inconsistent with hypothesis 2 WGDA and BGAN models: Experiment series 2. Percent inconsistent with each model and number of quotes per transaction.

CONCLUSION: In both series of experiments, improvements of the outstanding bids and offers by different buyers and sellers are more common than improvements by the same buyer and seller. Overall, the evidence clearly favors BGAN Hypothesis (2B) over WGDA Hypothesis (2W) and ZI Hypothesis (2Z).

TRANSACTIONS PARTNERS

All three models generally agree on the two remaining hypotheses, although the ZI model implies a somewhat weaker version of the following:

HYPOTHESIS (3): Early (later) transactions will be between high-value (low-value) bidders and low-cost (high-cost) sellers.

The weaker ZI model implication is that the high-value buyers and low-cost sellers are more *likely* to transact early.

Hypothesis (3) implies that the order in which buyers transact should be in decreasing order of their unit valuations, and the order in which sellers transact should by in increasing order of their unit costs. As a result, the Spearman Rank Correlation Coefficient between buyer (seller) valuation rank and transaction order should be significantly negative (positive).

EXPERIMENT SERIES 1: Table 3(a) presents estimates of this rank correlation coefficient for each of the three experiments separately and for the pooled data. For buyers, the estimated coefficients are negative (as predicted) for three of the four estimates, but only the estimate for experiment 3pd204 is significantly different from zero. For sellers, all four of the estimated coefficients are positive (as predicted), but only the estimate for experiment 3pd204 is significantly different from zero.

The rank correlation coefficients estimated for the 1000 simulated zero-intelligence markets ranged widely. For buyers, the estimated coefficients ranged from -0.59 to 0.35, with 80% of the estimates in the range $[-0.35, 0.06]$. For sellers, the estimated coefficients ranged from -0.26 to 0.75, with 80% of the estimates in the range [0.09, 0.50]. We cannot reject the hypothesis that the Table 3(a) estimates are generated by the ZI trader algorithm (the 5% lower critical value for sellers is 0.02).

A related implication of Hypothesis (3) is that gains from trade should be smaller for later trades in each period. Let the seconds left in the period at trade t be denoted as S_t, and define the gains from trade on trade at time t as $G_t = R_t - C_t$, where R_t is the valuation of the buyer in the transaction at time t and C_t is the cost of the seller in the transaction at time t. The implication that gains G_t should decrease as the period progresses can be tested by estimating the following equation:

$$G_t = \delta_o + \delta_1 S_t + e_t \,, \tag{7}$$

TABLE 3 (a) Rank correlation between transaction number and ranking of buyer and seller unit valuations: Experiment series 1.

Experiment Number	Buyer Rank Corr. Coeff.	p-value for 0 Corr.	Seller Rank Corr. Coeff.	p-value for 0 Corr.	N
3pd204	-0.33	0.037	0.33	0.037	41
3pd205	-0.07	0.652	0.08	0.580	45
3pd206	0.01	0.977	0.07	0.648	42
Pooled	-0.11	0.212	0.16	0.076	128

TABLE 3 (b) OLS estimation of Eq. (7): Experiment series 1.

Experiment Number	$\hat{\delta}_0$ (Intercept)	$\hat{\delta}_1$ (Seconds Left)	N	$\bar{R}^2$	D-W Stat
3pd204	1.00[1]	0.009[1]	41	0.22	2.0
	(0.18)	(0.002)			
3pd205	1.36[1]	0.002	45	0.00	2.5
	(0.19)	(0.003)			
3pd206	1.24[1]	0.004	42	0.02	2.4
	(0.21)	(0.003)			
Pooled	1.21[1]	0.005[1]	128	0.06	2.3
	(0.11)	(0.002)			

[1] Denotes significantly different from zero at the one percent level. (Standard errors in parentheses).

where e_t is the residual.[16] Hypothesis (3) implies that the gains from trade are decreasing as the seconds remaining in the period (S_t) are decreasing, which implies $\delta_1 > 0$. OLS estimation of this model is given in Table 3(b); all estimates are positive, and significant positive relationships are identified in two of the four regressions (experiment 3pd204 and all experiments pooled).

[16]The transaction ranking $T = 1, 2, 3, \ldots$ was also used instead of S_t as the explanatory variable in Eq. (7), with similar results.

EXPERIMENT SERIES 2: As explained previously, each agent had multiple units to trade in the experiments of this series. PLATO double auction rules require that buyers trade their highest valuation unit first, their second highest valuation unit second, and so on. Similarly, sellers are constrained to sell their lowest cost unit before selling higher cost units. For this reason, the institutional rules impose Hypothesis (3). Nevertheless, we tested Hypothesis (3) with these experiments *by only examining the first unit traded by each agent* to determine if the five buyers trade their first units in decreasing order of their resale values, and if the five sellers trade their first units in increasing order of their costs. When pooling across all 27 experiments, the rank correlation coefficients are significantly different from zero and have the sign predicted by Hypothesis (3) (although not impressively so, with estimates of -0.13 for buyers and 0.08 for sellers). An iterative maximum-likelihood GLS estimation of Eq. (7) using pooled data from all 27 experiments (GLS is used because of significant positive first-order disturbance autocorrelation) rejects the hypothesis that $\delta_1 = 0$ in favor of $\delta_1 > 0$, as implied by Hypothesis (3):

$$\underset{\text{(std. errors)}}{} G_t = \underset{(0.47)}{2.21} + \underset{(0.001)}{0.007} S_t, N = 506, \text{estimated autocorrelation} = \underset{(0.04)}{0.50} \quad (8)$$

CONCLUSION: Weak evidence exists that high-value buyers tend to transact before low-value buyers, and that low-cost sellers tend to transact before high-cost sellers. Weak evidence also exists that gains from trade decrease as more units transact, so we conclude that a weak version of Hypothesis (3), such as the version implied by the ZI model, is supported.

EFFICIENCY

HYPOTHESIS (4): Most potential gains from trade will be exhausted.

The efficiency of the three Series 1 random value markets is high and is similar to the efficiency of the initial periods in standard double auction markets (experiments 3pd204, 3pd205, and 3pd206 are 92.3%, 95.3%, and 89.9% efficient, respectively). Average efficiency is 93%, and 25 of the 45 periods (55%) achieved 100% efficiency. However, most of the ZI simulation markets achieved efficiency levels that exceeded the efficiency levels of the three experiments. The lower fifth percentile of the simulation market efficiencies is 93.9%, which is exceeded only by experiment 3pd205.

CONCLUSION: Hypothesis (4) is supported. These random value DA markets are also efficient, although they are not usually as efficient as comparable simulated ZI markets.

SUMMARY AND CONCLUSIONS

Because double auction markets are very important in field settings and because their strong equilibration properties have been extensively documented in the laboratory, we begin our study of price formation by looking at the most relevant existing data from double auction market experiments. The data analysis is shaped by three theoretical models that describe price formation within a single trading period: BGAN, WGDA, and ZI. The models agree in predicting allocational efficiency near 100% (Hypothesis 4), and the data broadly confirm this prediction. The models also agree in predicting that traders with larger potential gains from trade—the low-cost sellers and high-value buyers—will transact before those with smaller potential gains (Hypothesis 3). The data confirm that this is true on average, although there is also considerable randomness in the transaction sequences. The theoretical models disagree sharply in their characterizations of the bid and ask sequences. Here we find much stronger support for BGAN Hypothesis (2B), which calls for traders to improve the bids or asks of other traders à la Bertrand, than for WGDA Hypothesis (2W), which calls for traders to improve on their own bids or offers as in a Dutch auction, or for ZI Hypothesis (2Z), in which transactions consistent with BGAN and WGDA are about equally likely. The models also disagree in their predictions of transaction price change autocorrelation, and we find that the data support the ZI model's negative autocorrelation Hypothesis (1Z) and reject both the positive autocorrelation called for in BGAN Hypothesis (1B) and the zero autocorrelation called for in WGDA Hypothesis (1W). Data from other laboratory DA markets leads us to conjecture that this negative correlation may be explained (in part) by subject inexperience.

The results reported here have a more general possible interpretation that should be investigated with the additional experiments described below. Note that the model that relies most heavily on trader rationality (WGDA) has the least ability to describe market behavior, while the ZI model requires very little trader rationality and yet describes market behavior as well as or better than the strategic WGDA and BGAN models. This suggests that the DA institution's *rules* lead to its remarkable performance, rather than the rationality and strategic play of its traders.

Nevertheless, we do not interpret this evidence as grounds for dismissing the WGDA model. In retrospect, there are two reasons for regarding the available data as unfavorable to the WGDA model. First, the experiments used inexperienced subjects, so the presumed common knowledge (of other players' contingent strategies as well as parameter distributions) of WGDA is not given a fair chance. Second, the data do not permit testing one of WGDA's more striking implications, the dependence of the split in gains from trade on the numbers of buyers and sellers remaining, because the experiments all employed equal numbers of buyers and sellers.

New experiments would be useful to further evaluate the predictive ability of the alternative theories. The new experiments should feature single-unit transactors whose costs/values are drawn from known simple distributions, as in the Series 1 experiments. The experiments should also employ experienced subjects and the number of buyers and sellers should systematically be varied. Then the WGDA model will have a better opportunity to show its ability to explain the data. If the increased subject experience does not reduce the negative autocorrelation in price changes, it may be useful to allow traders to both buy and sell units and determine if the intertemporal arbitrage opportunities persist.

Several additional model implications can be tested in these future experiments. For example, one can examine the timing of transactions across a period. The WGDA model suggests substantial delay before the initial transaction and many trades toward the end of the period. In contrast, the BGAN and ZI models suggest greater uniformity of transactions across the period. Another discriminating test involves the source of market efficiency losses. The WGDA model only allows infra-marginal efficiency losses (i.e., untraded units with available gains from trade at the equilibrium price). The BGAN and ZI models also allow for infra-marginal efficiency losses, but can also allow inefficiencies from the trade of extra-marginal units (i.e., units that are not profitable to trade at the equilibrium price). Additional experiments can be used to test these implications.

Data from such experiments should make clear the strengths and weaknesses of the models. At that point the models can be modified to better explain the data. For example it might turn out that WGDA provides a good characterization of events in later trading periods of experiments with experienced subjects. In that case, theorists might fruitfully seek ways to weaken the model's common-knowledge assumptions without altering the basic view of the double auction as a waiting game punctuated by dutch auctions. Alternatively, if BGAN continues to do well, theorists might wish to incorporate a more sophisticated rationality into its game against nature view. Regularities in the data might also suggest completely new models of the price formation process.

We see price formation as a frontier territory for economists, rich in potential for policymakers as well as for theorists. The work presented here represents a necessary first step in exploring this territory.

ACKNOWLEDGMENTS

The authors thank Jim Cox and Ron Oaxaca for use of their data and Shawn LaMaster for programming assistance. John Rust, Arlie Williams, Robert Wilson, and participants at the Fall 1990 Economic Science Association meetings and the Santa Fe Institute conference provided helpful comments on an earlier draft. The usual caveat applies.

REFERENCES

1. Cox, J. C., and R. L. Oaxaca. "Using Laboratory Market Experiments to Evaluate Econometric Estimators of Structural Models." Working Paper, University of Arizona, 1990.
2. Easley, D., and J. O. Ledyard. "Theories of Price Formation and Exchange in Oral Auctions." This volume.
3. Fama, E. "The Behavior of Stock-Market Prices." *J. Bus.* (1965): 34–105.
4. Fama, E. "Efficient Capital Markets: A Review of Theory and Empirical Work." *J. Fin.* (1970): 383–423.
5. Friedman, D. "On the Efficiency of Double Auction Markets." *Amer. Econ. Rev.* (1984): 60–72.
6. Friedman, D. "How Trading Institutions Affect Financial Market Performance: Some Laboratory Evidence." Working Paper 221, University of California at Santa Cruz, 1990.
7. Friedman, D. "A Simple Testable Model of Double Auction Markets." *J. Econ. Behav. Organ.* (1991): 47–70.
8. Garcia, G. B. "On the Dynamics of Prices to Equilibrium in a Double-Auction Market." Working Paper, University of Chicago Business School, 1980.
9. Gode, D. K., and S. Sunder. "Allocative Efficiency of Markets with Zero-Intelligence (ZI) Traders: Market as a Partial Substitute for Individual Rationality." Working Paper, Carnegie-Mellon University, Graduate School of Industrial Administration, 1991.
10. Gode, D. K., and S. Sunder. "Lower Bounds for Efficiency in Double Auctions." This volume.
11. Granger, C., and O. Morgenstern. *Predictability of Stock Market Prices.* Massachusetts: D.C. Heath and Company, 1970.
12. Plott, C. R. "Industrial Organization Theory and Experimental Economics." *J. Econ. Lit.* **20** (1982): 1485–1527.
13. Roll, R. "A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market." *J. Fin.* (1984): 1127–1139.
14. Smith, V. L. "Microeconomic Systems as an Experimental Science." *Amer. Econ. Rev.* **72:5** (1982): 923–955.
15. Smith, V. L., and A. Williams. "The Effects of Rent Asymmetries in Experimental Auction Markets." *J. Econ. Behav. Org.* (1982): 99–116.
16. Smith, V. L., and A. Williams. "An Experimental Study of Alternative Rules for Competitive Market Exchange." In *Auctions, Bidding and Contracting: Uses and Theory,* edited by R. Engelbrecht-Wiggins, M. Shubik, and R. Stark, 307–334. New York: University Press, 1983.
17. Williams, A., and V. L. Smith. "Cyclical Double Auction Markets with and without Speculators." *J. Bus.* (1984): 1–33.

18. Wilson, R. "On Equilibria of Bid-Ask Markets." In *Arrow and the Ascent of Modern Economic Theory*, edited by G. Feiwel, 375–414. United Kingdom: MacMillan, 1987.

John H. Kagel and William Vogt†
Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260; †Permanent address, Department of Economics, Stanford University, Stanford, CA, 94305

TEN
Buyer's Bid Double Auctions: Preliminary Experimental Results

1. INTRODUCTION

This paper presents results from an ongoing experimental investigation of the buyer's bid double auction (BBDA) (as termed by Satterthwaite and Williams,[23,24] hereafter referred to as SW). The BBDA is a two-sided clearinghouse trading mechanism (call market) in which buyers and sellers simultaneously submit sealed bids and offers for a single unit of a commodity. What is interesting about the BBDA is that it represents a double auction trading mechanism for which there is a clear game-theoretic solution in terms of a Bayesian Nash equilibrium with incomplete information. The BBDA is a special case of what SW[25] call a k-DA auction—the case of $k = 1$. In investigating the BBDA experimentally, we hope to provide insight into how real buyers and sellers respond to the mechanism, with a view to developing a descriptively accurate model of double auction behavior.

Friedman[11] distinguishes between clearinghouse trading mechanisms and the more familiar (in terms of experimental implementation) continuous double auction trading mechanism. What is common to both mechanisms is that they implement two-sided auctions in which there are several buyers and several sellers. A clearinghouse mechanism is a discrete trading mechanism in which buyers and sellers submit one set of bids and offers per trading period, after which the market clears according to a well-defined set of rules for who trades and at what prices. In contrast, a continuous double auction mechanism permits trade at any time in a trading period, with buyers and sellers free to continuously update their bids and offers throughout the trading period.[1]

In implementing the BBDA experimentally, buyers' and sellers' valuations are randomly drawn from a known distribution, with new random draws in each trading period. This is common practice in experimental evaluation of one-sided auctions[7,17] but is rarely employed in experimental studies of double auction trading mechanisms. Rather, double auction experiments commonly employ stationary aggregate supply and demand schedules (supply and demand schedules that remain fixed across market periods), with occasional, unannounced shifts in the schedules (however, see Cason and Friedman,[4] for analysis of a continuous double auction experiment with random valuations).[2] A random value environment is natural for investigating Bayesian Nash equilibrium theories of price formation in markets where traders have incomplete information regarding each other's valuations. One might also argue that it is more representative of field settings where supply and demand schedules are rarely, if ever, completely stationary from one period to the next and traders do not know each other's reservation prices. Further, a random value setting is the appropriate setting for examining the Hayek[14] hypothesis—that markets are capable of achieving (close to) the competitive equilibrium price and quantity as determined by traders' underlying costs and limit prices—as this hypothesis is intended to apply to markets where traders have incomplete information regarding each others' valuations. With stationary supply and demand schedules and the same set of traders, a situation is created in which traders effectively acquire complete information regarding market-clearing price and quantity after several trading periods.

2. THE BBDA MECHANISM

In the BBDA, each buyer (seller) draws a single private valuation $x_{2i}(x_{1i})$ from a known probability distribution $F_2(F_1)$ defined on the interval $[\underline{x}, \overline{x}]$. In the BBDA, buyers who get to trade earn profits equal to $(x_i - p)$, where p is the market price. Sellers who get to trade earn profits equal to $(p - x_i)$. Buyers and sellers who do not trade earn zero profits.

In the BBDA, price is selected at the upper endpoint of the closed interval determining the market-clearing price, with all trade occurring at this price. This

[1] We, along with SW, use the term double auction to refer to two-sided auctions. Friedman[11] uses it to refer to the continuous double auction trading mechanism.

[2] There are a number of hybrid implementations with buyers' and sellers' individual valuations changing from period to period but with aggregate supply and demand schedules stationary (see Aliprantis and Plott,[1] for example) or with stationary supply and demand schedules whose intercepts shift randomly, by the same amount, in each period so that the competitive equilibrium price varies randomly while equilibrium quantity is stationary (see McCabe, Rassenti, and Smith[20]).

is determined as follows: All bids and offers are arranged in non-decreasing order $s_1 \leq s_2 \leq, \ldots, \leq s_{2m}$, where m is the number of buyers and sellers in the market (and the number of buyers is assumed to equal the number of sellers). Price is set at $p = s_{m+1}$. Trade occurs among all sellers whose offers are *strictly less* than p and all buyers whose bids are *greater than or equal* to p. SW show that in the case where $s_m < s_{m+1} < s_{m+2}$, p is a market-clearing price; otherwise, there is excess demand. In the case of excess demand, the available units are allocated to buyers starting with the high bidder and continuing down the list of bidders; when the point is reached where there are tied bids and not enough units remaining to satisfy them, units are assigned randomly to the remaining eligible bidders.

SW[23] demonstrate that each seller in the BBDA has a dominant strategy to set his offer equal to his reservation value, in response to which each buyer has an incentive to bid less than his reservation value. This strategic misrepresentation causes the BBDA to be *ex post* inefficient. However, the amount of misrepresentation of buyers must be small when the market becomes large. With uniform distributions of buyers' and sellers' valuations, SW have computed the exact increase in the expected gains from trade, as it depends on the number of traders, for the BBDA compared to a fully efficient outcome (gains that would be generated if traders honestly reported their reservation values). These increase rapidly from 92.6% efficiency in auctions with two buyers and sellers ($m = 2$) to 99.6% efficiency in markets with eight buyers and sellers ($m = 8$) (see SW,[23] Table 5.1). The buyer's bid function underlying this rapid increase in efficiency, given a uniform distribution of buyers' and sellers' valuations, is

$$b_i = \frac{m}{m+1} x_{2i}$$

where b_i and x_{2i} are buyer i's bid and valuation, respectively. In contrast, using a simple fixed-price rule, with prices set at the midpoint of the distribution functions underlying buyer's (seller's) valuations (costs), efficiency increases at a substantially slower rate.

Figure 1 illustrates two cases of the BBDA in terms of supply and demand schedules with $m = 4$. In Figure 1(a), buyers 4 and 3 and sellers 1 and 2 trade at a price set by seller 3's offer (note that seller 3 does *not* trade here). In Figure 1(b), again buyers 4 and 3 and sellers 1 and 2 trade, but in this case price is set at buyer 3's bid.[3]

[3]The BBDA is a special case (where $k = 1$) of the more general k-DA mechanism described in Rustichini, Satterthwaite, and Williams[22] and SW.[25] Smith, Williams, Bratton, and Vannoni,[26] using stationary supply and demand schedules, implement a sealed bid-offer (SB-O) trading mechanism for a double auction that is close in spirit to the k-DA mechanism with $k = .5$ (their PQ mechanism). It too is a clearinghouse trading mechanism, but buyers and sellers have multiple units for sale and must submit a *single* bid or offer for *all* units. Although the two mechanisms are similar, there would appear to be important strategic differences between the multiple-unit SB-O game and the single-unit k-DA with $k = .5$.

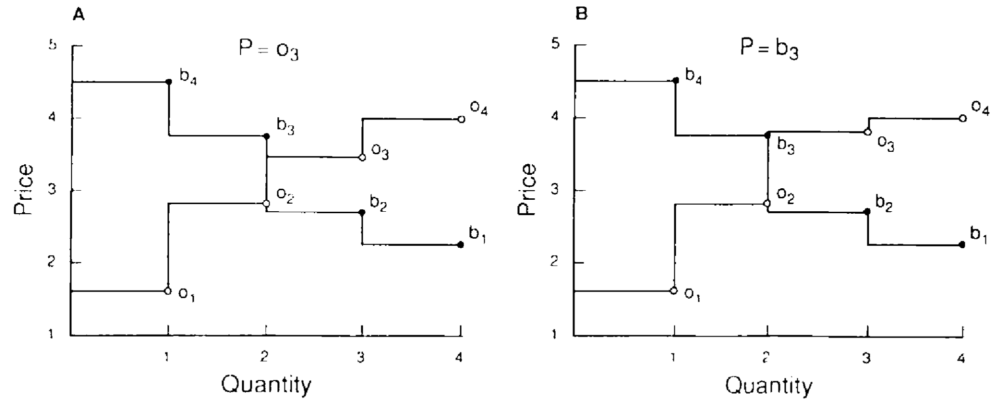


FIGURE 1 Price and Quantity Determination in BBDA.

We have also investigated a *modified BBDA* (referred to as MBBDA). As in the original BBDA, sellers who trade do not set the market price. Rather, price is always set by the buyer with the lowest bid to the left of the intersection of the supply and demand schedules. In terms of Figure 1, in both cases price is set by buyer 3.[4]

Under the MBBDA, sellers continue to have no incentive to bid above their costs, as this has no effect on market price and only serves to price them out of the market when they might otherwise be able to trade profitably. Sellers do, however, have incentives to bid below cost, as some degree of price cutting may open up profitable trading opportunities.[5] Further, buyers in the MBBDA have a somewhat stronger incentive to under-reveal demand, as they always set the market price and this affects their tradeoff between the probability of trading and the profit earned conditional on being the trader who sets the market price. We have no equilibrium characterization for the MBBDA, and there may well be no pure strategy, symmetric equilibrium (see SW,[25] section 5). Unfortunately this was not

[4]In cases of ties resulting from excess demand or excess supply, buyers and sellers who trade are determined randomly from among the tied bids and offers.

[5]For example, consider the case where $m = 2, F_1 = F_2 = U[0,5]$, and buyers 1 and 2 bid \$5.00 and \$0.00, respectively. Assume that seller 2 has a cost of \$4.50 and that seller 1 offers at \$3.00, irrespective of his cost. Then it pays for seller 2 to offer at below cost (below seller 1's offer), as he wins the item but receives a price of \$5.00, for a positive profit. Bidding below cost is risky for sellers, and opportunities for positive profits as a result of offering below cost appear to be quite circumscribed, and to be inversely related to m. However, the example does indicate that a dominant strategy of offering at cost no longer exists.

TABLE 1 Experimental Design

Session	Bidding Procedures: Single	Multi	Dual	Dual Switch	Mechanism	Experience	Improved Instruction
1	3-10[1]	none	12-23	none	BBDA	no	no
2	none	5-10	15-29	none	BBDA	no	no
3	none	none	3-14	17-29	BBDA	yes(1,2)	no
4	4-11	none	14-21	24-31	MBBDA	no	yes
5	none	3-10	13-20	22-29	MBBDA	no	yes
6	none	none	2-14	16-21[2]	MBBDA	yes(4,5)[3]	yes

[1] All periods not appearing in the table were dry runs.

[2] A computer failure caused session six to end prematurely. It had been planned and announced that the session would run for 25 live periods. It ran only for 19.

[3] The subjects in session six consisted of 14 experienced subjects from sessions four and five plus 2 inexperienced subjects. Five rounds of the multi-markets were run at the beginning of the session to orient the inexperienced subjects. These rounds are not reported.

DEFINITIONS

SINGLE During the single-market periods, all sixteen subjects participated in one large $m = 8$ market.

MULTI. The multi-market periods consisted of four separate $m = 2$ markets. The subjects were randomly divided each period into four markets of two buyers and two sellers each.

DUAL. This technique combined the two above. Each period, each subject participated in both an $m = 2$ and an $m = 8$ market. The subject had the same resale value or cost in both markets and was paid off in only one of the markets. The market to be paid off in was determined randomly each period.

DUAL SWITCH. During these periods, the subjects participated in a dual market as described above; however, the subjects who had been buyers in earlier periods became sellers and vice versa.

known prior to introducing the modification.[6] Our motivation for implementing the MBBDA was that it is close in spirit to the BBDA, but instructions to subjects regarding the trading rules would be simplified, as it is only necessary to explain a single leading case like Figure 1(b), and not two leading cases along with a number of special tie-breaking cases as well.[7] We report results from the MBBDA as an instructive sidetrack, since intuitively and behaviorally (at least for our experimental subjects) it is hard to distinguish from the BBDA.

3. EXPERIMENTAL DESIGN

Table 1 shows the six experimental sessions that have been conducted to date. Sessions 1–3 employed the BBDA mechanism, with session 3 calling back experienced subjects from sessions 1 and 2. Sessions 4–6 employed the MBBDA mechanism, with session 6 calling back experienced subjects from sessions 4 and 5.

The focus of the experimental design was to test predictions about the convergence of market efficiency to 100% as m increases. Our target values were $m = 2$, the minimal number required to get unique theoretical predictions, and $m = 8$ where equilibria are close to 100% efficiency under the BBDA.

Experimental sessions were conducted using a *dual-market technique*. Under the dual-market technique, using the same private valuation, bidders first bid in a market with $m = 2$, but before the market-clearing price is established, they again bid in a market with $m = 8$. Only one of the two markets is paid off; which one is determined randomly after both sets of bids and offers have been submitted.[8] The dual-market technique minimizes the extraneous variability resulting from changes in private valuations and changes in the subject population as m increases. That is, it operationalizes responses to *ceteris paribus* changes in the number of buyers and sellers since the same subjects bid with exactly the same valuations, only with m changing. The dual-market technique has been employed to good effect in one-sided auctions, reliably resulting in increased bidding with increased numbers of bidders in first-price private value auctions.[2,8]

Each bidder was randomly assigned to a (new) small market grouping in each auction period, with four small markets being conducted simultaneously. All subjects bid in the same large market. In order to simplify instructions as much as possible, auctions with inexperienced subjects began with several periods of bidding in a single market before introducing the dual-market technique.

In implementing the BBDA mechanism, we faced two important design decisions. First, although players will never suffer losses when playing according to the

[6] We, along with SW, were fooled, at first blush, into thinking that the equilibrium outcome for the MBBDA would constitute a rather straightforward extension of the BBDA.

[7] Experimental instructions are available from author on request.

[8] This eliminates any possibility of portfolio effects with subjects hedging their bets between the two auctions.

theory, we knew from past experience with one-sided auctions that some sellers were likely to offer to sell at prices below their costs and some buyers were likely to bid at prices above their valuations. We decided to put no restrictions on bids and offers, giving players a starting capital balance of $5, with gains and losses added to the balance.[9] We also had substitute bidders standing by in case a bankruptcy occurred (should a player's cash balance be 0 or less). This never happened. Second, we had to decide what information to report back to subjects. Within the theory, in equilibrium, this should have no effect on subjects' behavior. However, in an experiment we know that play does not immediately settle on an equilibrium outcome, and we really have no idea how different information flows might affect behavior. We decided to report back all bids and offers in the market the subject was participating in, along with the underlying private valuations (individual subject ID numbers were, however, suppressed).

Subjects were recruited primarily from undergraduate classes in economics. Instructions were copied from Plott[21] with appropriate modifications for the institutional characteristics of the BBDA and MBBDA mechanisms. In experimental sessions 4–6 the auction mechanism was explained with the aid of supply and demand diagrams like those shown in Figure 1. In addition, in going from sessions 1–3 to 4–6, the instructions were much more explicit about the possibility of losing money and how to avoid it. This change was made in response to questions commonly asked in sessions 1–3 about how one could lose money. The objective here was to formalize this element of the instructions.

4. EXPERIMENTAL RESULTS

We summarize our findings in the form of a series of conclusions. The supporting evidence underlying these conclusions is reported along with each conclusion.

4.1 MARKET EFFICIENCY

Table 2 reports mean efficiency levels for all dual-market periods for the session as a whole and for the last several periods of each auction session, after subjects have become accustomed to the procedures. The latter recognizes the fact that early-period choices are often reflective of some sort of adjustment process. These last-period data involve the last five auction periods in sessions 1 and 2 and the last five periods before and after role reversals in sessions 3–6 ("last 10" periods).

[9] Bids and offers were restricted to lie in the interval [0,5] in sessions 1–3. There were no restrictions in sessions 4–6. There are practical reasons for restricting bids and offers to lie within some reasonable interval of the upper and lower bounds of the distributions F_1 and F_2. The starting capital balance of $5.00 was used in lieu of paying a participation fee. Bystanders who did not get to play received $10.

In addition, the reader will recall that sessions 3 and 6 employed subjects who had participated in an earlier experimental session.

Several theoretically based efficiency measures are offered as well. The first is the efficiency predicted under the BBDA mechanism for the private valuations realized in each experimental session.[10] These are best viewed as idealized efficiency measures, upper bounds on the efficiency likely to be realized in practice. Also reported are efficiency measures under a *fixed-price mechanism* with the market price set at $2.50, the midpoint of the distribution underlying buyers' and sellers' valuations. In computing these fixed-price efficiencies, we assume that buyers (sellers) follow the dominant strategy of truthfully revealing their valuations (costs). Although the fixed price rule is not a market-clearing mechanism, it does provide one reasonable lower-bound measure of efficiency. (In cases where the market did not clear with the fixed-price rule, the computer randomly rationed trade on the long side of the market.) Also reported are efficiency measures for budget-constrained zero-intelligence (ZI) traders; buyers (sellers) who bid a random number drawn from a uniform distribution between their valuation (cost) and $0.00 ($5.00). Efficiency levels for ZI traders provide a second lower-bound measure of efficiency. This measure is attractive because of the simplicity of the underlying decision rule and the remarkably high efficiency levels (between 95–100%) reported for ZI traders under continuous double auction pricing rules.[12,13]

> **Conclusion 1:** Observed efficiency is typically lower than idealized efficiency (the BBDA prediction). However, observed efficiency is higher than under the fixed-price rule and much higher than with ZI traders.

Looking at last-period data, markets achieve average efficiency levels of 88.3% and 89.5% with $m = 2$ and 8, respectively. This compares quite favorably with the BBDA prediction of 92.4% efficiency with $m = 2$, but less favorably with the 99.6% efficiency predicted with $m = 8$.

Actual efficiency is considerably higher than the 76.5% average efficiency achieved under a fixed-price rule with $m = 2$, and somewhat higher than the 85.3% average efficiency under the fixed-price rule with $m = 8$. One way to view the fixed-price rule is as a clumsy instrument that is bound to achieve 100% efficiency as m increases, as the dominant strategy of bidding (offering) at valuation (cost) is completely transparent.[11] In contrast, the BBDA mechanism opens up the possibility of achieving substantially higher efficiency with small numbers of traders, while at the same time opening up the possibility that the trading rules are sufficiently complicated that traders fail to respond to the strategic possibilities, or respond incorrectly to them. Consequently, what this comparison shows is that although real players do not respond in an idealized fashion to the BBDA mechanism, "mistakes" are not large enough or frequent enough to offset the promised improvements in efficiency.

[10] This same measure is employed in the MBBDA sessions as well.

[11] This prediction should, of course, be tested experimentally.

TABLE 2 Market Efficiency[1]

	$m = 2$				$m = 8$			
Session	Observed[2]	Predicted[3]	$P = 2.5$[4]	ZI	Observed[2]	Predicted[3]	$P = 2.5$[4]	ZI
1	83.7	85.9	67.1	29.1	85.2	99.4	82.6	37.1
1 last 5[5]	79.9	88.1	52.3	28.1	88.3	99.1	70.0	35.7
2	93.1	88.1	73.3	27.7	89.1	99.5	82.1	36.1
2 last 5[5]	92.6	87.8	74.4	27.5	92.0	99.8	82.0	35.8
3	86.9	94.1	75.6	26.4	87.4	99.7	83.0	39.4
3 last 10[6]	89.9	97.9	80.8	25.9	84.1	99.6	89.5	40.7
4	93.9	93.4	85.0	32.3	88.6	99.7	89.3	37.9
4 last 10[6]	90.1	94.7	84.3	31.1	88.9	99.5	92.5	38.0
5	82.6	93.8	78.8	29.8	82.8	99.8	87.4	36.8
5 last 10[6]	86.7	92.2	82.5	31.1	87.4	100.0	89.0	34.5
6	88.8	93.9	83.4	29.8	95.1	99.6	88.9	34.8
6 last 10[6]	90.5	93.7	84.4	31.7	96.3	99.5	88.5	32.5

[1] Dual-market periods only.

[2] The sum, across periods, of observed consumer and producer surplus divided by the maximum possible consumer and producer surplus.

[3] The sum, across periods, of the consumer and producer surplus that would have been earned playing according to the benchmark BBDA theory, divided by the maximum possible consumer and producer surplus.

[4] The sum, across periods, of the consumer and producer surplus that would have been earned with a fixed-price mechanism $P = 2.5$, divided by the maximum possible consumer and producer surplus (assuming that the subjects fully reveal demand and supply).

[5] Last 5 dual-market periods.

[6] Last 5 dual-market periods before role reversal and last 5 dual-market periods following role reversal.

Human subjects do substantially better than ZI traders whose market efficiency averages 29.3% and 36.2% with $m = 2$ and 8, respectively. These efficiency levels are considerably lower than those achieved by budget-constrained ZI traders under continuous double auction rules (95–100% efficiency as reported in Gode and Sunder[12]). We suspect that the main difference between ZI traders here and in continuous double auctions is that clearinghouse trading rules are completely unforgiving with respect to bids and offers that fail to achieve mutually beneficial trades. In contrast, continuous double auction rules, which permit repeated bids and offers until a mutually beneficial trade is achieved, forgive such mistakes.

One way to look at ZI traders is that they represent an extreme form of mistake responding to a given set of trading rules. As such, Gode and Sunder attribute the high efficiency levels that human traders achieve under continuous double auction rules as resulting primarily from the trading rules, in conjunction with a budget constraint, rather than having anything to do with human traders' intelligence or perceptiveness. Under this interpretation, the substantially higher efficiency levels achieved by human, compared to ZI, traders under the BBDA mechanism is indicative of the human traders' responsiveness to the trading rules.

Conclusion 2: Efficiency levels increase with increases in the number of traders for idealized traders, ZI traders, and under the fixed-price rule. However, there is only a very limited tendency for actual efficiency to increase with more traders.

There is only a slight, statistically insignificant, increase in average efficiency in markets with $m = 8$ versus $m = 2$ for actual traders. That there is no consistent pattern here is reflected in the fact that efficiency increases as m increases in the last periods of sessions 1, 5, and 6, but decreases in the other sessions. (Similar results hold for the session total data, except session 5 shows a decrease in efficiency as m increases.)

Market inefficiencies result from three different factors: (1) extra-marginal units being traded, trades which will not occur under the Pareto efficient outcome,[12] (2) infra-marginal units not trading, trades that must occur to achieve the Pareto efficient outcome, and (3) "wrong" units trading, a lower valued buyer (higher cost seller) trading in place of a higher valued buyer (lower cost seller). Inefficiency rose on all three counts as m increased: (1) extra-marginal units were traded in 13% of all market periods with $m = 2$ compared to 28% of all periods with $m = 8$, (2) infra-marginal units failed to trade in 13% of all market periods with $m = 2$ compared to 35% with $m = 8$, and (3) the wrong units traded in 11% of all market periods with $m = 2$ compared to 68% with $m = 8$. However, these frequency counts do not measure the amount of damage (lost consumer and producer surplus as a percent of total surplus) resulting

[12] With sellers offering at cost and buyers bidding at or below their valuations, there will be zero extra-marginal units traded under the BBDA clearinghouse rules.

TABLE 3 Predicted and Actual Prices and Quantities[1]

	$m = 2$				$m = 8$			
Session	Prices		Quantitiy		Prices		Quantity	
	Actual	Predicted[2]	Actual	Predicted[2]	Actual	Predicted[2]	Actual	Predicted[2]
1	2.80	2.39	.79	.54	2.48	2.26	3.75	3.58
2	3.04	2.33	.98	.70	2.64	2.52	3.73	4.07
3	2.98	2.51	.87	.67	2.64	2.47	3.76	3.36
4	3.10	2.36	.97	.75	2.93	2.94	4.19	3.81
5	2.69	1.90	.88	.73	2.58	2.27	3.31	3.69
6	3.12	2.19	.97	.88	2.79	2.61	4.26	4.21

[1] Dual-market periods only.

[2] Predicted values based on BBDA mechanism with idealized traders.

TABLE 4 Sellers' Misrepresentation (sellers' offer less cost)

	$m = 2$				$m = 8$			
Session	Mean[1] (S_m)	Frequency % Negative	Zero	Positive	Mean[1] (S_m)	Frequency % Negative	Zero	Positive
1	-.05 (.13)	39.6	9.4	51.0	.03 (.11)	32.3	14.6	53.1
1 last 5[2]	-.18 (.23)	42.5	12.5	45.0	.08 (.19)	30.0	27.5	42.5
2	.19 (.09)	9.1	16.7	74.2	.21 (.10)	10.0	18.3	71.7
2 last 5[2]	.25 (.15)	7.5	22.5	70.0	.26 (.15)	5.0	25.0	70.0
3	-.12 (.14)	39.0	11.5	49.5	.08 (.13)	39.5	10.5	50.0
3 last 10[3]	-.23 (.16)	46.3	11.3	42.5	.17 (.14)	45.0	8.8	46.3
4	.09 (.08)	18.8	2.3	78.9	.14 (.09)	21.9	1.6	76.6
4 last 10[3]	.09 (.09)	18.8	2.5	78.8	.19[4] (.10)	22.5	2.5	75.0
5	.23[4] (.11)	21.9	14.1	64.1	.31[4] (.12)	17.2	10.2	72.7
5 last 10[3]	.11 (.12)	25.0	16.3	58.8	.25[4] (.11)	17.5	13.8	68.8
6	.14[4] (.07)	15.8	29.0	55.2	.09 (.09)	18.4	26.3	55.2
6 last 10[3]	.07 (.08)	21.3	21.3	57.5	.09 (.09)	23.8	17.5	58.8

[1] Mean across subject means; S_m = standard error of the mean.

[2] Last 5 dual-market periods.

[3] Last 5 dual-market periods before role
reversal and last 5 dual-market periods following role reversal.

[4] Significantly different from 0 at 10% level, 2-tailed t-test.

from these inefficient trades. But clearly the damage per trade has to be getting smaller, on average, as m increases. This is consistent with the fact that average efficiency actually rose slightly with increased m.

> **Conclusion 3:** Actual prices and quantities tend to exceed predicted prices and quantities. This tendency is most pronounced with $m = 2$.

Table 3 reports actual and predicted prices and quantities, where predicted values are based on the BBDA mechanism with idealized traders. With $m = 2$, actual prices average close to 68 cents above predicted prices (29.6% of mean predicted price), and mean quantities average .20 more than predicted (27.8% of mean predicted quantity). For $m = 8$, actual prices average 17 cents above predicted prices (6.6% of mean predicted price), and mean quantities average .05 more than predicted (1.2% of mean predicted quantity). The higher than predicted prices here are reflective of sellers' tendency to offer at above cost and buyers' tendency to bid closer to their valuations than predicted. Individual buyer and seller behavior is discussed in the next section.

4.2 INDIVIDUAL SUBJECT BEHAVIOR

In what follows, we characterize the behavioral foundations for the deviations between actual and idealized behavior in the BBDA. We begin with seller behavior.

> **Conclusion 4:** The majority of seller's offers deviate from the dominant strategy prediction of full cost revelation. This cost sellers an average of 8 cents per period, or 10% of the dominant strategy earnings. Had all sellers followed the dominant strategy, with no adjustment in buyers' bids, average market efficiency would have been 95.7% with $m = 2$ and 95.0% with $m = 8$.

Table 4 shows mean seller misrepresentation (sellers' offers less their costs) and the frequency distribution of these deviations (negative seller misrepresentation involves offering at below cost). Costs to sellers for misrepresenting are computed in terms of an individual seller unilaterally offering at cost.[13] Note that costs of deviating from the dominant strategy usually involve opportunity costs rather than negative profits, as out-of-pocket losses are unlikely to result from offering at below cost.

The deviations from the dominant strategy prediction are not unexpected, as deviations from the dominant bidding strategy are reported in one-sided, single-unit, second-price auctions[17,18] and in one-sided, multiple-unit, uniform-price auctions.[6]

[13] In making these computations, we employ a norm of offering at cost for the MBBDA auctions.

As with the one-sided auctions, the dominant bidding strategy is far from transparent in the BBDA auction, and there is small expected cost associated with deviating from the dominant strategy. The relative frequency of offering at below cost was reduced in sessions 4–6 where the instructions emphasized the fact that offering at or above cost would guarantee not losing money. Changing the number of buyers and sellers in the market had little systematic effect on the extent of seller misrepresentation.

Looking at individual subjects, around 30% of all sellers deviated from offering at cost by 10 cents or less more than 90% of the time; i.e., about 30% of all sellers were reasonably close to offering at cost almost all the time. Further, looking at the cost of misrepresenting, about 50% of all sellers would *not* have improved their earnings at all had they unilaterally offered at cost all the time.[14] We suspect that it is this relatively weak and inconsistent relationship between deviations from the dominant offer strategy and reduced earnings, in conjunction with the fact that the dominant strategy is not transparent, that is responsible for sellers not consistently offering at cost.

Our counterfactual efficiency measures show that average efficiency would have increased to around 95% had sellers offered at cost. For $m = 2$, this would be higher than the predicted efficiency of 92.4% with traders playing according to the BBDA predictions. This indicates that buyers' bids were closer to valuation than predicted with $m = 2$, which was indeed the case (see below). These efficiency calculations are counterfactual in the sense that we do not know what response buyers would take in response to sellers offering at cost. What these efficiency measures do indicate is that sellers offering at cost would not be sufficient, by itself, for efficiency levels to increase with increases in the number of traders.

Conclusion 5: Buyers consistently bid more than predicted under the BBDA with $m = 2$. This cost buyers an average of 20 cents per period, or almost half of what they would have earned bidding according to Eq. (1). For $m = 8$, overbidding was much more modest, costing buyers around 6 cents per period, less than 15% of what they would have earned, on average, bidding according to (1). Had buyers followed the bidding function 1, with no adjustment in sellers' offers, market efficiency would have been reduced to 83.1% per period with $m = 2$, but would have increased to 94.8% with $m = 8$.

Table 5 shows predicted buyer misrepresentation (buyers' valuations less their bids) based on the risk-neutral bid function (1), actual buyer misrepresentation, and the frequency distribution of buyer misrepresentation (negative buyer misrepresentation involves bidding above valuation). Costs to buyers for misrepresenting are computed in terms of an individual buyer unilaterally adopting the bid function (1).

[14]In the MBBDA sessions, two sellers would have *reduced* their earnings slightly if they always offered at cost. These two sellers are included in the 50% measure.

TABLE 5 Buyers' Misrepresentation (Buyers' valuation less bid)

	$m = 2$						$m = 8$					
Session	Mean Deviation (S_m)			Frequency %			Mean Deviation (S_m)			Frequency %		
	Predicted[1]	Actual[2]		Negative	Zero	Pos.	Predicted[1]	Actual[2]		Negative	Zero	Pos.
1	.76	.01	(.09)	25.0	17.7	57.3	.25	-0.11	(.14)	33.3	14.6	52.1
1 last5[3]	.79	.01	(.19)	15.0	30.0	55.0	.26	-.30	(.25)	40.0	25.0	35.0
2	.85	-.02	(.09)	63.3	6.7	30.0	.28	.01	(.09)	57.5	10.8	31.7
2 last 5[3]	1.00	.11	(.13)	60.0	5.0	35.0	.33	-.01	(.15)	60.0	10.0	30.0
3	.76	.01	(.08)	45.0	8.5	46.5	.25	-.08	(.08)	49.0	6.0	45.0
3 last 10[4]	.73	.03	(.10)	41.3	11.3	47.5	.24	-.02	(.12)	45.0	8.8	46.3
4	NA[7]	.20[5]	(.11)	7.8	6.3	85.9	NA[6]	.18[5]	(.10)	12.5	1.6	85.9
4 last 10[4]	NA[7]	.17	(.15)	8.8	5.0	86.3	NA[7]	.13	(.13)	13.8	1.3	85.0
5	NA[7]	.11	(.11)	18.0	14.1	68.0	NA[7]	.12	(.10)	18.0	13.3	68.8
5 last 10[4]	NA[7]	.06	(.15)	18.8	13.8	67.5	NA[7]	.11	(.11)	18.8	13.8	67.5
6	NA[7]	.20[6]	(.07)	17.8	10.5	71.7	NA[7]	.09	(.10)	20.4	10.5	69.1
6 last 10[4]	NA[7]	.20[6]	(.07)	13.8	16.3	70.0	NA[7]	.06	(.11)	17.5	15.0	67.5

[1] Based on equation (1) in the text.

[2] Mean across subject means; S_m = standard error of the mean.

[3] Last 5 dual-market periods.

[4] Last 5 dual-market periods before role reversal and last 5 dual-market periods following role reversal.

[5] Significantly different from 0 at 10% level, 2-tailed t-test.

[6] Significantly different from 0 at 1% level, 2-tailed t-test.

[7] NA Not applicable. No clear theoretical prediction.

Based on the bid function (1), buyers should have bid 75 to 80 cents less than their valuations, on average, with $m = 2$, and 25 to 30 cents below valuations with $m = 8$. Buyers bid much closer to their valuations than this, however, particularly with $m = 2$. It is not clear that buyers should bid according to (1), given the extent of seller misrepresentation reported. And it is possible that the smaller than predicted underrevelation on the part of buyers constitutes a strategic response to sellers offering above costs. (This can, of course, be tested by having live buyers play against computerized sellers who follow the dominant strategy.) Nevertheless, the extent to which buyers overbid relative to (1), particularly with $m = 2$, cost them an average of 20 cents per period, or a little over 50% of predicted profits. Note that most of these costs were in the form of opportunity costs, rather than negative earnings. Further, bidding above valuation was reduced substantially in auction series 4–6, probably as a result of the clear and explicit warnings that such bidding might result in losses.

The overbidding relative to the risk-neutral Nash equilibrium (RNNE) prediction reported here is not unlike the overbidding relative to the RNNE prediction found in one-sided, first-price, sealed-bid auctions. The overbidding in first-price auctions is most extreme in auctions with small numbers of bidders, in which case it is not uncommon for bidders to take home 50% or less of the predicted RNNE profits (see Kagel and Roth,[19] Table 4). Some of the overbidding in one-sided, first-price auctions is no doubt a result of risk aversion, and some a result of buyers' misperceptions and/or inexperience (see Kagel,[16] for a review of the experimental literature on this point). Some of these same forces are likely to be at work in the BBDA as, other things equal, risk aversion will generate bidding above (1).

Buyers are sensitive to the strategic implications of increasing numbers of buyers and sellers as they bid more, on average, when m increases (a mean increase in bids, relative to valuations, of \$0.10, which is significantly different from 0 at the 10% level).[15] Sensitivity to increasing numbers of bidders is also reported in one-sided, first-price, private-value auctions (see Battalio, Kogut, and Meyer,[2] for example).[16] Nevertheless, this increased bidding did not result in systematic increases in efficiency here, either because it was not consistent enough and/or because at times it resulted in bidding above valuations, which is likely to result in reduced efficiency.

Our counterfactual efficiency measures show that average efficiency would have been reduced to 83.1% had buyers followed the bidding function (1) with $m = 2$, and increased to 94.8% with $m = 8$.

These efficiency calculations are counterfactual in the sense that we do not know what response sellers might make in response to buyers bidding less. These counterfactual efficiency measures suggest that the failure of efficiency levels to

[15] Using the last-period data from each session as the unit of observation, $t = 2.04$ with 5 degrees of freedom using a paired t-test.

[16] Battalio, Kogut, and Meyer[2] explicitly compare the effects of the dual-market bidding procedure with within-session crossover procedures in evaluating the effects of increasing numbers of bidders. They report no systematic differences under the two procedures.

systematically increase with increases in m is at least partly due to overbidding on the part of buyers with $m = 2$, as it is simply impractical to expect bidding to increase very much, starting from such high levels, following increases in m (as buyers will not knowingly suffer losses).

5. SOME REMARKS ON THE RELATIONSHIP BETWEEN THEORY AND EXPERIMENTAL EVIDENCE

We view the results reported here as more in the nature of a pilot study, or an initial inquiry into the BBDA.[17] As experimenters we are not interested in simply falsifying or confirming a theory but rather in understanding the basic behavioral processes underlying what we observe. Theory provides an important benchmark regarding fully rational behavior as well as a guide to interesting experimental treatments that have definite directional implications. The BBDA, or k-DA, auction is most attractive in these respects.

On the basis of results reported in simpler one-sided auctions we never expected the point predictions of the BBDA to be satisfied. There is an *a priori*, as well as an empirical, basis for this expectation. Bayesian Nash equilibria impose very strong information-processing requirements and posit extremely sophisticated play on the part of traders. Few people expect most human agents to be able to satisfy these requirements regardless of the extent of their experience with a particular economic environment. The idea that professional traders, by virtue of their superior experience, will consistently behave in agreement with the theory is problematic as well. Cech and Ball[5] review the data on subject pool effects between student and market professionals in laboratory experiments which follow standard experimental procedures. With one exception, the studies reviewed show no closer conformity to the predictions of the theory on the part of professional traders. We suspect that there are two reasons for this: (1) learning tends to be situation specific, while laboratory settings strip away many of the contextual clues professionals employ in field settings, and (2) laboratory markets, which are based on theoretical work, are often not fully representative of the environment encountered in the field.[3,9]

In addition, under the BBDA there is little of what psychologists call reinforcement feedback that would promote playing closer to the equilibrium predictions (recall, for example, that 50% of all sellers would not have improved their average earnings had they unilaterally offered at cost *all* the time). Finally, for a purist, completely satisfactory tests of the theory become even more difficult once it is recognized that the symmetry assumptions usually required for deriving clean theoretical predictions are not likely to be precisely satisfied, either because agents have different risk preferences and/or because they have different information-processing capacities that are difficult, if not impossible, to completely control for.

[17] This section is stimulated by the thoughtful remarks offered in section 5 of SW.[25]

So what do experimenters fall back on to investigate and evaluate a theory? How do we rationalize our inquiries? First, it is useful to posit alternative, nonrational models to provide alternative benchmarks. With an alternative benchmark we can compare the organizing ability of the fully rational model to the nonrational alternative, thereby getting some idea of the relative drawing power of economic theory. Second, one can rely on tests of the comparative static implications of the theory, for even though point predictions are often not satisfied, comparative static predictions often are, at least directionally. Admittedly, such tests involve a leap of faith as once the point predictions of a theory fail to be precisely satisfied we have no theory regarding "near" theories that can be used to rationalize these comparative static predictions. Nevertheless, these tests are important since, in some respects, the comparative statics are what we're really interested in anyway. And, to many, economic theory is designed to embody primary tendencies and was never meant to be taken literally, so that breakdowns in comparative static predictions are more damaging than breakdowns in point predictions. Third, one can compare the effects of different institutions in economically interesting contexts. This usually proves interesting in its own right (for example, comparing the efficiency and price characteristics of different kinds of clearinghouse mechanisms or a clearinghouse mechanism with the continuous double auction).[18] Fourth, one needs to explore the effects of procedural variations on experimental outcomes, as it is not uncommon for results to be sensitive to different procedural implementations.

6. SUMMARY AND CONCLUSIONS

We have reported initial results from a buyer's bid double auction experiment. Using the benchmark of fully rational behavior, the theory exhibits a number of shortcomings. But efficiency measures are uniformly high, much higher than predicted with zero-intelligence traders and consistently higher than predicted under a fixed-price mechanism. Thus, the theory outperforms these alternative benchmarks. The most significant failing of the theory in our minds is that efficiency fails to increase significantly as the number of traders increase, rather it increases moderately and the direction of change is erratic across experimental sessions. The fact that efficiency does not increase can be largely attributed to buyers bidding substantially more than predicted in auctions with small numbers of traders and to there being a practical upper bound on how much bids can increase without exceeding buyers' valuations. (The extent to which bids exceed the risk-neutral Nash equilibrium prediction with small numbers of traders is reminiscent of results reported in one-sided, first-price, sealed-bid auctions.) Subjects receive little consistent feedback for such unambiguous point predictions as sellers following the (non-transparent)

[18] See Smith et al.[26] and Hong and Plott[15] for nice examples of this in the context of multiple-unit double auctions with stationary supply and demand schedules.

dominant strategy of bidding at cost. This inconsistent feedback must seriously retard learning. A similar lack of consistent feedback applies to buyers who choose to explore best replies to sellers' actions. It remains to be seen whether one can devise strategies that will enhance the learning process.[19]

As noted in the introduction, the results presented here are part of an ongoing experimental investigation of the BBDA. A number of avenues of continuing research suggest themselves. First, we plan to conduct additional BBDA auctions with an enhanced set of instructions similar to those used in the MBBDA auctions. This will test for the sensitivity of our results to procedural modifications. Second, we plan to look at bidder behavior in BBDA auctions with computerized sellers following the dominant bidding strategy. This should make for a more stable environment for learning to take place, as well as determine whether the limited discounting of bids relative to valuations results from buyers' response to sellers' offering at above cost. Finally, we plan to run a parallel series of continuous double auctions using random valuations and to compare efficiency and prices relative to the competitive equilibrium outcome across institutions.

ACKNOWLEDGMENTS

An earlier version of this paper was presented at the Santa Fe Institute Conference on Double Auctions, June 1991. Research support has been provided by the Economics Division of the National Science Foundation and the REU initiative of the NSF. The paper has benefitted from the comments of conference participants, an anonymous referee, the editors, and discussions with Jack Ochs and Mark Satterthwaite. The usual caveat applies.

[19]SW[25] provide a number of suggestions on how to do this. They argue that it would be easier for traders to learn best responses to permanent reservation values than to learn the optimal response function for a set of randomly drawn values and suggest clever means for implementing this suggestion. There are, however, potential incentive problems associated with their suggestions. Buyers (sellers) with low (high) valuations (costs) are likely to lose interest as they will be rarely able to earn profits. In contrast, buyers (sellers) with relatively high (low) valuations (costs) have little pressure to behave strategically. This points up the fact that price is determined by at most two of the $2m$ bids and offers which, as SW note, is the basis for the inconsistent reinforcement feedback in the first place.

REFERENCES

1. Aliprantis, C. D., and C. R. Plott. "Experiments with Competitive Equilibrium in Overlapping Generations Economies." Mimeographed, California Institute of Technology, Pasadena, CA, 1991.
2. Battalio, R. C., J. Kogut, and J. Meyer. "Individual and Market Bidding in a Vickrey First-Price Auction: Varying Market Size and Information." In *Advances in Behavioral Economics*, vol. 2, edited by L. Green and J. H. Kagel. Norwood, NJ: Ablex Publishing, 1990.
3. Burns, P. "Experience and Decision Making: A Comparison of Students and Businessmen in a Simulated Progressive Auction." In *Research in Experimental Economics*, vol 3, edited by V. L. Smith. Greenwich, CT: JAI Press, 1985.
4. Cason, T., and D. Friedman. "An Empirical Analysis of Price Formation in Double Auction Markets." This volume.
5. Cech, P., and S. B. Ball. "Subject Pool Choice and Treatment Effects in Economic Laboratory Research." Discussion Paper No. 9142, General Motors Research Center for Strategy in Management, Kellogg Graduate School of Management, 1991.
6. Cox, J. C., V. L. Smith, and J. M. Walker. "Expected Revenue in Discriminative and Uniform Price Sealed Bid Auctions." In *Research in Experimental Economics*, vol. 3, edited by V. L. Smith. Greenwich, CT: JAI Press, 1985.
7. Cox, J. C., V. L. Smith and J. M. Walker. "Theory and Individual Behavior in First-Price Auctions." *J. Risk and Uncertainty* **1** (1988): 61–99.
8. Dyer, D., J. H. Kagel, and D. Levin. "Resolving Uncertainty About the Number of Bidders in Independent Private-Value Auctions: An Experimental Analysis." *Rand J. Econ.* **20** (1989): 268.
9. Dyer, D., and J. H. Kagel. "The Winner's Curse and Bidding in the Commercial Construction Industry: A Case Study." In preparation.
10. Easley, D., and J. Ledyard. "Theories of Price Formation and Exchange in Double Oral Auctions." This volume.
11. Friedman, D. "The Double Auction Institution: A Survey." This volume.
12. Gode, D., and S. Sunder. "Human and Artificially Intelligent Traders in Computer Double Auctions." Mimeographed, Carnegie-Mellon University, 1991.
13. Gode, D., and S. Sunder. "Lower Bounds for Efficiency in Double Auctions." This volume.
14. Hayek, A. "The Use of Knowledge in Society." *Amer. Econ. Rev.* **35** (1945): 519–30.
15. Hong, J. T., and C. R. Plott. "Rate Filing Policies for Inland Water Transportation: An Experimental Approach." *The Bell J. Econ.* **13** (1982): 1–19.
16. Kagel, J. H. "Auctions: A Survey of Experimental Research." In *The Handbook of Experimental Economics*, edited by J. H. Kagel and A. E. Roth. Princeton University Press, forthcoming.

17. Kagel, J. H., R. M. Harstad, and D. Levin. "Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study." *Econometrica* **55** (1987): 1275–1304.
18. Kagel, J. H., and D. Levin. "Independent Private Value Auctions: Bidder Behavior in First-, Second-, and Third-Price Auctions with Varying Numbers of Bidders." Mimeograph, University of Houston, 1990.
19. Kagel, J. H., and A. E. Roth. "Comment on Harrison versus Cox, Smith and Walker: Theory and Misbehavior in First-Price Auctions." *Amer. Econ. Rev.*, forthcoming.
20. McCabe, K., S. Rassenti, and V. Smith. "Designing a Uniform-Price Double Auction: An Experimental Evaluation." This volume.
21. Plott, C. R. "Industrial Organization Theory and Experimental Economics." *J. Econ. Lit.* **20** (1982): 1485–1587.
22. Rustichini, A, M. Satterthwaite, and W. Williams. "Convergence to Price Taking in Simple Market." Discussion Paper 914, Center for Mathematical Studies in Economics and Management Science, Kellogg Graduate School of Management, 1990.
23. Satterthwaite, M. A., and S. R. Williams. "The Rate of Convergence to Efficiency in Buyer's Bid Double Auction as the Market Becomes Large." *Rev. Econ. Studies* **56** (1989): 477–498.
24. Satterthwaite, M. A., and S. R. Williams. "Bilateral Trade with the Sealed Bid k-Double Auction: Existence and Efficiency." *J. Econ. Theor.* **48** (1989): 107–133.
25. Satterthwaite, M. A., and S. R. Williams. "The Bayesian Theory of the k-Double Auction." This volume.
26. Smith, V. L., A. W. Williams, W. K. Bratton, and M. G. Vannoni. "Competitive Market Institutions: Double Auction vs. Sealed Bid Offer Auctions." *Amer. Econ. Rev.* **72** (1982): 58–77.

Kevin A. McCabe,* Stephen J. Rassenti,† and Vernon L. Smith‡
*Department of Accounting, Carlson School of Management, University of Minnesota, Minneapolis, MN 55455; †Economic Science Laboratory, University of Arizona, Tucson, AZ 85721; and ‡Economic Science Laboratory, University of Arizona, Tucson, AZ 85721

ELEVEN
Designing a Uniform-Price Double Auction: An Experimental Evaluation

INTRODUCTION

Call markets have been used on many European (and Israeli) Stock Exchanges, and for the daily opening on the New York, American, and Tokyo Stock Exchanges in all outstanding listed issues. In some versions the bids (offers) are verbal, in others written. In some cases the procedure is Walrasian: orders may be modified after a trial price is given (Whitcomb[12]; also see Jarecki[5] for the Walrasian mechanism implemented with a unanimity stopping rule in the London bullion market). In France, new issues are offered in fixed quantity using modified, uniform-price, sealed-bid auction rules.[10] However, the dominant form of auction in current financial markets is the continuous double auction (DA).

Price volatility is now a major source of dissatisfaction among investors in stocks. This raises the cost of capital to firms and indirectly hurts the economy as a whole. As a consequence, trading practices and rules have become suspect, such as the role of specialists, program trading, and the lack of market integration. Less

The Double Auction Market, Eds. D. Friedman and J. Rust,
SFI Studiew in the Sciences of Complexity,
Proc. Vol. XIV, Addison-Wesley, 1993

attention has been given to the form of the auction; i.e., alternatives to the continuous double auction where orders arrive, and contracts are executed, sequentially. With few exceptions, most market centers have adapted the computer to existing oral DA practices[2] and have not attempted to systematically adapt trading institutions to the new forms of trading made possible by computer technology. A new trading institution proposed and implemented by Wunsch Auction Systems (WAS) departs from this trend, and exploits the unique possibilities made feasible by the computer in coordinating single-price trading.

Our interest in developing trading system alternatives to DA began as an academic exercise in 1987. We conjectured that some of the features of that institution—such as the "improvement rule" requiring admissible bids and offers to narrow the bid-ask range, and the public display or outcry of all bids and offers—might be incorporated into a uniform-price auction that retained the remarkable efficiency properties of DA yet captured the nondiscriminatory, zero-volatility characteristics of the uniform-price, sealed-bid auction. After this research was well advanced, we discovered that similar considerations had motivated Steven Wunsch to leave Kidder Peabody and form his own company (Wunsch Auction Systems, Inc.) to develop the same kind of trading system for common stock. Since 1990 we have had close intellectual ties with Wunsch, who began operations with the new system, in competition with the New York and other stock exchanges, in the spring of 1991. In December 1991 two of us (Rassenti and Smith through their company, Cybernomics) were instrumental in getting Wunsch to move to Phoenix and become the new Arizona Stock Exchange. Consequently, both a laboratory and a field experiment occurred almost simultaneously.

The research reported here is part of a larger program to design efficient markets for complex economic system problems such as natural gas and electric power networks.[6,7,8] In such situations there are many markets whose prices must be simultaneously coordinated. It is nearly an impossible task for a simple telephone market or for multiple independent double auction markets at various locations to efficiently coordinate trading and also avoid excessive transactions costs. One solution is a feedback call market with continuous computer-optimized updating of the tentative allocation. These very valuable potential applications lend great weight to the need for studying call market variants.

In this paper we study the alternative form of auction which allows continuous orders (like a DA) but executes contracts at previously defined points in real time as in a call market. Since contracts are executed at a single point in time, there is one market price. This motivates the name of our proposed auction: The Uniform-Price Double Auction or UPDA. UPDA is conceptually equivalent to a clearinghouse institution reported in Friedman,[3] although there are some implementational differences.

UPDA is a hybrid auction which combines the advantage of having more bids and offers in determining price (at the "call") with the advantage of continuous order updating, which reduce price and transaction uncertainty. UPDA itself does not provide the same degree of immediacy as DA, but the specialist system can be employed to provide immediate execution for anyone unable to wait for the completion of the "call."

Our interest in UPDA has evolved from our general interest in auction design. More than 20 types of single- and two-sided auctions, half of them synthetic, have been studied at the Economic Science Laboratory. Most of these institutions are included in Friedman's taxonomy.[4] We call this research program "Market Engineering" since our purpose is to study new institutional designs using the laboratory. The use of the laboratory is necessary when complex extensive forms cannot be readily reduced to formal models.

It should be evident that the economist's analytical skill in acquiring an *ex post* understanding of the incentive and efficiency properties of particular contractual and institutional arrangements does not imply that we can use that understanding to design successful arrangements *ex ante* for some particular new application. This is because we don't know how to weigh the importance of various elements in an institutional design. For example, the Dutch auction allocations have been shown to be less efficient than other auctions (English, sealed bid), but Dutch flower auctions are enormously faster than other auction forms and thereby offer lower transactions' cost, which helps to account for their use in the sale of small ticket (and perishable) items like cut flowers and plants.

Our view is that the technology of laboratory experimentation can facilitate the interplay between the evaluation and modification of proposed new exchange institutions before field implementation. This paper is part of the exploratory process of investigating this conjecture. Laboratory experiments allow one to investigate the incentive and performance properties of alternative exchange institutions, and, with respect to institutional design, they provide a low-cost means of trying, failing, altering, trying, etc. This process uses theory, loose conjecture, intuitions about procedural matters, and, most important, repeat testing to understand and improve the features of the institutional rules being examined.

UPDA

UPDA accepts orders in the form of bids and asks. Orders arrive with the ID of the person making the order; the time the order was made, the order type (bid/ask), the limit buy/sell price, and the maximum transaction quantity; i.e.,

$$\text{Order} = (\text{ID, time, bid/ask, price, quantity}).$$

When supply is fixed in advance, as in a new issues market, only bids are accepted.

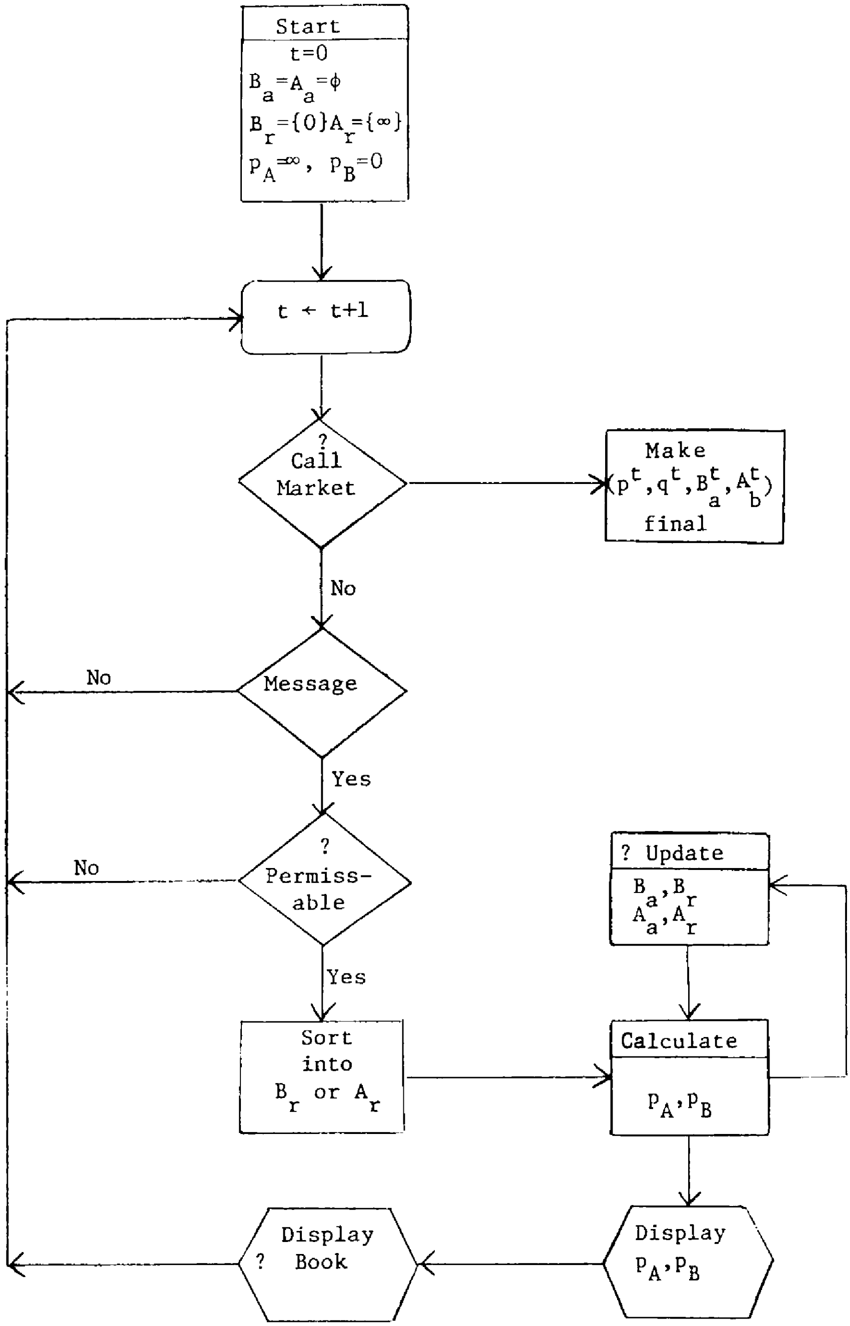


FIGURE 1 Flow Chart of UPDA.

We refer to the sequence of bids and asks, sorted by time, as the order flow, denoted F^t, defined by: $F^t = \{\text{Orders: time} \leq t\}$. We break this order flow down into four mutually exclusive subsets:

$$\begin{aligned}
B_a^t &= \{\text{the set of bids tentatively accepted for trade}\};\\
B_r^t &= \{\text{the set of bids rejected for trade}\};\\
A_a^t &= \{\text{the set of asks tentatively accepted for trade}\};\ \text{and}\\
A_r^t &= \{\text{the set of asks rejected for trade}\}.
\end{aligned}$$

The set of rejected bids always contains a dummy bid of 0 while the set of rejected asks always contains a dummy ask of ∞. In addition, we maintain these sets sorted by price: highest to lowest for both bid sets and lowest to highest for both ask sets. All ties are broken by time priority. These four sets are collectively known as the "book." At any point in time the tentative contract, (p^t, q^t, B_a^t, A_a^t), consists of a market price, denoted p^t, the sets of accepted bids and asks, and q^t, the number of accepted bids and asks. At any point in time there also exists a standing bid price p_B^t which would guarantee any seller a tentative contract if he met its terms, and a similar standing ask price $p_A^t > p_B^t$ which guarantees any buyer a tentative contract.

For convenience in the discussion which follows, we also define:

$$\begin{aligned}
LaB &= \text{price of the last acceptable bid } (q\text{th element of } B_a^t);\\
LaA &= \text{price of the last acceptable ask } (q\text{th element of } A_a^t);\\
FrB &= \text{price of the first rejected bid (1st element of } B_r^t);\ \text{and}\\
FrA &= \text{price of the first rejected ask (1st element of } A_r^t).
\end{aligned}$$

We give a flow chart for UPDA in Figure 1. Question marks indicate where we investigate alternative institutional rules for UPDA. First, when do we call the market, i.e., make the tentative contract final? The field favors a fixed point in time. This rule is simple, but it invites endgame strategies. There are many alternatives. Call rules may be based on tentative trade volume (double auction can be thought of as a special case of UPDA with a matched quantity call), price differences since the previous call, bid and offer arrival rates, randomly generated points in time, or hybrids of any of these. For the purpose of this study, we chose a fixed-time rule and an endogenous call based on elapsed time of inactivity.

CALL RULE:

a. Fixed-Time Call (exogenous close): The market is called when the auction has been open a predetermined amount of time.

b. Fixed-Interval Call (endogenous close): the market is called when no new tentatively acceptable bid or offer arrives for a specified amount of elapsed time.

The next set of rules deal with processing orders. The message test simply checks for the existence of incoming messages. If there exists a message, we must decide under what (if any) terms it is permissible. Here we plan to use one of two rules.

MESSAGE RULE:

a. Costless Improving Order Submission: The order is permissible if it is a new order or improves the terms of an existing order. There is no time-related cost for submitting orders.

b. Costly Order Submission: The order is permissible if it is a new order or improves the terms of an existing order at a predetermined premium dependent on time t. A non-improving change in the terms of a order is permissible for a penalty (e.g., $2 \times$ premium) dependent on time.

Existing laboratory experiments have documented the value of improvement rules for market efficiency. However, in the field it may be necessary to allow changes in orders. Making such changes costly reduces the incentive to try to manipulate prices with false orders which are changed or withdrawn before the call (as in the opening of the Toronto Exchange). Rule b features timed premiums (as implemented in the Wunsch auction). The intention is to provide an incentive to submit meaningful orders early in the process and so avoid a congested inefficient endgame.

The next rule to consider is the update rule: how do the contents of the sets B_a, B_r, A_a, and A_r evolve. The form of the standing bid p_B and ask p_A (prices at which agents are guaranteed a tentative contract) implicitly defines how updating must occur and what the final uniform price p should be. We consider two possible rules. The both-sides (2S) rule allows any agent to beat the terms of accepted orders on his own side of the market or meet the terms (p_B or p_A) of the other side—whichever is less demanding. The other-side (1S) rule forces an agent to accept the terms of the other side of the market (p_B or p_A) to become accepted.

UPDATE RULE:

a. Update based on $2S$ rule:

$$\begin{aligned} p_B &= \max\{LaA - 1, FrB\}; \\ p_A &= \min\{LaB + 1, FrA\}; \\ p &= 1/2(p_A + p_B). \end{aligned}$$

b. Update based on $1S$ rule:

$$\begin{aligned} p_B &= \text{2nd high } (LaB \cup B_r); \\ p_A &= \text{2nd low } (LaA \cup A_r); \\ p &= 1/2(LaB + LaA). \end{aligned}$$

TABLE 1 Order Arrival Sequence

Time	Bid	Ask
1	42	
2		52
3		44
4	46	
5	48	
6		46

Note: $LaA - 1$ means one monetary unit below the last accepted ask. A new seller must pay $LaA - 1$ to displace the seller who offered LaA. Similarly, $LaB + 1$ means one monetary unit above the last accepted bid—the price that a new buyer must pay to displace the incumbent buyer at LaB.

Under either rule, bids and asks are updated in two steps:

$$\text{Step 1} \quad \begin{aligned} &B_a \leftarrow B_a \cup FrB \text{ iff } FrB \geq p_A; \\ &A_a \leftarrow A_a \cup FrA \text{ iff } FrA \leq p_B. \end{aligned}$$

$$\text{Step 2} \quad \left. \begin{aligned} &B_a \leftarrow B_a \setminus LaB \\ &A_a \leftarrow A_a \setminus LaA \end{aligned} \right\} \text{ iff } LaB < LaA.$$

The 2S rule corresponds to the Wunsch implementation of a bid/ask feedback mechanism for UPDA. All accepted bids (asks) are greater (less) than or equal to all rejected bids (asks). The 1S rule was conceived by us in response to observed behavior in auctions using 2S. It ensures a greater price spread than the first rule and allows marginal inconsistency in favor of overcoming price inertia as volume increases. It increases the incentive of agents in the rejection sets to form new matched trades that become accepted. It also most nearly corresponds to the standard continuous DA rules. In order to illustrate the difference between these rules, consider Table 1, a sequence of bids and offers for one unit each of a single commodity. The first trade occurs at time 4 and the book (regardless of update rule) looks as in Table 2. All bids and offers above the double line are tentatively accepted; those below rejected. However, the bid-ask spread would not be identical (see Table 3).

For example, this means that to submit an acceptable bid under the 2S rule, a bidder need only bid $LaB + 1 = 47$ to displace the currently contracted buyer at 46, but under the 1S rule he must bid at least 52 (the FrB) to displace her. In like

TABLE 2 Book: Time 4

Bid	Ask
46	44
42	52

TABLE 3 Bid/Ask: Time 4

	2S Rule	1S Rule
Standing Ask, p_A	47	52
Standing Bid, p_B	43	42
Price, p	45	45

TABLE 4 Book: Time 5

2S Rule		1S Rule	
Bid	Ask	Bid	Ask
48	44	46	44
46	52	48	52
42		42	

TABLE 5 Bid/Ask: Time 5

	2S Rule	1S Rule
Standing Ask, p_A	49	52
Standing Bid, p_B	46	46
Price, p	47.5	45

TABLE 6 Book: Time 6

2S Rule		1S Rule	
Bid	Ask	Bid	Ask
48	44	48	44
46	46	46	46
42	52	42	52

TABLE 7 Bid/Ask: Time 6

	2S Rule	1S Rule
Standing Ask, p_A	47	52
Standing Bid, p_B	45	42
Price, p	46	46

manner under 2S, a new seller must offer $LaA - 1 = 43$ to displace the contracted seller at 44, while under 1S the seller must offer at most 42. Note that the spread (bid 42)/(asked 52) under 1S would also be the standing spread in a continuous DA at time 4, because the bid at 44 and ask at 46 would have already formed a binding contract. Consequently, 1S is a "natural" rule system for generalizing DA.

Now consider the book and spread for 1S and 2S which result at time 5 after a bid of 48 has arrived (Tables 4 and 5). Notice the displacement that has occurred in the 2S rule but not the 1S rule. In the former case the standing ask has decreased and the standing bid increased as a result of the new bid, while in the latter case only the standing bid has increased. Also, the price is inside the spread in 2S but outside the spread in 1S. Consider the consequence of a new ask at 46 which hits the standing bid at time 6 (Tables 6 and 7). The books once again look identical although the spread is much greater under the 1S rule. The tentative volume of trade is always the same under both rules.

It is this 1S rule which we believe may be a key innovation in designing a uniform-price double auction. If the market price of the current tentative contract is off the competitive equilibrium, one side must adjust its prices while the other side stands pat. Under these circumstances the 2S rule can foster a painfully slow adjustment process where agents of the same side leapfrog and displace each other by epsilon price increments without quantity increases. Under the 1S rule you cannot displace someone on your own side from a tentative contract by simply bettering his price: you must usually agree to the terms of the first rejected agent on the opposite side of the market.

The final design decision concerns how much information is revealed to the agents during the call. We have again limited ourselves to two alternatives—reveal the entire book or only the parts of the book that pertain directly to the agent concerned.

INFORM RULE:

a. Limited Information (closed book): The agent sees only his own bids or asks that have been tentatively accepted and rejected.

b. Full Information (open book): The agent sees the entire book B_a, B_r, A_a, and A_r sorted by price with a separating line drawn between accepted and rejected orders. The orders which she owns are highlighted to be easily identifiable.

The Wunsch Auction uses rule b. In both cases we constantly publish p_A^t and p_B^t as reference points for the agents, even though they are implicitly contained in the book when it is published.

EXPERIMENTAL DESIGN AND PROCEDURES

We adopt a design used in Campbell, LaMaster, Smith, and Van Boening.[1] This environment has been used extensively with different institutions including double auction (DA), double Dutch (DD), and tâtonnement auctions (TAT). Figure 2 illustrates the normalized structure of supply and demand. In each period each buyer is assigned (randomly) to one of the demand steps B1–B5, and each seller is assigned (randomly) to one of the supply steps S1–S5, shown in Figure 2. The odd-numbered agents have six units to trade while the even-numbered agents have three units each. Then a random constant from the interval [1.00, 4.90] was added to (or subtracted from) all values. This introduces uncertainty with respect to market-clearing prices but keeps equilibrium quantity constant (at eighteen units) and in double auction experiments this environment yields large price volatility. In Figure 2 we also graph the theoretical price for each of the fifteen periods of an experiment. Since subjects are never informed as to the supply and demand structure, nor do they know on which step they reside, they remain uncertain of the relationship between the equilibrium price and their assigned values.

The random shifts in supply and demand create price uncertainty. Furthermore, the 20-cent equilibrium tunnel creates a sizeable bargaining gap, promoting strategic uncertainty. These two elements contribute to price volatility. We noticed that in our baseline continuous double auction experiments, more than 1/3 of the contract prices lie outside the equilibrium tunnel. Whereas in previous work,[9] we found that the call market, DD, was always able to discover a uniform price in the equilibrium tunnel.

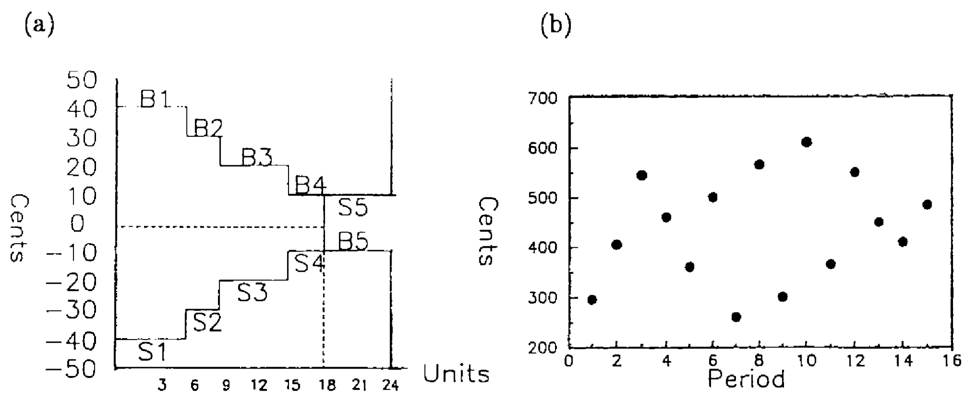


FIGURE 2 Environment E2. (a) Basic supply and demand configuration (resale values and costs are normalized with the mid-point c.e. price equal to zero). (b) C.E. price for periods 1–15.

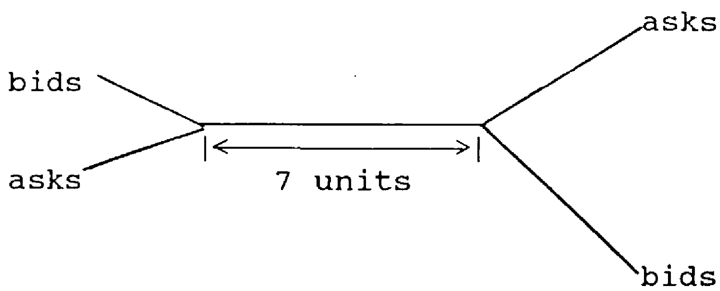


FIGURE 3 Environment.

It is interesting to note that in this environment there exist many competitive/Nash equilibria. They consist of selecting some uniform price in the equilibrium tunnel, and having at least one more winning bid and offer tied at that price than the largest number of units traded by any individual. This prevents price manipulation on the part of any individual, and encourages simple concession to competitive market forces. In our environment all that is required is that at least seven bids and asks be tied at some price in the competitive price tunnel (see Figure 3).

Once this critical volume of tentative trades at a single price has accumulated, no single agent (with a maximum of six units to trade) can shift the supply or demand to adjust the price. (This can occasionally promote price stickiness outside

TABLE 8 UPDA variants using six cells of the $2 \times 2 \times 2$ factorial design

	UPDATE Rule b (1S)		UPDATE Rule a (2S)	
	INFORM		INFORM	
CALL	Rule a	Rule b	Rule a	Rule b
Rule a	8 exps.	8 exps.	8 exps.	8 exps.
Rule b	8 exps.	8 exps.		

the equilibrium tunnel.) Such equilibria, usually exemplified by very little inframarginal revelation of value and cost on the part of buyers and sellers, are natural attractors in a uniform-price call market.

In this paper we compare the behavior of UPDA with DA (the favored field institution) and DD (a call market known to perform well in this environment). We also compare a number of UPDA variants using six cells of the $2 \times 2 \times 2$ factorial design in Table 8. All of these experiments employ the costless order-improving MESSAGE rule.

These experiments were run with undergraduate business majors at the University of Arizona (UA). Subjects were required to make a 2.5-hour commitment to the experiment. They were paid five dollars for showing up plus their salient earnings. When first recruited all subjects were completely without experience in laboratory market experiments. UPDA was programmed on the PLATO computer system as part of the CALL MARKET software system at UA. The subjects independently read their instructions on their assigned computer terminals. Inexperienced groups usually took 30–40 minutes to complete the instructions and begin trading. Experienced groups halved that time. Appendix A contains a typical screen for a buyer in UPDA with rules UPDATE b, INFORM b, and CALL a.

In testing a potential field institution, it is important that subjects be given experience with the institution to see if gaming or strategic behavior emerges. For this reason we ran at least five inexperienced groups from which three experienced groups were randomly selected in each cell of our design.

RESULTS

The parametric environment chosen to motivate subjects in these experiments allows us to make many "independent" observations (one each period) on the same basic market game. Since both prices and relative bargaining strengths are randomly fluctuating, it becomes very difficult for subjects to strategize on the basis of knowledge of the environment. Apparently most of the learning of UPDA subjects can be attributed to reflecting on how to use an institution between experiments, since most cells (institutional designs) display no consistent statistical increase in efficiency from period 1 through period 15. (The only institution that displays consistently strong intra-period learning is the baseline double auction in the first few periods.)

Given our initial experience with our pilot experiments (we happened to choose the worst variant to begin with), we were quite surprised to see some variants of UPDA performing very nearly as well as the baseline 300-second PLATO double auctions we conducted and the double Dutch auctions reported in McCabe, Rassenti, and Smith.[9] The best efficiency performance came from experienced subjects using an endogenous-close, other-side (1S), price update rule and an open book, but they only completed 35 out of 45 potential auctions with one failed auction because of inability to negotiate an initial contract in the allotted time.

The best performance from inexperienced subjects occurs using an endogenous-close, other-side, price update rule, and a closed book. In fact, in the first five periods, it strongly outperforms double auction. Taking inexperience as a proxy for volatile market conditions suggests that sometimes UPDA may be more efficient than the continuous double auction. Double Dutch also displays this property. Finally we note that a very close second best variant of UPDA with both experienced and inexperienced subjects used a fixed-close, other-side price rule and a closed book. For a field application this would be our favored UPDA variant, since not only does it handle varying experience levels, but it can accommodate calls to be made at certain times and with less transactions costs (fewer bids and offers in a shorter time). In some applications, however, the closed book may suffer credibility questions (e.g., the Wunsch Auction was sold largely because of its open book).

The tables in Appendices B and C summarize the results for the six design treatments completed, the baseline double auction, and the double Dutch auction. The subjects in these UPDA experiments do, indeed, display a strong tendency to converge each period to one of the previously discussed competitive/Nash equilibria. A majority of the winning bids and offers are usually tied at the market-clearing uniform price, and in about 85% of the 680 data periods observed, the remaining minority of the untied winning bids and offers reveal less than 10% of the surplus actually collected by the subjects. UPDA is typically characterized by an initial transient phase in which the tentative uniform price is manipulated by the earliest successful bids and offers. But as soon as the tentative trading volume increases to the largest number of units any one player can trade (6), price inertia sets in. The best variants of UPDA can overcome this inertia when the starting price is outside the equilibrium tunnel.

We further organize our results as a series of findings for which we have gathered strong empirical evidence. Each of these findings is followed by a brief discussion of its implications. We begin with a couple of findings concerning execution time which reflect the data contained in Appendix C.

FINDING T1: Under the endogenous-close call rule, open-book auctions took much longer to execute than closed-book auctions.

In fact, with experienced subjects, the closed-book auctions take an average of 264 seconds to execute, less than the 300 seconds allowed for all calls in which the fixed time call was in place. Providing information clearly invites prolonged negotiations, but, as we shall see, does not necessarily affect efficiency in a consistent manner.

FINDING T2: The time to the first trade under the fixed-time call rule ($\mu = 40$) was much greater than under the endogenous-close call rule ($\mu = 15$).

The interesting thing to note is that under the endogenous-close rule, 45 seconds was allowed before the market was called (more than the 40-second average it took to negotiate the first contract under the nonbinding 300-second time constraint of the fixed-time call) yet only an average of 15 seconds was consumed to negotiate the first contract and restart the clock. It seems that as soon as there is an impression of a binding time constraint, messages are hastier. Some times not hasty enough as, indeed, there were three occasions on which the first contract did not transact in time and the auction immediately ended. With hindsight we envision that a more efficient version of the endogenous-close institution might allow longer periods of time to negotiate the first few tentative contracts with the time scale shrinking as quantity increases and price inertia builds.

The following findings are concerned with the period-by-period efficiency of the institutional variants of UPDA which use a fixed-time close.

FINDING F1: When using the both-sides update rule (as does the Wunsch Auction in the field), open-book auctions produce greater trading efficiency than closed-book auctions.

This finding was counter to our intuition that less information should reduce strategic manipulation, especially endgame plays, and favor an outcome close to a competitive/Nash equilibrium. Apparently the book information improves the quality of the feedback more than it encourages gaming. Wunsch made the right choice in opening the book given the update rule chosen.

FINDING F2: Regardless of information conditions the other-side update rule produces greater trading efficiency than the both-sides update rule.

Our pilot experiments with a group of graduate students using the both-sides update rule with a closed book and a fixed close led us to believe that UPDA would not perform very well compared to the double auction or double Dutch. Because of this experience we conceived the other-side price update rule and the endogenous close as possible fixes. Later we were to find that the pilot variant of the UPDA institution is the worst performer of all.

FINDING F3: When using the other-side update rule, a closed book produces greater trading efficiency than the open book.

This result is in contrast to F1 and closer to our intuition that less information is better. Since the price always reflects the other-side's position, information feedback is already good and opening the book tends to cause more endgame congestion and attempted manipulation.

Finally, we consider a finding (E1) concerning the effect of the information condition when using the endogenous close.

FINDING E1: When using the other-side update rule with the endogenous close, experienced subjects produce higher trading efficiencies with the open book than without it, but inexperienced subjects produce the opposite result.

From finding T1 we know that opening the book greatly increases the transactions costs: number of bids and offers made and auction duration. Yet the endogenous close has no endgame (there is always a timed opportunity to respond to any new tentative contract), and experienced subjects learn to use the book only to ensure allocative efficiency. Inexperienced subjects, on the other hand, seem to play each restart interval as if they were approaching a fixed close and misuse the book.

What we have learned from these single-market UPDA experiments will allow us to proceed with confidence in designing call-market variants for complicated economic network systems. We will incorporate a fixed-close, no-book, other-side rule into our next series of natural gas and electric power networks. Although overcoming political impediments seems their first priority, we think the new call-market exchanges that have appeared or are soon to come could learn from our findings and execute their own laboratory experiments. Subtle rule changes such as the other-side update rule can improve efficiency, raise the institution's acceptability by subscribing agents (who are used to continuous double auctions), and hence, reduce the risk of costly field failure.

APPENDIX A: BUYER SCREEN EXAMPLE

id	BID		ASK	id
3	406	1	351	2
▶1	400	2	369	3
4	400	3	378	2
▶1	400	4	385	5
3	398	5	386	3
▶1	388	6	397	4
2	387	7		
		8		

The table above is a sample of what is known as the logbook of the CDC. The columns labelled id show you who submitted the BIDs and ASKs while the inner columns show you exactly what prices were submitted. Notice that all BIDs which belong to you have been pointed out by an ▶. The logbook will be continuously updated by the CDC.

Time: 290

Unit	Value	BID	Accept
1	425	400	✓
2	425	400	✓
→ 3	425	388 ( )	

DATA to BID SELLER ASK: 397
BUYER BID: 388

HELP to change BID unit #.
COPY repeats BID of 388.
TAB increases BID by 5.

APPENDIX B: EFFICIENCY IN THE UPDA MARKETS

The tables in this appendix report efficiency in the UPDA markets. Besides means they include triples (n_1, n_2, n_3) which describe: the frequency n_1 with which efficiency exceeded 93.75% (minimum efficiency when the most valuable fifteen units traded); the frequency n_2 with which efficiency exceeded 84.38% (the minimum efficiency when the price was in the equilibrium tunnel and at least fifteen units traded) but was less than 93.75%; and the frequency n_3 with which efficiency was below 84.38%. In many of the experiments, fewer than fifteen periods of trading occurred in the 2.5 hours allotted. This was sometimes due to excessively long negotiations, and occasionally due to somebody taking a very long time ($>$ 45 minutes) to read the instructions. Whenever this happened it is indicated by the actual number of periods in parentheses beside the experiment name. In three instances when the endogenous CALL (rule b) was in place, the first or second trade was not negotiated in time and trading ended immediately. These cases are indicated by a negative sign followed by the period number in parentheses, and were not included in the tabulation of mean efficiency.

TABLE 9 A. $\underset{(n_1,n_2,n_3)}{\text{eff.}}$

CALL b: endog. close INFORM b: open book UPDATE b: other side					CALL b: endog. close INFORM a: closed book UPDATE b: other side				
Exp. #	1–5	6–10	11–15	4–12	Exp. #	1–5	6–10	11–15	4–12
Inexperienced									
1	88.4	91.0		88.2	3	96.0	88.2	95.5	92.3
(10,–7)	2,2,1	3,0,1		4,0,2	(14)	4,1,0	3,0,2	3,1,1	7,0,2
2	82.2	79.0		78.5	5	96.4	99.6	97.5	98.6
(7)	0,4,1	0,0,2		0,1,3	(12)	5,0,0	5,0,0	2,0,0	9,0,0
17	92.2	89.3		89.3	12	93.8	91.0	97.3	93.2
(10,–6)	2,2,1	2,1,1		2,3,1	(13)	2,3,0	2,2,1	3,0,0	5,3,1
19	88.6	95.8		93.7	18	89.0	94.8	100	92.9
(9)	2,1,2	3,1,0		4,1,1	(11)	2,1,2	4,1,0	1,0,0	6,1,1
21	86.8	94.0	86.0	94.5	20	93.2	98.2		95.7
(11)	2,2,1	4,0,1	0,1,0	6,0,2	(10)	3,1,1	4,1,0		6,1,0
ALL	87.6	91.2	86.0	89.9	ALL	93.7	94.4	96.7	94.5
	8,11,6	12,2,5	0,1,0	16,5,9		16,6,3	18,4,3	9,1,1	33,5,4
Experienced									
8	89.8	100	93.8	99.2	7	98.0	95.8	98.0	96.3
	4,0,1	5,0,0	3,2,0	9,0,0		5,0,0	4,1,0	5,0,0	8,1,0
23	93.6	98.5		97.0	15	84.4	93.0	88.6	89.9
(10,–10)	3,2,0	4,0,0		5,1,0		3,0,2	4,0,1	3,1,1	6,1,2
24	92.0	98.0	98.0	95.5	25	98.0	91.6	93.6	92.4
(11)	3,1,1	4,1,0	1,0,0	6,1,1		4,1,0	3,1,1	4,0,1	6,1,2
ALL	91.8	98.8	94.5	97.4	ALL	93.5	93.5	93.4	92.9
	10,3,2	13,1,0	4,2,0	20,2,1		12,1,2	11,2,2	12,1,2	20,3,4

TABLE 9 B. $\text{eff.}_{(n_1,n_2,n_3)}$

	CALL a: fixed close INFORM b: open book UPDATE b: other side					CALL a: fixed close INFORM a: closed book UPDATE b: other side			
Exp. #	1–5	6–10	11–15	4–12	Exp. #	1–5	6–10	11–15	4–12
Inexperienced									
34	82.0	81.0	95.2	86.9	4	89.2	95.8	97.3	93.9
	0,3,2	2,0,3	4,1,0	4,2,3	(14)	1,3,1	3,2,0	3,1,0	5,3,1
35	85.8	97.0	96.2	94.8	6	95.8	90.2	88.8	92.3
	2,1,2	4,1,0	4,1,0	6,3,0		3,2,0	3,1,1	2,1,2	6,1,2
36	72.0	84.2	97	83.1	11	89.8	85.8	91.0	89.6
	0,1,4	1,1,3	4,1,0	3,2,4		3,1,1	2,1,2	2,2,1	4,3,2
38	90.8	81.8	88.8	86.3	13	93.8	96.0	96.8	95.8
	2,2,1	2,1,2	1,3,1	4,2,3		3,2,0	4,1,0	4,1,0	7,2,0
39	89.6	96.8	94.8	93.9	14	89.6	87.6	93.0	88.3
	2,2,1	5,0,0	4,1,0	7,2,0		2,1,2	3,0,2	1,4,0	3,3,3
41	95.8	91.4	92.0	92.9					
	5,0,0	2,3,0	2,1,1	5,4,0					
ALL	86.0	88.7	94.0	89.7	ALL	91.6	91.1	93.2	92.0
	11,9,10	16,6,8	19,9,2	29,15,10		12,9,4	15,5,5	12,9,3	25,12,8
Experienced									
40	92.0	90.8	89.4	87.4	9	96.0	98.0	92.0	95.9
	2,3,0	3,0,2	2,1,2	3,3,3		4,1,0	4,1,0	3,1,1	7,1,1
42	98.0	94.8	93.2	94.4	16	93.0	92.6	95.2	91.3
	5,0,0	3,2,0	3,2,0	6,3,0		3,1,1	3,1,1	2,3,0	4,5,0
43	96.6	97.2	92.4	94.6	22	94.6	97.0	92.0	95.3
	3,2,0	4,1,0	2,3,0	5,4,0		3,1,1	4,1,0	2,2,1	6,2,1
ALL	95.5	94.2	91.7	92.1	ALL	94.5	95.9	93.1	94.2
	10,5,0	10,3,2	7,6,2	14,10,3		10,3,2	11 ,3,1	7,6,2	17,8,2

TABLE 9 C. $\underset{(n_1,n_2,n_3)}{\text{eff.}}$

	CALL a: fixed close INFORM b: open book UPDATE b: both sides					CALL a: fixed close INFORM a: closed book UPDATE b: both sides			
Exp. #	1–5	6–10	11–15	4–12	Exp. #	1–5	6–10	11–15	4–12
Inexperienced									
26	94.0	91.6	92.4	92.7	44	94.6	91.6	83.8	91.1
	3,2,0	4,0,1	2,1,2	6,2,1		3,2,0	3,1,1	0,3,2	4,4,1
27	91.0	84.4	91.0	88.6	45	86.0	82.3	67.0	80.9
	3,0,2	2,0,3	3,1,1	5,0,4	(-10)	2,1,2	0,2,2	0,0,5	1,2,5
29	95.6	91.0	91.0	93.8	46	86.2	89.6	81.6	87.2
	4,1,0	4,0,1	3,1,1	8,0,1		3,0,2	2,2,1	1,1,3	3,3,3
31	81.4	82.4	78.5	81.6	48	80.2	81.8	79.4	85.8
(14)	2,0,3	1,3,1	1,0,3	2,3,4		2,1,2	3,0,2	1,1,3	5,1,3
32	83.4	97.2	81.2	85.3	48	82.2	79.0	92.6	84.4
	2,1,2	4,1,0	2,1,2	5,2,2		1,0,4	0,4,1	2,2,1	2,5,2
ALL	89.1	89.3	87.2	88.4	ALL	85.8	84.8	80.9	85.8
	14,4,7	15,4,6	11,4,9	26,7,12		11,4,10	8,9,7	4,7,14	15,15,14
Experienced									
30	88.6	96.6	98.2	94.9	47	79.0	94.8	88.2	96.1
	3,0,2	4,1,0	5,0,0	7,1,1		3,0,2	4,1,0	3,0,2	8,1,0
33	90.8	82.8	92.4	86.4	50	80.2	89,4	79.2	82.0
	1,4,0	4,0,1	3,1,1	5,3,1		2,0,3	1,3,1	1,1,3	2,3,4
37	86.0	83.8	91.2	87.2	51	75.0	49.0	68.2	63.4
	2,1,2	0,3,2	3,1,1	2,5,2		0,2,3	0,0,5	0,2,3	0,3,6
ALL	88.5	87.7	93.9	89.5	ALL	78.1	77.7	78.5	80.5
	6,5,4	8,4,3	11,2,2	14,9,4		5,2,8	5,4,6	4,3,8	10,7,10

TABLE 9 D. $\underset{(n_1,n_2,n_3)}{\text{eff.}}$

	Double Auction					Double Dutch			
Exp. #	1–5	6–10	11–15	4–12	Exp. #	1–5	6–10	11–15	4–12
Inexperienced									
417	93.0	97.0	95.4	96.8	dd5	97.1	100	98.8	100
	4,0,1	4,1,0	3,2,0	7,2,0		4,1,0	5,0,0	5,0,0	9,0,0
418	86.2	97.2	99.2	97.4	dd6	90.2	98.8	100	97.5
	1,2,2	4,1,0	5,0,0	7,2,0	(13)	2,1,2	5,0,0	3,0,0	8,0 1
419	88.4	95.0	98.8	96.6	dd7	94.2	91.9	97.9	93.5
	4,0,1	3,2,0	5,0,0	7,2,0	(11)	3,1,1	2,3,0	1,0,0	4,4 0
420	75.4	84.6	92.4	88.6	dd8	94.0	94.0	100	96.2
	0,3,2	1,1,3	2,2,1	2,4,3	(11)	3,2,0	2,3,0	1,0,0	5,3,0
421	83.8	96.0	95.6	96.6					
	2,1,2	3,2,0	3,2,0	6,3,0					
ALL	85.4	93.9	96.3	95.2	ALL	93.9	96.2	99.2	96.9
	11,6,8	15,7,3	18,6,1	29,13,3		12,5,3	14,6,0	10,0,0	26,7,0
Experienced									
422	95.6	99.0	93.8	96.6					
	3,2,0	5,0,0	4,0,1	8,0,1					
423	100	95.4	95.4	96.6					
	5,0,0	2,3,0	2,3,0	5,4,0					
ALL	97.8	97.2	94.6	96.6					
	8,2,0	7,3,0	6,3,1	13,4,1					

APPENDIX C: MEANS AND STANDARD DEVIATION OF PRICE DIFFERENCE

The tables in this appendix report means and standard deviations of price difference (p^*) from equilibrium and execution time (t^*) for first, second, and final contracts in each experiment.

TABLE 10 A. $(\frac{\mu}{\sigma})$

CALL b: endog. close INFORM b: open book UPDATE b: other side							CALL b: endog. close INFORM a: closed book UPDATE b: other side						
Exp. #	p1	t1	p2	t2	pf	tf	Exp. #	p1	t1	p2	t2	pf	tf
Inexperienced													
1	−0.1	21	−4.4	46	−6.0	364	3	−17.9	10	−13.9	24	−2.7	363
(10,−7)	8.5	11	7.6	12	7.7	147	(14)	23.4	3	20.0	7	5.0	159
2	−4.2	16	−5.5	45	−2.1	442	5	1.8	19	1.7	29	7.3	389
(7)	14.6	7	16.4	18	7.8	251	(12)	13.0	9	13.1	12	8.7	91
17	8.7	15	11.9	32	11.1	485	12	−8.1	14	−6.1	28	−9.6	331
(10,−6)	14.4	10	15.1	18	14.2	128	(13)	17.6	8	11.9	14	10.3	89
19	−20.3	15	−17.0	28	−0.5	476	18	0.3	9	2.6	23	3.7	370
(9)	31.2	10	21.4	14	8.9	85	(11)	17.1	5	13.2	8	7.3	103
21	−0.4	11	−3.4	20	−6.7	317	20	3.7	11	5.3	19	3.6	332
(11)	22.6	5	15.0	14	9.6	90	(10)	9.0	5	9.6	8	5.1	187
ALL	−3.1	15	−3.6	33	−1.6	411	ALL	−3.3	13	−2.9	25	0.0	351
	19.9	9	15.7	15	9.9	152		16.7	6	14.0	10	7.6	133
Experienced													
8	9.8	12	8.7	23	7.3	291	7	0.4	14	−0.1	22	−2.9	207
	13.4	9	11.2	9	10.0	150		12.9	9	13.0	11	6.3	70
23	−8.3	15	−7.1	32	−5.1	395	15	−9.5	17	−5.6	33	−2.9	257
(10,−10)	11.1	9	9.6	9	7.4	106		11.9	10	12.3	17	7.5	133
24	−2.5	15	−0.3	27	2.5	514	25	−5.2	12	−7.1	30	−5.6	328
(11)	7.7	9	8.6	10	8.6	107		11.2	7	11.6	16	7.7	121
ALL	1.3	14	1.8	27	2.6	388	ALL	−4.8	14	−4.3	28	−3.8	264
	11.0	9	9.9	9	8.7	122		12.0	9	12.3	15	7.2	111

TABLE 10 B. $\binom{\mu}{\sigma}$

	CALL b: fixed close INFORM b: open book UPDATE b: other side							CALL b: fixed close INFORM a: closed book UPDATE b: other side					
Exp. #	p1	t1	p2	t2	pf	tf	Exp. #	p1	t1	p2	t2	pf	tf
Inexperienced													
34	−10.3	13	−8.2	23	−5.8	283	4	−3.1	14	−1.3	25	7.1	296
	19.0	8	14.1	13	8.3	23	(14)	9.0	7	8.3	14	8.7	5
35	4.5	11	6.9	29	6.3	267	6	−3.9	11	−2.0	19	−2.9	269
	10.3	7	11.5	19	6.8	24		8.9	5	9.8	7	7.0	13
36	4.1	25	5.9	43	−0.2	291	11	−7.6	36	−10.7	54	−8.9	292
	15.1	16	14.8	19	8.7	12		11.8	23	14.3	21	11.0	12
38	8.2	68	6.8	102	0.9	285	13	10.1	27	10.6	53	4.7	290
	17.7	54	15.3	53	7.3	14		18.7	12	15.8	31	8.3	16
39	13.5	9	8.6	18	0.1	291	14	−2.6	12	−5.3	29	−1.0	288
	19.0	3	15.7	7	8.6	9		11.1	9	13.4	18	9.1	10
41	−10.3	13	−8.2	27	3.6	283							
	15.6	6	12.8	11	6.8	25							
ALL	1.6	21	1.9	40	0.8	283	ALL	−1.4	20	−3.8	36	−0.2	287
	16.4	23	14.1	25	7.8	19		12.4	13	12.6	20	8.9	12
Experienced													
40	−8.9	17	−7.6	33	−3.3	280	9	4.1	23	4.8	43	5.9	292
	15.3	9	14.7	14	8.4	20		9.2	13	8.4	19	7.8	12
42	−3.3	20	−3.0	37	−3.5	291	16	17.9	51	13.4	90	6.1	288
	9.2	14	9.7	17	6.7	11		21.5	42	17.0	56	8.6	14
43	−1.9	6	−14.4	15	−3.8	272	22	1.9	52	2.6	84	4.9	285
	10.8	3	19.6	7	7.5	26		10.0	36	9.9	48	10.2	20
ALL	−4.7	14	−8.3	28	−3.5	281	ALL	8.0	42	6.9	72	5.6	288
	12.1	10	15.2	13	7.6	20		14.7	33	12.4	44	8.9	16

TABLE 10 C. $\binom{\mu}{\sigma}$

	CALL b: fixed close INFORM b: open book UPDATE b: other side							CALL b: fixed close INFORM a: closed book UPDATE b: other side					
Exp. #	p1	t1	p2	t2	pf	tf	Exp. #	p1	t1	p2	t2	pf	tf
Inexperienced													
26	−9.3	15	−3.9	30	−2.7	291	44	11.4	103	9.3	122	4.0	298
	18.5	9	9.6	10	8.4	9		16.5	96	14.8	89	10.0	5
27	−5.6	29	−5.3	49	−2.1	279	45	17.6	146	18.4	170	4.2	300
	11.3	33	11.0	41	8.9	33	(−10)	26.9	89	23.0	86	11.0	2
29	−4.7	11	−4.6	18	−5.6	274	46	−15.6	28	−13.5	49	−8.2	300
	15.9	5	15.8	6	8.4	24		12.6	28	12.8	36	8.1	3
31	−11.1	28	−14.3	59	−15.4	288	48	−13.0	91	−13.5	113	−9.3	300
(14)	16.8	28	18.9	44	18.5	9		11.2	73	10.7	75	6.4	1
32	−1.1	15	2.3	34	7.1	284	49	−8.7	53	−7.7	71	10.3	300
	14.1	6	16.4	15	13.3	18		11.0	45	11.1	52	9.1	1
ALL	−6.4	20	−5.2	38	−3.7	283	ALL	−1.9	83	−1.7	101	−4.0	300
	15.5	20	14.8	28	12.2	21		19.2	81	17.5	81	10.8	3
Experienced													
30	14.8	24	8.5	38	3.0	273	47	15.5	173	15.5	187	6.6	299
	15.9	15	12.0	15	6.8	24		16.9	82	16.6	72	7.7	2
33	5.6	79	6.2	110	6.9	293	50	−20.1	205	−13.3	233	−8.8	300
	13.8	71	13.5	66	11.2	10		23.8	70	9.9	41	6.3	1
37	−4.2	43	−3.9	61	−2.5	278	51	−14.7	150	−11.7	169	−12.4	300
	12.3	36	12.7	37	10.0	25		30.3	77	22.6	68	14.0	1
ALL	5.4	49	3.6	70	2.5	281	ALL	−6.4	176	−3.2	196	−4.9	300
	14.1	47	12.8	44	9.5	21		28.5	78	21.4	66	12.8	1

TABLE 10 D. $\binom{\mu}{\sigma}$

Exp. #	Double Auction p1	t1	p2	t2	pf	tf	Double Dutch (Not Applicable)
Inexperienced							
417	−8.3	25.4	−4.1	52.7	0.4	293	
	7.1	36.0	13.8	62.8	7.8	9.6	
418	−0.3	17.2	−4.8	30.8	0.7	297	
	13.5	13.2	9.8	14.2	9.3	4.2	
419	−12.1	19.2	−9.7	35.0	4.2	283	
	19.1	13.9	19.6	14.2	6.6	15.5	
420	−12.3	21.2	−7.3	35.9	−5.9	295	
	17.4	22.4	13.0	31.2	10.3	4.2	
421	4.3	13.5	0.6	33.2	−2.8	289	
	15.2	9.3	13.9	19.1	11.2	12.1	
ALL	-5.7	19.3	−5.1	37.5	−0.7	291	
	16.6	21.2	14.4	33.7	9.2	10.1	
Experienced							
422	−3.2	15.5	0.9	30.8	1.6	296	
	10.7	10.4	9.0	11.7	8.6	3.0	
423	−0.3	44.3	−1.0	61.4	−0.1	297	
	10.6	30.8	10.3	31.2	3.5	2.8	
ALL	−2.9	29.9	−0.1	45.6	0.8	297	
	10.7	23.0	9.7	23.6	6.6	2.9	

REFERENCES

1. Campbell, Joseph, Shawn LaMaster, Vernon L. Smith, and Mark Van Boening. "Off-Floor Trading, Disintegration, and the Bid-Ask Spread in Experimental Markets." *J. Business* **64** (1991): 495–522.
2. Domowitz, Ian. "Automating the Continuous Double Auction in Practice: Automated Trade Execution Systems in Financial Markets." This volume.
3. Friedman, Daniel. "How Trading Institutions Affect Financial Market Performance: Some Laboratory Evidence." Working Paper vol. 221, Department of Economics, University of California, Santa Cruz, March, 1991.
4. Friedman, Daniel. "The Double Auction Institution: A Survey." This volume.
5. Jarecki, Henry G. "Bullion Dealing, Commodity Exchange Trading and the London Gold Fixing: Three Forms of Commodity Auctions." In *Bidding and Auctioning for Procurement and Allocation*, edited by Y. Amihud, 146–154. New York: New York University, 1976.
6. McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith. "Designing 'Smart' Computer-Assisted Markets: An Experimental Auction for Gas Networks." *Euro. J. Pol. Econ.* **5** (1989): 259–283.
7. McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith. "Experimental Research on Deregulated Markets for Natural Gas Pipeline and Electric Power Transmission Networks." In *Research in Law and Economics*, vol. 13, edited by R. O. Zarbe and V. P. Goldberg, 161–189. Greenwich, CT: JAI Press, 1991.
8. McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith. "Smart Computer-Assisted Markets." *Science* **254** (1991): 534–538.
9. McCabe, Kevin A., Stephen J. Rassenti, and Vernon L. Smith. "Designing Call Auction Institutions: Is Double Dutch the Best?" *Econ. J.* **Jan.** (1992).
10. McDonald, J., and B. Jaquillat. "Pricing of Initial Equity Issues: The French Sealed-Bid Auction." *J. Business* **47** (1974): 37–47.
11. Smith, Vernon L. "Experimental Studies of Discrimination Versus Competition in Sealed-Bid Auction Markets." *J. Business* **Jan.** (1967): 56–84.
12. Whitcomb, David, K. "An International Comparison of Stock Exchange Trading Structures." In *Market Making and the Changing Structure of the Securities Industry*, edited by Y. Amihud, T. S. Y. Ho and R. A. Schwartz, 237–255. Lexington, MA: Lexington Books, 1985.

Laura Clauser† and Charles R. Plott‡
†Department of Economics, Occidental College, Los Angeles, CA 90041 and ‡California Institute of Technology, Division of Human & Social Sciences, Caltech 228-77, Pasadena, CA 91125

TWELVE
On The Anatomy of the "Nonfacilitating" Features of the Double Auction Institution in Conspiratorial Markets

Effective conspiracies are more easily formed in posted-offer and in sealed-bid markets than in double auction markets. A feature of double auctions is isolated as the possible source of the behavioral differences. The double auction presents conspirators with continuous temptations to defect from conspiratorial agreements. It also fosters a second type of competition among sellers for access to buyers caused by the fact that only one quotation can be exposed to the market at any instant in time. An extreme case of the restricted exposure feature is the New York rule that requires that only the (first) best offers are exposed to the market and can be replaced only by better offers. This second type of competition can be interpreted as a coordination problem for volume allocation that could interact with other features of the process to undermine conspiracy. The research demonstrates that this second type of competition cannot account for observed differences. When the second type of competition is removed through the creation of a special type of market organization called The Individualized Seller Market Double Auction, the conspiracies still have little or no effectiveness.

The Double Auction Market, Eds. D. Friedman and J. Rust,
SFI Studies in the Sciences of Complexity,
Proc. Vol. XIV, Addison-Wesley, 1993

INTRODUCTION

This study focuses on a question that emerges naturally from literature on the relationship between the details of market institutions and market performance. The existing experimental literature has demonstrated that the successful operation of a market conspiracy is profoundly influenced by the structure of the markets in which they operate. Conspiracies operating within the framework of the double auction do not tend to be successful in influencing prices and quantities to their advantage. On the other hand, conspiracies operating within a sealed-bid environment or within a posted-price environment are successful. The contrast of performance is pronounced.

The "success" of a conspiracy can be analyzed at five different levels.[2] Do potential agents recognize their common self-interest and explore and discuss conspiracies? Does a conspiracy actually evolve in the sense that agents manage to reach an agreement? Is the agreement implemented? Does the implemented agreement actually influence the market? Are the effects to the advantage of the conspirators? The analysis in this paper is primarily concerned with these various possibilities and their relationship to the details of market organization.

Previous studies have provided the following results upon which the study is built. First, it appears that in all markets studied to date, the opportunity to conspire is quickly followed by attempts to conspire. The data show that an immediate recognition of the potential benefits of conspiracy are not beyond the cognitive, perceptual, and moral capacities of humans in general. Somewhat surprising is the fact that the potential for conspiracy translates itself into an actual conspiracy in the sense that a common strategy can be quickly recognized, articulated, and agreed upon. Implementation is usually attempted, but successful implementation and the success of implementation once the implementation is complete are different matters. What follows after implementation is attempted is dependent upon the details of the market organization.

Studies have tended to focus on agents that do not have a long-term relationship or substantial experience in the sense of dozens or more conspiracies. Similarly, studies have not focused on circumstances in which the conspiracy might be enforced by institutions and incentives external to the market. A strong presumption exists that relationships involving time and additional institutional linkages can serve to facilitate successful collusion. Because of this presumption, studies have focused on settings in which neither time nor substantial institutional machinery operate to facilitate conspiracy. Instead, the focus has been on the influence of facilitating devices under circumstances in which the presumption suggests that successful conspiracies will *not* otherwise exist. The questions posed address the features of institutions and organizations that make successful conspiracy easy and why.

1. THE EXPERIMENTAL LITERATURE

Conspiracies were first studied in a laboratory experimental setting by Isaac and Plott.[2] The markets were (parametrically) stationary with each period (day) of market activity involving the same underlying demand and supply configurations. The markets were organized as single-unit double oral auctions with no quotation "improvement rules." The single-unit feature, as opposed to the multiple-unit double auction, is that all quotations must be for single units. The absence of quotation improvement rules means that any new bid (ask) replaces the standing bid (ask) regardless of whether or not it is an improvement. Offer improvement rules (the "New York" rule) means that for a bid (ask) to seize the floor, it must be higher (lower) than the standing bid (ask). In other words, only the "best" offers are exposed to the floor. In the Isaac and Plott experiments, only a single bid and a single ask could occupy the floor at any time, but they could be replaced at any time by any other quotation without regard to whether or not an improvement existed.

Between periods (days) the buyers and sellers were in different rooms. Under the conspiracy treatment conditions, one side (e.g., sellers) was allowed to formulate a joint strategy for the upcoming period. The explicit language was:

> *Except for the bids and their acceptance and except during periods of recess, you are not to speak to any other subject.* During periods of recess you are free to discuss whatever you wish as long as you stay within the confines of the rules: you cannot discuss side payments or make physical threats, and you can't reveal the detailed quantitative information on your payoff charts. Other than those specific things you can discuss all aspects of the market fully.

The results of Isaac and Plott are striking. Conspiracies readily form. Some effects of the conspiracy can be detected in the market. However, the most important possible effect of conspiracies, the phenomenon that is anticipated by theories, laws, and regulations, was not observed. The conspiracy did little to help the conspirators. Conspiracies quickly unraveled, and prices converged to prices near competitive levels.

The study of conspiracies was carried a step further by Isaac, Ramey, and Williams[3] who studied posted-offer markets, and by Isaac and Walker[4] who studied sealed-bid markets. Posted offer involves each seller privately choosing a price and quantity. Once chosen neither can be changed. All offers are exposed to the buyers at the same time. In random order, buyers are allowed to purchase from the seller of their choice. In the absence of conspiracy, posted-offer processes are known to result in prices that are somewhat higher than the competitive equilibrium and have a tendency to converge to near the competitive equilibrium from above. Under conditions of conspiracy, the posted-offer process does not show such a tendency toward the competitive equilibrium. By contrast to the nonconspiracy case, under conspiracy, prices frequently approach the levels predicted by cartel theory and the sellers are advantaged by conspiracy.

Sealed-bid processes are similar to posted-price processes. Bids are tendered in private and cannot be changed once tendered. However, unlike posted-price processes, sealed bids involve the sale of only one unit of a product to the highest bidder. Conspiracy works to the advantage of the conspirators under the sealed-bid processes. In terms of market performance under conspiracy, both posted-price processes and sealed-bid processes exhibit cartel-like performance, but performance under the double auction is competitive.

2. POSSIBLE SOURCES OF INSTITUTIONAL INFLUENCE

The stylized facts are easily stated. People who are not particularly experienced in participating in markets and in conspiracies find themselves involved in a successful conspiracy when participating in one type of market. However, these same people (same subject pool) find themselves attempting to conspire but failing to be effective when participating in a market that is organized differently. Presumably, some feature of the market is instrumental in bringing about the difference in performance. The natural question to pose is "What feature makes such a profound difference?".

Incentive theory and game theory suggests a line of investigation. Notice that the double auction involves only a *temporary commitment* to bids and asks. As such, the double auction presents a continuous temptation to cheat on any previous agreement. Such continuous temptation does not exist under posted prices. The double auction also involves additional competition among the sellers for *access* to buyers. As such, the double auction presents a special coordination problem to conspirators that the posted-price process does not. The coordination problem is related to the fact that only a single ask at a time is exposed to the buyers. A seller wishing to have an opportunity to sell must tender an ask that replaces the ask of the seller that is currently exposed. Any seller that does not compete for access to buyers will have limited access to buyers, and with limited access to buyers, a seller gets a limited market share. Somehow the conspiracy must solve a coordination problem to determine the amount sold by each seller at any agreed-upon price. Under posted prices the allocation of buyers to sellers tends to be determined without any special institution or agreement.

The nature of a conspiracy makes clear both the nature of the temptation and the nature of the coordination problem. The temptation to defect is continuous. Under the double auction process (but not under posted offer or sealed bid), bids and asks are tendered in real time. Since any agent is free to accept the bid/ask of any other agent at any time, any party to a conspiracy has an opportunity to defect at any time. As is the case with any prisoner's dilemma, defection can generate (possibly temporary) personal gains. The opportunity to defect presents itself every moment.

The roots of the coordination problem are also clear. An effective seller's conspiracy will place price above marginal cost so an excess supply exists. The allocation of demand at such a relatively high price is critical to each seller's individual profits. In the double auction and especially the multiple-unit double auction in which an ask can be accompanied by a quantity available at the asking price, access to the buyers, and thus individual sellers volumes, are competitively determined. A seller wanting to have an ask exposed to the market knows that his/her ask will be exposed for a limited time before it is replaced by the ask of a competitor. The success of the seller in making a sale is dependent upon the action of some buyer during the short duration that the seller's ask is exposed. A temptation exists for the seller to try to induce the buyer to take action by undercutting the previous ask. If the New York rules are in place, especially in the case of the multiple-unit double auction, a seller cannot even gain exposure unless he/she undercuts the standing ask which is more than likely the price agreed upon by the cartel. It would seem that a conspiracy must find some mechanism for giving sellers access to buyers and for coordinating the allocation of volume among sellers in some agreeable fashion. Under the rules of the double auction, such coordination requires more than a simple agreement on price.

Neither the continuous temptation nor the access/exposure to buyers (volume coordination) problems exist in the posted-price organization or in the sealed-bid process. In both posted-price and sealed-bid processes, there is no continuous temptation to defect. If a conspirator ever commits to an agreed-upon action, the commitment is final. After the initial commitment there is no opportunity to defect. The market access (coordination) problem is similarly nonexistent. All sellers have equal access to buyers. All sellers will be tendering the same price, so unless some bias exists in the process, or in the way that buyers allocate their purchases across sellers, the volume is probabilistically equally allocated among sellers.

In a sense the existence of an access/exposure to buyer's (coordination) problem, if indeed it is a problem, would seem to reflect a lack of sophistication on the part of conspirators. Conspirators do not recognize the incompleteness of their agreement and do not attribute the incentive system and associated problems of maintaining their agreement to the incompleteness of that agreement. They recognize that they want prices high, but they fail to recognize that higher prices must necessarily be accompanied by lower volume. They focus their collective attention and agree on one dimension of the phenomena, the high prices, but they fail to explore and agree upon the resolution of the other dimension, a reduction of volume. For example, a conspiracy that limits each seller to a prespecified volume and leaves prices unspecified might be more effective. Nevertheless, without volume restrictions, market access becomes a critical issue which the double auction provides no help in resolving, while both the posted-price and the sealed-bid organizations do.

The discussion above brings into focus the main question posed by this paper. Is the access-to-buyers problem (represented by the single outstanding price property) either directly or indirectly responsible for the failure of conspiracies in the double auction institution? That is, if the access-to-buyers problem is removed within the multiple-unit double auction framework, do the conspiracies still fail? If we discover that the removal of the access-to-buyers problem is not accompanied by successful conspiracies, then we will have discovered the general location of the conspiracy disrupting feature that we seek to identify. The conspiracy disrupting feature must reside in other features of the pricing environment such as the temporary commitment feature and the resulting continuous temptation, the interactions with buyers, etc.

3. THE INDIVIDUALIZED SELLER MARKET DOUBLE AUCTION (ISMDA)

The experimental design used in this study involved a change in the double auction to allow equal access to buyers for all sellers. This was achieved by opening a separate "market" for each seller.[1] Each seller could accept bids or tender any ask desired at any time in his/her own market, but could not tender asks or accept bids in the market of other sellers. Buyers could tender bids in the market of any or all sellers or could accept asks in any or all markets. The only overriding rules were: (i) the New York improvement rules applied to each market independently, and (ii) after 15 seconds market exposure, a bid or ask could be canceled. Thus, the values of bids and asks could differ across markets and buyers had the freedom to cancel an offer made to one seller if he/she wanted to make an offer to a different seller without being exposed to the possibility of buying in both markets. The 15-second rule allowed adequate time for a seller, or buyer, to execute a transaction.

This new market institution, which we will call the Individualized Seller Market Double Auction (ISMDA), can be used to test for the source of the organizational influences observed under conditions of conspiracy. The ISMDA organization removes all problems of access to buyers that might exist for sellers. That is, the competitive access feature of the double auction has been replaced by the unrestricted access feature of posted price. Relative volume can be determined in much the same way as it is under posted prices. If sellers ask the same price, then buyers will simply allocate themselves across sellers, and if no bias exists, the expected volumes of all sellers should be the same. Thus, the ISMDA is the double auction institution. It is characterized by the temporary commitment to asks of the double auction, and thus it has the continuous temptation property of the double auction, but it does not have the access to buyers problem (the coordination problem). The ISMDA has all of the other features of the double auction.

From another perspective, ISMDA is like a posted-price process with the permanent commitment to asks replaced by the temporary commitment of the double auction. The ISMDA has the unrestricted access to buyers aspect of the posted-price institution instead of the competitive access of the double auction.

[1]Similar market organizations have been used in some previous experiments. See for example Miller and Plott,[7] or Lynch, Miller, Plott, and Porter.[6]

4. INSTITUTIONAL COMPARISON AND EXPERIMENTAL DESIGN

The analysis of this paper has focused on two features of otherwise very different appearing institutions. The institutional variables identified so far are the mode of access to buyers (competitive vs. unrestricted) and the commitment to price quotations (temporary vs. permanent). Yet there are many additional features of the two central institutions (double auction and posted prices) which could be of key importance. Double auctions operate in real time, and the timing and magnitude of quotation changes might be important. Posted-price processes operate faster than double auctions. The single pricing decision in the posted-price institution is more important than any particular quotation decision in the double auction. Buyers are less active in the posted-price process than in the double auction, and there would seem to be less latitude for seller beliefs about buyer behavior to effect seller decisions under the posted-price process. Briefly stated, the micro-market structure is very complex, and a large number of features exist that one might imagine could cause behavioral differences.

The method of dealing with the institutional complexity is to proceed on the bases of some central assumptions that will be maintained throughout the analysis and interpretations. In particular, a *Hypothesis of Institutional Equivalence* will be invoked. The hypothesis will assert that certain institutional differences will have no impact on the variables of interest.

The Hypothesis of Institutional Equivalence consists of the three assumptions that follow:

1. The only substantive differences in the institutions are:
   i. the mode of access to the buyers (competitive vs. unrestricted); and
   ii. the commitment to price quotations (temporary vs .permanent).
2. ISMDA is the posted-price institution with the double auction price quotation feature of (ii).
3. ISMDA is the double auction with the posted-price access feature of (i).

The role of the Hypothesis of Institutional Equivalence is to enable us to make institutional comparisons. It allows us to model complex institutions in terms of their features. Figure 1 contains the two variables identified by assumption (1). The double auction is located in the upper left box of the figure as the combination of the temporary quotation commitment feature and competitive access-to-buyers feature, while posted prices are located in the lower right. The ISMDA institution is located in the upper right and is defined as the double auction with the posted-prices access feature or, equivalently, as the posted price with the double auction commitment feature. The lower left is an institution in which the prices are posted, but only the lowest posted price is exposed to the market (competitive access).

Access to Buyers

Quotation Commitment		Competitive	Unrestricted
	Temporary	Double Auction	ISMDA =DA + Posted Price Access =Posted Price + DA Quotation Commitment
	Permanent	Posted Price with Market Access Only Given to Lowest Priced Seller	Posted Price

FIGURE 1 Institutional Comparisons.

Of course, this latter institution could be very inefficient if all price quotations are above the competitive equilibrium and not the same across sellers. Only the low-priced seller could sell, and when that seller sold his/her desired volume, buyers could not purchase even at prices quoted by (high-priced) sellers. Other modes of competitive access might be imagined such as the first to enter or the maximum volume offered, etc. This process is included for completeness.

A brief summary of previous results and stylized facts placed in the context of the Hypothesis of Institutional Equivalence helps identify the key feature of any new experimentation. Figure 2 contains the summary. First, collusion is never effective in the absence of much repeated interaction unless very special institutions are present.[2] Pilot experiments under ISMDA demonstrated that in the absence of conspiracy under ISMDA, prices converge to the competitive equilibrium as one would naturally expect. Thus the area outside the conspiracy box in Figure 2 is shaded as conditions under which near competitive performance is observed. The only known effective conspiracies are those founded in posted-price and sealed-bid

[2] See Grether & Plott[1] for the appropriate example.

institutions. The lower left of the figure is included as a conjecture/assumption supported by the general pattern of previous results. The concept of "competitive access" under permanent quotation conditions is a bit ambiguous. For purposes of analysis it is assumed that access is determined by the best quotation. A strong presumption exists that effective conspiracies would not be eliminated by a practice of excluding higher priced agents from the market in a posted-price institution. Thus by conjecture (assumption), the analysis and interpretation of results are based on the assumption that the results of Isaac, Ramey, and Williams will extend to the lower left box. However, because of the ambiguity associated with the concept of "competitive access under conditions of permanent quotation commitments," a more complete examination of this topic is left for further research.

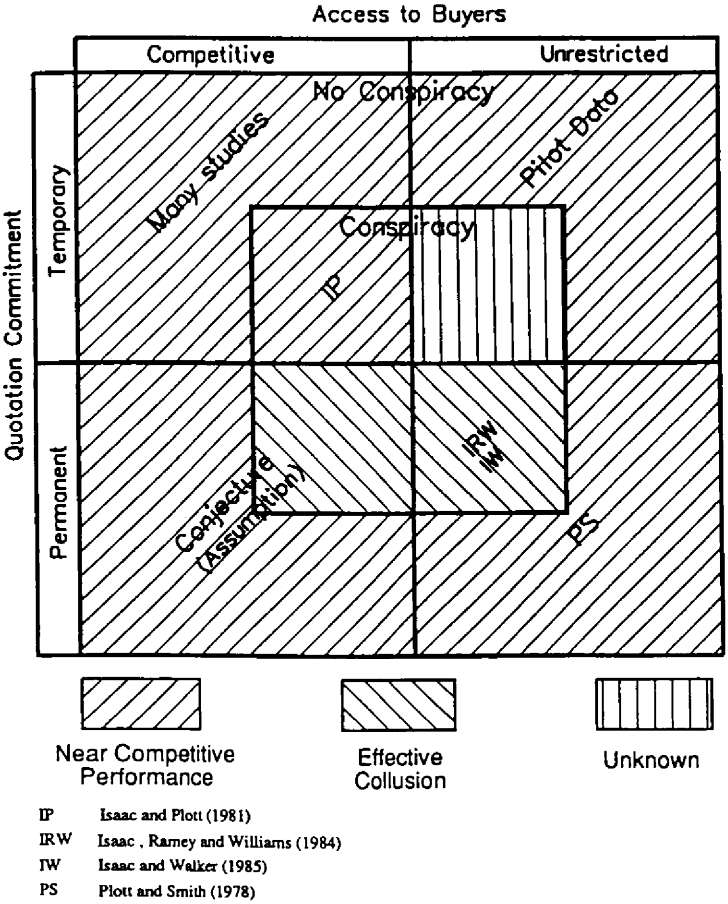


FIGURE 2 Previous Results of Institutional Performance.

Given the pattern of known results and assumptions listed above, open questions are those that exist under the ISMDA institution operating under conditions of conspiracy. Thus, the appropriate experiment to conduct is ISMDA when agents are allowed to conspire. If a successful collusion is observed, we can then conclude that the competitive access to buyers of the double auction is a key feature that acts as a nonfacilitating device for conspiracy. If successful collusion is not observed, then we can conclude the opposite—that access to buyers is not a facilitating device. Furthermore, under Hypothesis of Institutional Equivalence, we can derive the even stronger conclusion that the temporary commitment to quotations and the accompanying continuous temptation to defect is the key institution that undermines conspiracy.

5. EXPERIMENTAL DESIGN, PARAMETERS AND PROCEDURES

The relevant experiment is easily identified. Conduct ISMDA under conditions of conspiracy. Four experiments were conducted at the California Institute of Technology. Subjects in the four experiments were from Caltech (experiments 2 and 4) and Occidental College (experiments 1 and 3). Most of the subjects had no previous experience in experimental markets.

All experiments involved eight people. Four were designated as buyers and four were designated as sellers. Preferences were induced by application of induced preference theory. The redemption values and costs of individual buyers and sellers are in Table 1. When aggregated to the market level of analysis, the individual parameters are as appear in Figure 3. Parameters were identical for all periods for all individuals.

Instructions were read to all subjects. These are available from the authors on request. Buyers were located in one room and sellers in a different room. Instructions were read separately to each group. The delicate part concerning the possibility of conspiracy was not read to buyers.

All participants were trained to use the computerized Multiple-Unit Double Auction program[5] by participating in a stand-alone computerized instruction package which demonstrates the location of keys and their function. Accounting and other instructions were administered verbally.

Each seller was assigned to a separate market. Sellers were unable to tender asks or to accept bids in the markets to which they were not assigned. Buyers were free to tender bids or accept asks in any of the four markets. Screen displays were such that the standing bid and ask were displayed in each market along with the identification of the agent who tendered the action. Similarly, a history screen revealed transactions in each market and the people involved. Thus, the markets were characterized by full information about bids and contracts. Of course, the redemption values and costs were private information.

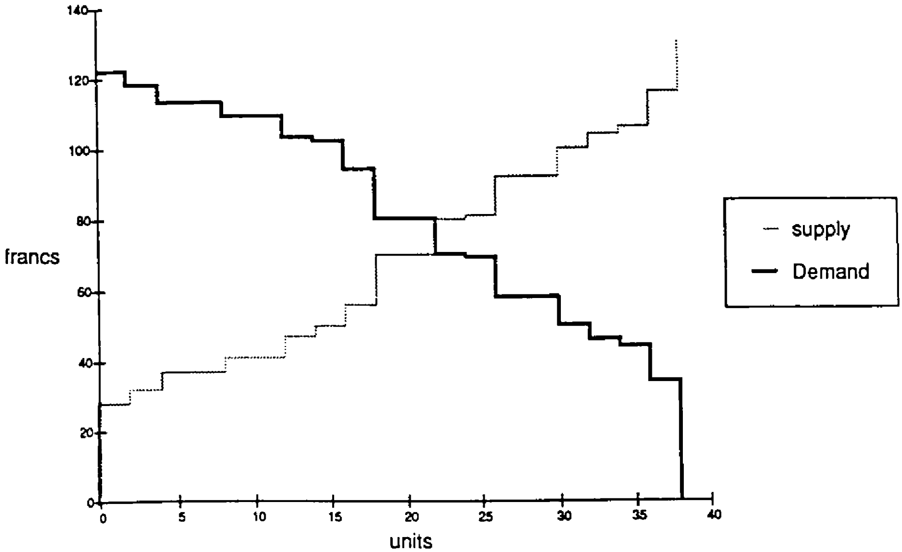


FIGURE 3 Demand and Supply.

TABLE 1 Induced Values in Dollars.

	Seller Costs		Buyer Redemption Values	
I.D. No./ Unit	0,1	2,3	4,5	6,7
1	.32	.28	1.18	1.22
2	.37	.37	1.13	1.13
3	.41	.41	1.09	1.09
4	.47	.50	1.03	1.00
5	.56	.70	.94	.80
6	.70	.81	.80	.69
7	.80	.92	.70	.58
8	.92	1.00	.58	.50
9	1.04	1.06	.46	.44
10	1.16			.34

Periods lasted for five minutes. Two minutes were allowed between periods for profit calculations and for discussion among the potential conspirators.

6. MODELS

Two models are of interest. These will be used as benchmarks against which the influence of the institutions can be assessed.

The first is the competitive model. As is clear from Figure 3, the competitive equilibrium is in the price range of [$.70, $.80] and the volume is 22 units. Price movements in the range of the competitive equilibrium would be evidence that the conspiracy is not successful. Of course, market efficiency should be 100% at the competitive equilibrium.

The second model is that of a profit-maximizing cartel (with no side payments). Joint profits of sellers are maximized at a price of $1.00 and a volume of 16 units. Table 2 contains the profits for each of the two types of sellers based on the assumption that market volume is split evenly between the two types. All sellers prefer the price of $1.00 and an equal split of volume to the competitive equilibrium price and volume. Compare profits for individuals 0 and 1 of $2.43 per period on average when the price is $1.00 to the competitive equilibrium profits of $1.87 each. Similarly, sellers 2 and 3 would prefer an average profit of $2.44 per period at a price of $1.00 to the competitive equilibrium of $1.74. These figures assume that all surplus goes to the sellers.

TABLE 2 Theoretical Profit Levels for Different Seller Types at Different Market Prices.

	Seller Types 0,1		Seller Types 2,3	
Price	Combined Profit	Combined Volume	Combined Profit	Combined Volume
1.22	.94	1	.90	1
1.18	1.67	2	1.62	2
1.13	3.14	4	3.22	4
1.09	4.34	6	4.42	6
1.03	4.54	7	4.59	7
1.00	4.86	8	4.88	8
.94	4.76	9	4.64	9
.80	3.94	12	3.48	10

A successful conspiracy that is unable to implement price discrimination would choose a price of $1.00. The market efficiency would be 90.82 percent. Of course, if price discrimination becomes possible, the conspiracy would do much better by simply "walking down" the demand curve.

7. RESULTS

The graphs of the time series of bids, asks, and contracts are contained in Figures 4 through 7. Dollars are on the vertical axis and clock time (seconds) is measured on the horizontal. Vertical lines indicate the open and close of periods. Horizontal lines indicate the price predictions of the two models ($1.00 for the cartel model and [$.70,$.80] for the competitive model). Contract prices in a given market are connected by a line. Thus, the figures contain four time-series lines of contract prices, one for each of the four different sellers, each of which is operating in a different market. Dots above the contract lines tend to be the asks, pooled across all four markets, and the dots below the contract lines tend to be the bids, pooled across all four markets.

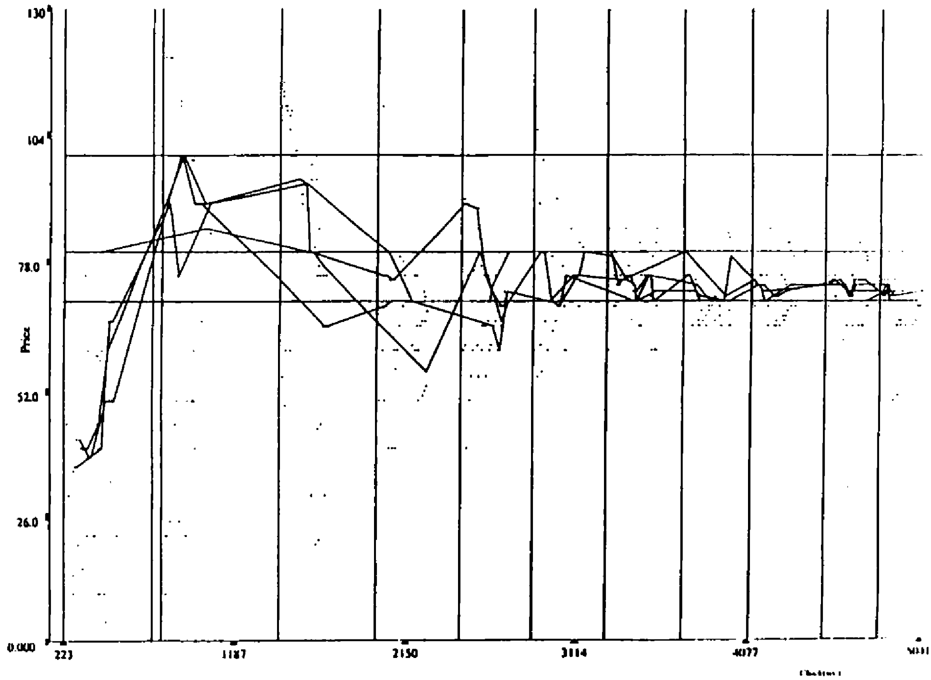


FIGURE 4 Experiment 1—Bids, Asks, and Contracts.

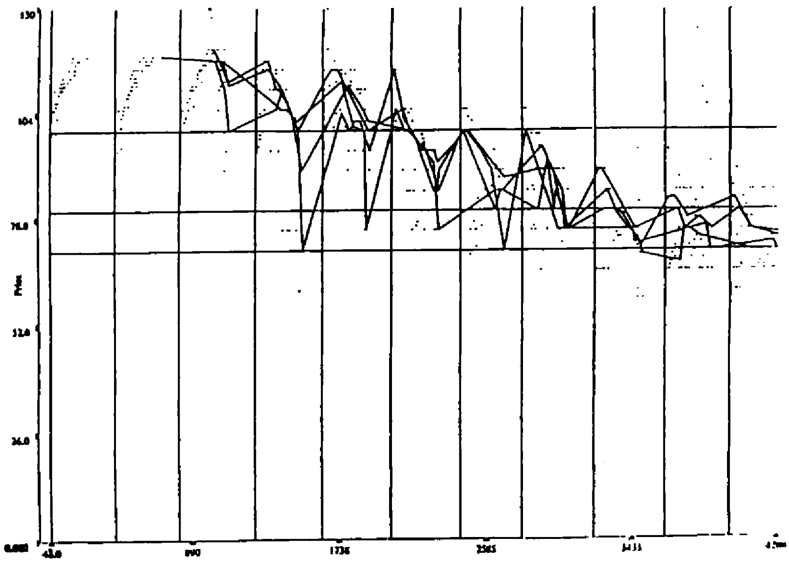


FIGURE 5 Experiment 2—Bids, Asks, and Contracts.

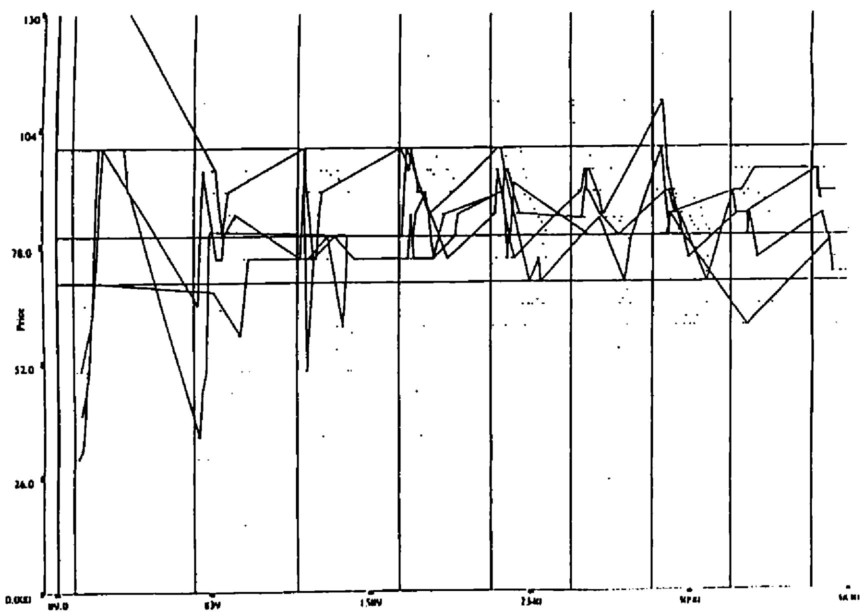


FIGURE 6 Experiment 3—Bids, Asks, and Contracts.

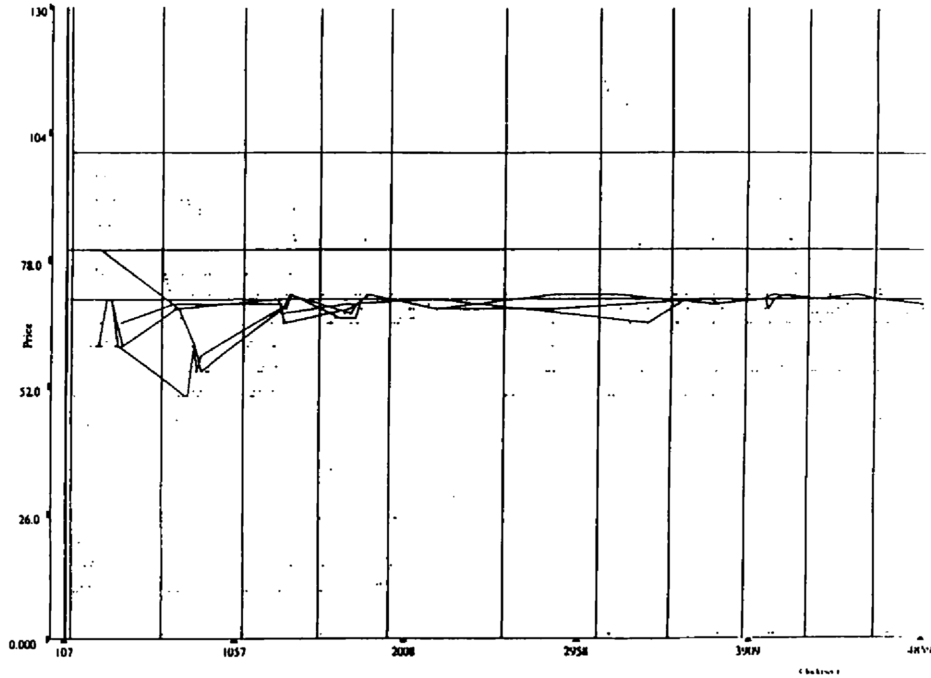


FIGURE 7 Experiment 4—Bids, Asks, and Contracts.

CONCLUSION 1

Conspiracies formed in all markets.

SUPPORT

The experimenter was present and could hear all conversations. In all experiments, conspiracies formed and attempted to maintain a price. In all experiments, except experiment 3, the conspiracy formed within the first few minutes of contact before the first period opened. In experiment 3 the conspiracy formed in period 5. In experiment 4 the conspiracy agreed to enforce a competitive equilibrium price of only $.70 after the second period. In retrospect the report could have been more precise if the experimenter had maintained a log of the number of seconds that elapsed before the words "fix price" or phrases with equivalent meanings were mentioned along with other suggestions of agreement, e.g., "o.k."

CONCLUSION 2

In all experiments volume was closer to the competitive equilibrium than that of the cartel model.

SUPPORT

Volume figures are in Table 3. In the last two periods of all experiments, the volumes are nearer the competitive equilibrium volume of 22 rather than the cartel volume of 16 units. The last five periods of the four experiments averaged values of 22, 20.6, 21.4, and 14, respectively. The volume in experiment 4 is a clear exception. Close examination reveals that some buyers were refusing to buy during periods 6, 7, 8, and 9. This might have been an attempt to collude by the buyers. The low volume was a result of these actions by the buyers and not the success of a conspiracy of sellers. Thus the data from experiment 4 also supports the conclusion.

CONCLUSION 3

In all experiments contract prices converged to the competitive equilibrium price range or were converging toward the range during the last few periods.

TABLE 3 Average Prices (Cents) and Volume

	Experiment 1		Experiment 2		Experiment 3		Experiment 4	
Period	Price	Volume	Price	Volume	Price	Volume	Price	Volume
1	56	24	0	0	79	18	69	22
2	90	15	118	1	84	25	63	18
3	81	22	114	13	104	26	68	21
4	69	21	104	17	86	21	69	21
5	73	23	103	19	87	22	69	21
6	79	19	96	23	85	21	69	11
7	73	23	89	19	86	21	71	11
8	73	23	83	20	85	21	69	7
9	72	22	79	23	87	22	70	14
10	73	19	76	21			70	20
11	72	23	74	20			70	17

TABLE 4 Profit per Period.

Theoretical	Competitive Model Buyers	Competitive Model Sellers	Cartel Model Buyers	Cartel Model Sellers
	632	632	174	974
	Experiment 1		**Experiment 2**	
Experimental Period	Buyers	Sellers	Buyers	Sellers
1	1075	60	0	0
2	291	716	4	86
3	418	756	-15	877
4	674	349	73	1030
5	660	499	66	1123
6	675	445	137	1053
7	633	576	354	849
8	651	458	443	733
9	690	553	545	685
10	607	409	583	641
11	706	526	542	599
	Experiment 4		**Experiment 5**	
Experimental Period	Buyers	Sellers	Buyers	Sellers
1	404	472	755	348
2	245	702	663	403
3	-187	1283	573	480
4	392	822	714	499
5	379	875	-57	489
6	411	773		
7	397	829		
8	412	677		
9	363	681		
10				
11				

TABLE 5 Efficiency in Percent.

	Experiment			
Period	1	2	3	4
1	89.32	0	73.89	87.34
2	78.88	6.8	74.92	84.33
3	92.88	68.2	86.71	83.31
4	80.93	87.26	96.84	95.97
5	92.33	94.07	99.20	34.17
6	83.70	94.15	93.67	59.97
7	95.65	95.17	97.23	54.35
8	87.74	93.04	86.16	78.16
9	98.34	97.31	86.16	97.31
10	80.70	96.84		97.31
11	97.47	90.27		87.10
12				

SUPPORT

Average prices for all periods of all four experiments are displayed in Table 3. In experiments 1, 2, and 4 prices are in the competitive price range of [$.70, $.80] by the last two periods. In experiment 3 three prices are closer to the competitive range than the cartel prediction of $1.00 in all periods but one. Experiment 4 is interesting because prices never exceeded the competitive range even though an active conspiracy was operating.

CONCLUSION 4

Conspiracies were not uniformly successful in getting seller profits above the competitive equilibrium. In no case were profits maintained near the cartel levels.

SUPPORT

Of the 36 periods reported, 18 involve seller profits that are greater than the competitive equilibrium profit levels. In only nine of these were seller profits closer to the cartel model than the competitive equilibrium. Generally profits were closer to the competitive equilibrium than cartel levels.

CONJECTURE

Conspiracies reduced market efficiency by contributing to efficiency variability.

SUPPORT

No controls exist that would permit convincing measurements to be made of differences in market efficiencies that might accompany conspiracies. The support only comes from a judgement by the authors about the properties of efficiency time-series characteristic of other markets. Typically efficiencies increase in a near monotone fashion until near 100%.

CONCLUSION 5

The continuous temptation to defect feature of the double auction exerts an influence capable of producing the "nonfacilitating" property independent of any influence of the competition among buyers for access to the market.

SUPPORT

Together conclusions 1 through 4 lead to conclusion 5. Conspiracies evolved but failed to be effective even though the buyer access problem had essentially the same features as the posted-price process. The market behavior of the individualized seller market double auction is similar to the market behavior under the standard double auction.

8. SUMMARY OF CONCLUSIONS

When operating within posted-price markets or sealed-bid markets, inexperienced people readily participate in and implement successful conspiracies. The same cannot be said of inexperienced people operating within the framework of a multiple-unit double auction. The research presented in this report provides a step in isolating the particular feature(s) of the organizations that foster this difference in behavior.

Two broad aspects are identified as potential key differences in the structures of the institutions. These are features of the double auction that are not present in the other two institutions: (i) a continuous temptation to defect from a conspiratorial agreement, and (ii) competition for access to buyers, fostered by rules which allow only one quotation to be exposed to the market. The primary question posed is whether the second of these features can account for the nonfacilitating nature of the double auction.

The answer suggested by the research is that the access-to-buyers feature of the double auction cannot account for the observed differences in behavior of the two institutions. If the Hypothesis of Institutional Equivalence is accepted, an even stronger conclusion is at hand. Some feature or features of the continuous temptation-to-defect property constitute the important conspiracy-breaking ingredient. In the presence of the continuous temptation to defect, rules that limit seller access to buyers have no overriding influence on the unraveling of a conspiracy. In the presence of the continuous temptation to defect, the conspiracies unravel whether the competitive access to buyers is in effect (MUDA) or not (ISMDA).

Of course, left somewhat unexplored is the possibility that special competitive-access processes might be devised that have an independent unraveling influence. At this stage of research, it is not obvious what form such similar devices might take. Posted prices remove the continuous temptation property and could be a natural institution to which to attach a new device, and the lower left corner of Figure 2 might be a place to start. Ideas along these lines have not been pursued, and so remain available for later studies with more appropriate technology and theory. In addition, the force of the Hypothesis of Institutional Equivalence needs to be dissected. While the temporary commitment property appears to contain a key to nonfacilitation, the property itself contains many variables. In particular the behavior of buyers and the beliefs of sellers about the behavior of buyers should not be overlooked as possibly being important.

In retrospect, the results reported here can be linked to other results to identify a possible continuity of observations. Smith[8] demonstrated that monopoly has a difficult time charging monopoly prices when operating within the double auction and does much better under posted prices. A monopolist has no access to buyer problems, even under the extreme competitive access rules such as the New York rule. In addition, the monopoly is involved in no conspiracy so there is no temptation to "defect" except perhaps from his/her previous decisions. It seems reasonable to suspect then that the sequential nature of actions and information, together with the possible (counter-speculative) activities of buyers, must be explored in greater detail before the central elements of the conspiracy-breaking properties of the institution can be isolated.

ACKNOWLEDGMENTS

The research support of the National Science Foundation and the Caltech Laboratory of Experiments in Economics and Political Science is gratefully acknowledged. Clauser is a senior at Occidental College and Plott is a professor at Caltech. The comments of Mark Isaac led to several changes in this paper and are greatly appreciated.

REFERENCES

1. Grether, David M., and Charles R. Plott. "The Effects of Market Practices in Oligopolistic Markets: An Experimental Examination of the Ethyl Case." *Econ. Inquiry* **22** (1984): 479–507.
2. Isaac, R. Mark, and Charles R. Plott. "The Opportunity for Conspiracy in Restraint of Trade." *J. Econ. Behav. Org.* **2** (1981): 1–30.
3. Isaac, R. Mark, Valerie Ramey, and Arlington W. Williams. "The Effects of Market Organization on Conspiracies in Restraint of Trade." *J. Econ. Behav. Org.* **5** (1984): 191–222.
4. Isaac, R. Mark, and James M. Walker. "Information and Conspiracy in Sealed Bid Auctions." *J. Econ. Behav. Org.* **6** (1985): 139–159.
5. Johnson, Alonzo, Hsing-Yang Lee, and Charles R. Plott. "Multiple Unit Double Auction User's Manual." Social Science Working Paper No. 676, California Institute of Technology, 1988.
6. Lynch, Michael, Ross M. Miller, Charles R. Plott, and Russell Porter. "Product Quality, Informational Efficiency, and Regulations in Experimental Markets." In *Res. Exp. Econ.*, vol. 4, edited by R. Mark Isaac, 268–318. Greenwich, CT: JAI Press, 1991.
7. Miller, Ross, and Charles R. Plott. "Product Quality Signaling in Experimental Markets." *Econometrica* **53** (1985): 837–872.
8. Smith, Vernon L. "An Empirical Study of Decentralized Institutions of Monopoly Restraint." In *Essays in Contemporary Fields of Economics in Honor of Emanuel T. Weiler (1914–1979)*, edited by George Horwich and James P. Quirk, 83–106. West Lafayette, IN: Purdue University Press, 1981.

Colin F. Camerer† and Keith Weigelt‡
†Graduate School of Business, University of Chicago, Chicago, IL 60637, e-mail: fac_camerer@gsbacd.uchicago.edu; and ‡Department of Management, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104

THIRTEEN
Convergence in Experimental Double Auctions for Stochastically Lived Assets

We report market experiments in which subjects trade stochastically lived assets that pay a dividend each period and live from period to period with a known probability (.85). Since the probability is like a discount rate, the experiments test whether prices converge to discounted dividend value. We find that prices converge slowly to the discounted dividend value, from below; experienced subjects traded at prices close to the discounted dividend value. The forces creating convergence in this setting appear to be fundamentally different than in most other settings: minimal rationality in information-rich auctions is insufficient to produce convergence. Instead, convergence seems to be caused by one speculator buying much of the supply and holding it for several periods, which causes prices to rise. When prices overshot the discounted dividend value, they were slow to converge downward. Trading volume and allocative efficiency were often low during convergence (due to the speculator's behavior).

The Double Auction Market, Eds. D. Friedman and J. Rust.
SFI Studies in the Sciences of Complexity,
Proc. Vol. XIV, Addison-Wesley, 1993

INTRODUCTION

We conduct experimental markets in which subjects trade a stochastically lived asset that pays a dividend each period and lives from period to period with a known probability. The probability that the asset lives is formally equivalent to a discount rate, so we can test whether prices converge to the value of discounted dividends.

Our main interest in this paper is convergence dynamics in double auctions for a stochastically lived asset. Rapid, reliable convergence to competitive equilibrium has been observed in most experimental double auction markets for commodities or single-period assets. Convergence is often slower, but still reliable, in most markets for multi-period assets (see Sunder[39]; the "bubble" experiments of Smith, Suchanek, and Williams,[37] among others are important exceptions we mention below). It appears that in simple settings, double auctions can magically transform well-specified preferences into competitive equilibrium prices, even when traders have only minimal amounts of information and rationality (see Gode and Sunder, this volume). We wanted to explore the limits of this ability to produce convergence.

We found that when subjects are inexperienced, convergence is typically slow and erratic (convergence paths vary across experimental sessions). We conclude that some features which generate convergence in simpler experiments are missing in our markets. For example, in commodity markets the cost and value of commodities is clearly induced by experimental design. By contrast, in our markets next-period prices are an important component of an asset's value in the current period. But the future-price component of asset value is created by traders' actions, rather than controlled by our design, so insightful traders must deduce it by calculation, or guess it after observation. We suspect convergence is slow because most traders lack such insights. Convergence was sparked, to some extent, by speculators who apparently figured out the equilibrium price, then bought units cheaply and held them until prices rose near the equilibrium. Thus, while double auctions work beautifully in simpler settings with minimal trader information and rationality, we conclude that in the absence of rational calculation by most traders, aggressive speculation by one or more rational traders is required when assets are stochastically lived.

A secondary motivation for our experiments is the debate about whether asset prices are rational forecasts of discounted dividend value (see reviews by Camerer[3] and Leroy[25]). In most studies with natural data, the discounted dividend value of an asset is not known, so judging its correspondence with asset prices is difficult. But we know the discounted dividend value in our experiments (because we control it), so departures from rational pricing are easily observable. While markets with inexperienced subjects converged erratically, prices did move, slowly, toward the discounted dividend value (sometimes overshooting it and crashing). In a second trading session, experienced subjects traded at prices quite close to the discounted dividend value.

TABLE 1 Experimental Design Parameters

Session	No. of Subj.	Location & Type of Subject	Dollars/ Franc	Endowment (in francs) Initial	Each Period	No. of Periods	Subjects—Made Forecasts?
Inexperienced Subjects							
1	10	NYU, MBA	$.001	7000	400	27	no
2	8	NYU, MBA	$.001	7000	400	30	no
3	8	NYU, MBA	$.001	7000	400	30	no
5	12	Penn, undergrad	$.002	6000	0	21	yes
6	9	Penn, undergrad	$.0015	6000	0	25	yes
8	9	NYU, MBA	$.001	6000	0	38	yes
11k	12	Penn, undergrad	$.001	10000	0	8	yes
12k	12	Penn, undergrad	$.001	10000	0	13	yes
13r	12	Penn, undergrad	$.001	10000	0	12	yes
Mechanism Experienced							
4m	9	Penn, undergrad	$.001	7000	400	20	yes
9m	9	Penn, undergrad	$.001	6000	0	14[1]	yes
Context Experienced							
7c	9	Penn, undergrad	$.0015	6000	0	30	yes
10c	9	Penn, undergrad	$.0015	7000	0	17[1]	yes

[1] Trading periods were 6 minutes long; all other trading periods were 4 minutes long.

THE DESIGN

Subjects traded assets which paid a dividend at the end of each period to the asset holder. In contrast to previous asset experiments, our assets were "infinitely lived" in a specific sense: At the end of each period, all the assets expired worthless with a probability .15, or they lived to the next period with probability .85.

In most sessions there were three dividend levels—high, medium, and low. The dividends earned by any one subject rotated across these levels in a three-period cycle (mimicking cyclical preferences due to life-cycle effects or stochastic taste changes). Subjects' dividend rotations were staggered so that one third of the subjects had high dividends in any one period, one third had medium dividends, and one third had low dividends.[1] (Actual dividend levels are shown in Table 2.)

This design has three advantages: (i) Demand, aggregated across traders, was the same every period so the discounted dividend value was constant, which made convergence easy. (ii) Since each subject's dividend level changed from period to period, subjects had a financial incentive to trade (fulfilling Smith's[36] "dominance" precept). (iii) We can measure the allocative efficiency of the market—did the high-dividend traders hold assets each period?—and test whether price fluctuations harm efficiency.

As is standard in economics experiments, in most sessions subjects knew only their own dividends in current and future periods (see Smith's "privacy" precept[36]), but they were told "dividends may be different for different investors." However, dividends were common knowledge in two sessions.

COMPETING THEORIES OF PRICES AND VOLUME

In this section we derive predictions about prices and trading volume in competitive equilibrium. Each subject was endowed with two asset units and several thousand "francs" of currency, at the beginning of a session. Traders cannot issue units or sell short by borrowing units and repaying them. Figure 1 shows demand and supply curves for a representative period of sessions 8–10c, which have nine subjects. (Other sessions are similar.) Note that endowments will not be equal each period because of trading in previous periods.[2]

Suppose initially that the asset lives only one period. Pricetaking traders supply 12 units at a price of zero (6–9 units from medium-dividend (M) traders and 3–6 units from low-dividend (L) traders) and 3 units at a price of 100 (by M-traders). Nine units are demanded by high-dividend (H) traders at a price up to 300. They also demand as many additional units as they can afford (within their budget constraint) at a price of 100.

[1] For example, subjects 1, 7, and 9 earn high dividends in period 1, medium dividends in period 2, low dividends in period 3, high dividends in period 4, and so on. Subjects 3, 5, and 6 first earn medium dividends in period 1, then low dividends, high dividends, medium dividends, etc. Subjects 2, 4, and 8 first earn low dividends in period 1.

[2] Define high-, medium-, and low-dividend traders to be H-, M-, and L-traders. At the beginning of a typical equilibrium period, at least twelve units will be held by M-traders (who were H-traders in the previous period) and three units will be held by L-traders (who were M-traders previously). The three other units will be held either by L- or M-traders. This indeterminacy in holdings is clumsy, but it arises from the fact that H-traders earn positive dividends for each extra unit they buy, which is necessary to generate excess demand at prices below the competitive equilibrium (see footnote [4]).

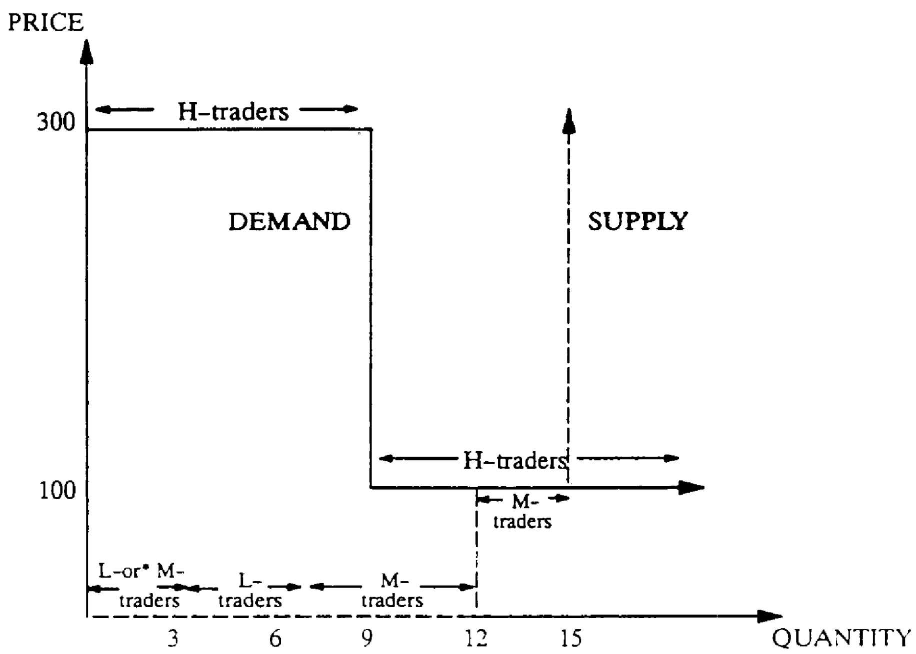


FIGURE 1 Supply and demand for a hypothetical one-period asset. Note: Assumes the parameters of experiments 8–10x with nine traders, with endowments equal to those predicted at the beginning of a period in equilibrium.
*The first three units are supplied by L-traders if they begin the period with six units, and by M-traders if they begin the period with fifteen units.

In a Walrasian competitive equilibrium, the price should be 100, since supply meets demand at that price and any lower price generates excess demand. Trading volume will range from 12 to 15 units (depending on whether M-traders sell marginal units to H-traders).[3] The double oral auctions we conducted are clearly not Walrasian. But prices in double oral auctions tend to converge to Walrasian competitive equilibria both empirically[23,36] and theoretically.[16,48]

The careful reader may wonder whether 100 is always the market-clearing dividend. It seems possible that units may be allocated at the beginning of a period, because of previous trading, in such a way that the demand and supply curves for that period cross at a price of 0 or 300 or cross vertically between 100 and 300, creating an indeterminate range of market-clearing dividends. However, 100 is always

[3]Note that trading volume will be lower in the first experimental period (six to nine units) because the H-traders begin with a total endowment of six units.

the market-clearing dividend. This can be shown by checking demand and supply curves for all possible pre-trade allocations of units, or by more formal arguments.[4]

DISCOUNTED DIVIDEND VALUE PREDICTIONS

Now recall that the asset actually lives from period to period with a probability .85. (In experiment 4m the probability was .8 instead; our analyses below adjust for this difference where appropriate.) In each period the market-clearing dividend is 100 (it is the same each period because aggregate demand is fixed). If subjects maximize expected utility and are risk-neutral, prices should be set by the equilibrium condition (formally derived in the Appendix)

$$P_t = 100 + .85E_t(P_{t+1}). \tag{1}$$

Intertemporal maximization in asset pricing theories typically yields an equilibrium condition like Eq. (1), with a discount rate $1/(1+r)$ instead of the probability .85. The expiration probability thus induces a probabilistic discounting which is formally equivalent (under risk neutrality) to discounting due to time preference.

Since probabilistic discounting is novel, a brief explanation of its merits is appropriate. Probabilistic and traditional discounting have a common theoretical basis in utility maximization. Probabilistic discounting reflects the ratio of expected utilities today and tomorrow; traditional discount rates reflect the ratio of marginal utilities of consumption today and tomorrow (perhaps because of discounting due to uncertainty, e.g., Fisher[13]). Expiration probabilities also have several practical advantages. They are easy to understand, require no additional accounting, and provide a natural way to end a session. Most importantly, Camerer and Weigelt[5]

[4]Consider the design we used most often, in session 8–13r with 12 traders and a fixed supply of 24 units (sessions 5–7 are almost identical, and the earlier sessions are similar, too). The fact that 100 is always the market-clearing dividend—in competitive equilibrium—hinges on two features we built into our design. First, there is always plenty of demand at a price of 100 because high-type traders can buy as many units as they can afford and collect a dividend of 100 on each. As a result, demand and supply can never cross at 0 because there is a fixed supply of units and excess demand at 100. Second, the curves cannot cross at 300 unless the number of unfulfilled units of demand at 300, n_u, is greater than or equal to the supply of units at a price of less than 300. But the number of units supplied at less than 300 is simply $24 - N_f$, where N_f is the number of units held by traders with demand at 300. Thus, a cross at 300 requires $N_u \geq 24 - N_f$, or $N_u + N_f \geq 24$. But, by design, $N_u + N_f = 16$ so the inequality cannot hold. Intuitively, there cannot be both a large enough *demand* for units at 300 (by traders who do not have them) and a small enough *supply* of units below 300 (because many units are held by traders with dividend values of 300).

Complications could arise if high-type traders are budget constrained, choking off excess demand for marginal units which pay them a dividend of 100, but budget constraints will only make the market-clearing dividend *lower* than 100, not higher. Furthermore, since traders accumulate francs throughout the experimental session from dividend earnings, their wealth grows and budget constraints become looser.

Finally, our proof assumes prices are determined by the intersection of demand and supply curves (in competitive equilibrium). A more complicated game-theoretic model might generate different predictions.

found that expiration probabilities induced discounting reasonably well in markets for two-period assets.[5]

Now we return to the discounted dividend value derivation. We cannot solve the pricing equation (Eq. (1)) without an assumption about how the expectation $E_t(P_{t+1})$ is formed. We suppose expectations are rational. If expectations are rational, then traders use Eq. (1) to calculate $E_t(P_{t+1})$, yielding

$$\begin{aligned} E_t(P_{t+1}) &= E_t[100 + .85E_{t+1}(P_{t+2})] \\ &= 100 + .85E_t(E_{t+1}(P_{t+2})). \end{aligned} \tag{2}$$

By the law of iterated expectations (e.g., Sargent,[34] p. 208), $E_t(E_{t+1}(P_{t+2})) = E_t(P_{t+2})$. Substituting into Eq. (2) yields a difference equation for expectations,

$$E_t(P_{t+1}) - .85E_t(P_{t+2}) = 100, \tag{3}$$

which has the solution(s) (e.g., Tirole[41,42])

$$\begin{aligned} E_t(P_{t+1}) &= \frac{100}{.15} + E_t(B_{t+1}); \\ E_t(B_{t+1}) &= \left(\frac{1}{.85}\right) B_t. \end{aligned} \tag{4}$$

Plugging the expectation $E_t(P_{t+1})$ from Eq. (4) into Eq. (1), calculating, and rearranging yields

$$P_t = 667 + B_t. \tag{5}$$

The first term (667) is the discounted dividend value (or "intrinsic value"). The second term is a "bubble," a component of the asset price which is unrelated to fundamentals (see Camerer[3]). If subjects in our experiments expect the arbitrary bubble term to grow at a rate of $1/.85 - 1$, about 18% per period, then the current bubble B_t equals the discounted (expected) present value of next period's bubble ($= .85(1/.85)B_t$), so the bubble component of prices satisfies the equilibrium condition (1).[6]

Rational bubbles of this sort are a theoretical possibility which have never been clearly observed outside the laboratory (see Garber[18,19]). In a companion paper,[6] we test for the presence of rational bubbles, using prices and subjects' forecasts of future prices. (We find some evidence of rational bubbles, but many dramatic prices increases and crashes which seem to be bubbles are actually volatile convergence paths which were not forecasted by subjects.)

[5] However, note that probabilistic discounting makes risk aversion a consideration. In our 1991 experiments, in designs which required relatively few traders to bid prices upward to competitive equilibrium, the fact that some traders were risk-averse inhibited convergence substantially. Thus, the success of probabilistic discounting is likely to depend subtly on design details.

[6] If expectations are adaptive, rather than rational, then prices converge to the discounted dividend level and rational bubbles cannot occur (cf. Lucas[26]).

DEVIATIONS FROM RISK-NEUTRAL DISCOUNTED DIVIDEND VALUE

The prediction that traders will price assets at the discounted dividend level relies on several assumptions. Violation of any of the assumptions could make prices deviate from the prediction. We review each important assumption in turn.

RISK NEUTRALITY. If subjects are risk-averse (risk-preferring), equilibrium prices will be below (above) the discounted dividend value 667 and rational bubbles will grow faster (slower) than 1/.85. We did not attempt to induce risk neutrality (as in Roth and Malouf,[33] or Berg et al.[2]) because: the methods for doing so may be unreliable[8,44]; the amounts at stake here are small (so risk neutrality should be a good approximation); and we can learn about convergence paths without knowing the subjects' degree of risk aversion.

EXPECTED UTILITY MAXIMIZATION. Even if subjects do not maximize expected utility (EU), as is assumed in deriving the equilibrium condition (1), EU-based predictions are likely to be good approximations. Most EU violations can be explained by subjects acting as if they use different utility functions when valuing different gambles.[27,28] In equilibrium, the asset is a gamble which is the same each period. Since completeness of preferences guarantees the utility function used to evaluate that gamble is unique, even if different utility functions are used to evaluate different gambles, the asset price will still have a unique equilibrium price (based on some utility function). In this sense, much evidence of EU violation is likely to be irrelevant here. Some other EU violations cannot be dismissed so easily. For instance, most choice theories assumes compound lotteries are equivalent to reduced single-stage lotteries. Empirically, people often overestimate the probability of a compound lottery which is a conjunction of events (such as an asset living eight periods), relative to its reduced probability ($.85^8$).[1] If our subjects make this error, they will overestimate the chance of receiving long dividend streams and overestimate the discounted dividend value.

MYOPIC EXPECTATIONS. Subjects may be extremely myopic, as if they think the probability of expiration is 1 or they simply don't realize the asset yields a probabilistic stream of dividends. Then prices will be closer to the marginal dividend value of 100 than to the discounted dividend value of 667.

NAIVE EXPECTATIONS. Since subjects know only their own dividend schedules, it will take time to develop rational expectations about market-clearing dividends (i.e., to learn that when they earn low dividends others earn high dividends). A plausible initial expectation is the naive belief that all subjects have the same dividend structure. With naive expectations, prices will be substantially lower than 667 (360 in sessions 1–3, 389 in session 4, and 259 in sessions 5–13).

TABLE 2 The Dividend Structure in Each Session

	High Dividend				Medium Dividend			Low Dividend
Session	1st	2nd	3rd	4th or more	1st	2nd	3rd or more	1st or more
1–3	300	100	100	100	*not applicable*			0
4m	300	100	100	100	50	50	50	0
5–7	200	200	200	100	200	100	0	0
8–13r	300	300	300	100	300	100	0	0

DESIGN DETAILS

Important design parameters are summarized in Table 1. Sessions are numbered in chronological order. Several design variables were changed across the sessions.[7]

First, subjects were either undergraduate and MBA students. Most of them were familiar with the discounted dividend model from finance classes; many of the MBA's worked full-time in Wall Street jobs.

Second, many subjects had not participated in previous market sessions, but some had. Subjects in two sessions had been in other double oral auction experiments; we call them "*m*echanism-experienced" subjects and denote the sessions 4*m* and 9*m*. Subjects in session 7*c* had been in session 5 or 6—we call them "*c*ontext-experienced"; those in session 10c had been in session 4m or 9m.

Third, we made the dividend structure common knowledge in sessions 11k–12k, by explaining the dividend rotation scheme without mentioning the identity of the subjects in each role. Common knowledge of heterogeneous dividends helps convergence to rational expectations equilibrium in multi-period asset markets[14]; we were curious whether common knowledge would help here too. Notice that without

[7]The design parameters in Table 1 show a variety of changes in each session. In an ideal experimental design, several treatment variables are chosen in advance combined in every possible way. Then repeated sessions are run for each combination of variables.

We began following this classical design model, only varying the number of subjects from ten to eight (which has made no difference in comparable experiments; see Smith,[36] p. 945). Then we ran two replications of the same design (sessions 2–3) on the same day. The results were surprisingly different. We concluded that variation between different sessions with the same parameters ("within-cell variance") was likely to be as large as the variation between sessions with different parameters ("between-cell variance"). So we replaced systematic parameter variation with a more exploratory design.

knowing the dividends of others, subjects must develop rational expectations of future prices by observation, rather than by calculation. Common knowledge enables them, in theory, to calculate the discounted dividend price.

Fourth, we changed the dividend rotation slightly in session 13r. Here's why: In sessions 1–12k, the dividend cycle was high–medium–low–high (HMLH). If trading is motivated mostly by dividend earnings, then one third of the subjects will be buying in the HMLH cycle (when they switch from L to H) and two thirds of the subjects will be selling (when they switch from H to M, and M to L). We thought this cycle might prevent prices from rising to the discounted dividend level, since in equilibrium twice as many subjects were selling each period as were buying. So in session 13r we changed the cycle to HLMH, causing twice as many subjects to buy (during L-M and M-H switches) as were selling (during H-L switches).

Before each period of sessions 4–13, we asked subjects to forecast the mean trading price in each of the next two trading periods.[8] Subjects earned $.10 for every forecast within 10 francs of the actual mean price. We used forecasts to test various theories of subjects' expectations (reported below).

Subjects were endowed at the beginning of the session with two units of the asset and several thousand "francs" of trading currency. Subjects could hold their assets or trade them for francs in a double oral auction. In sessions 1–4m, subjects were given more francs before each period.

After the end of each period a ball was drawn publicly from a bingo cage to determine whether the assets expired. If assets expired, subjects were re-endowed with two more assets if enough time remained in the session. (The re-endowment should not affect trading strategies since it is independent of previous behavior.)

Subjects earned the sum of capital gains (or losses) from trading and dividends from holding assets. Total franc earnings were converted to dollars at an exchange rate (see Table 1) and subjects were paid in cash. They earned about $30 each.

RESULTS

The top graphs in Figures 2–14 show a time series of mean prices (solid line) in each period, in each session. Table 3 shows mean prices and efficiencies. In the figures, the dotted horizontal line shows the discounted dividend value, 667. Vertical lines show periods when assets expired. The bottom of each figure shows the maximum number of assets held by any subject in each period. (These data are used later to explain convergence.)

Compared to other auction experiments, these data converge to equilibrium prices slowly and unreliably. There is wide variation in convergence paths. No simple generalization about convergence paths applies to all sessions. When subjects were

[8] Others have found that eliciting predictions about market behavior does not affect the equilibrium subjects reach.[24,37,47]

inexperienced, prices typically stayed near the marginal dividend of 100 for 5–15 periods, then moved toward 667 (sometimes overshooting)—but prices in session 5 (Figure 6) never moved. Subjects with experience in other auctions (Figures 5 and 10) converged more quickly. Context-experienced subjects (Figures 8 and 11) converged most quickly of all, to about 100 francs above the discounted dividend value. There was some convergence (prices ended closer to 667 than they began) in every session but one.

Common knowledge of dividends in sessions 11k–12k did not uniformly speed convergence (see Figures 12–13). In session 11k a single subject quickly bid 500, began buying units which he held through period 5, and prices converged within five periods. But in session 12k prices did not converge substantially quicker, or more smoothly, than in other sessions. We suspect common knowledge plays *some role*, perhaps interacting with trader rationality or experience, as in Forsythe and Lundholm.[14] But its visible effect in the data is weak enough to cast doubt on the presumption that traders will calculate equilibrium prices when they can and converge quickly to them.

Behavior in session 13r, with a different dividend cycle designed to speed convergence, was very peculiar (see Figure 14). Trades began immediately around 1200—twice the discounted dividend value—and prices only fluctuated by a few francs in the entire session. Since it is unlikely that the initial price difference is caused by the change in dividend design (which was *not* common knowledge), we are not sure how dividend rotation matters.

Ending prices were widely scattered around the discounted dividend value. But convergence to that value depends on several assumptions—competitive

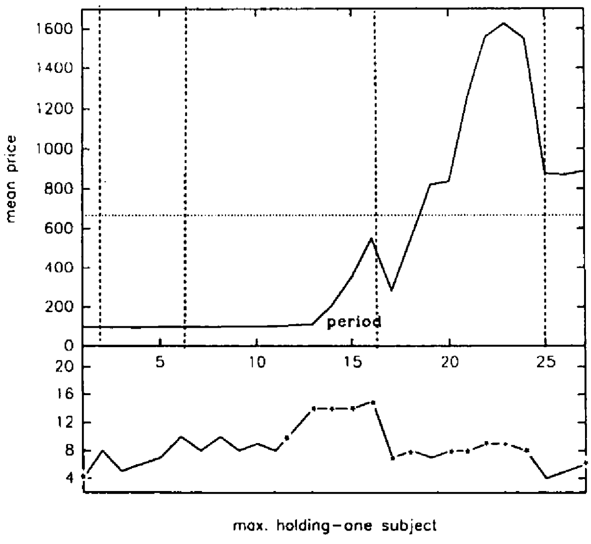


FIGURE 2 Time series of mean prices. Session 1: inexperienced subjects.

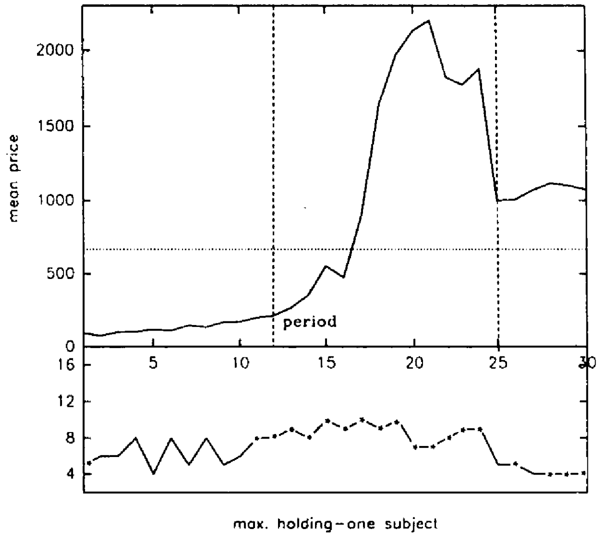


FIGURE 3 Time series of mean prices. Session 2: inexperienced subjects.

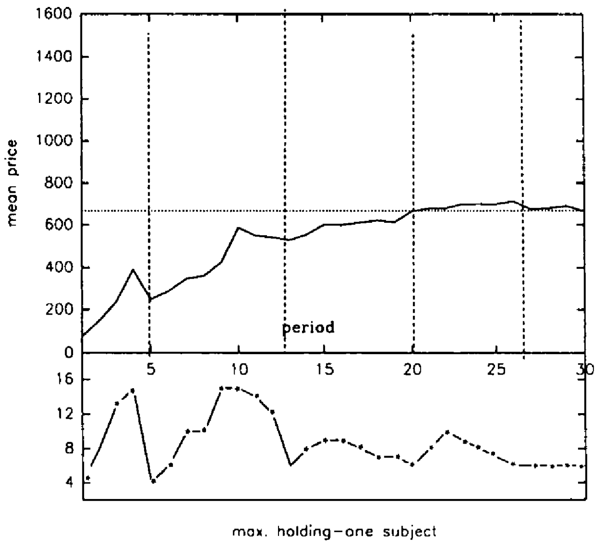


FIGURE 4 Time series of mean prices. Session 3: inexperienced subjects.

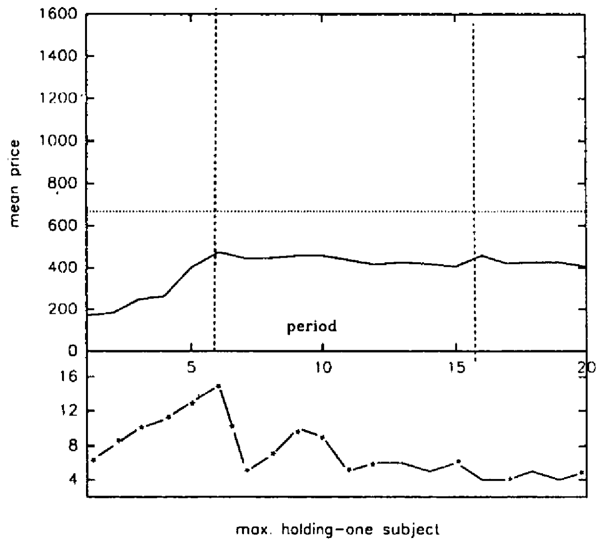


FIGURE 5 Time series of mean prices. Session 4m: experienced subjects.

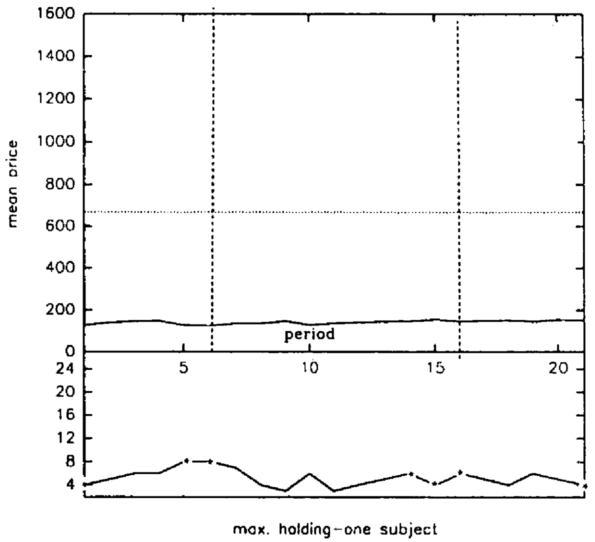


FIGURE 6 Time series of mean prices. Session 5: inexperienced subjects.

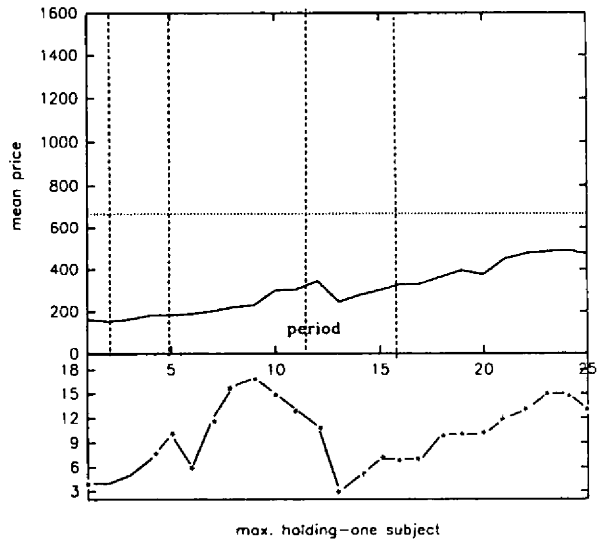


FIGURE 7 Time series of mean prices. Session 6: inexperienced subjects.

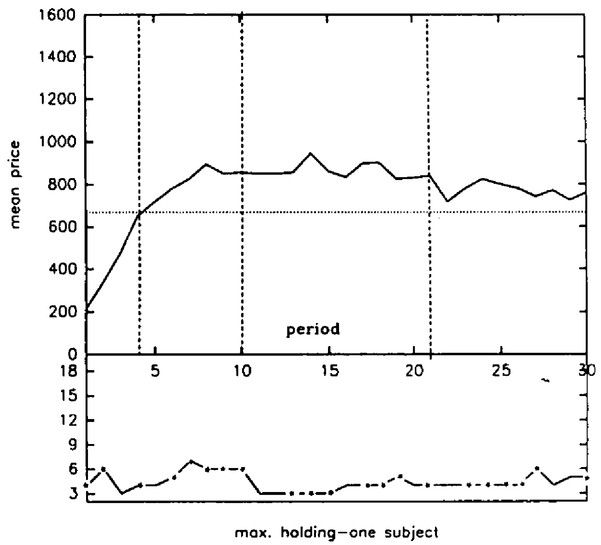


FIGURE 8 Time series of mean prices. Session 7c: content-experienced subjects.

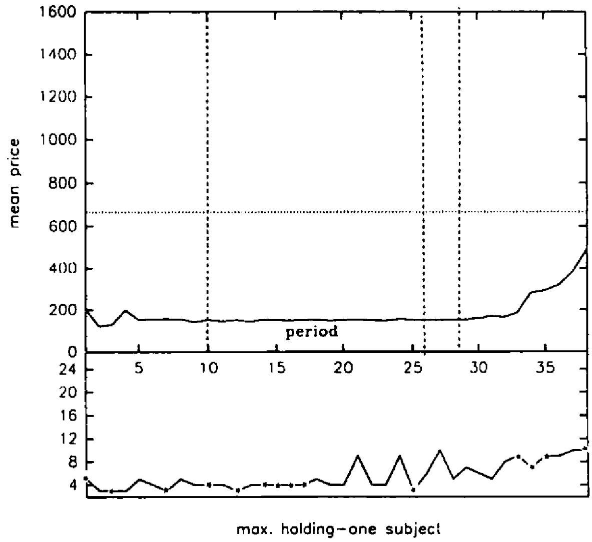


FIGURE 9 Time series of mean prices. Session 8: inexperienced subjects.

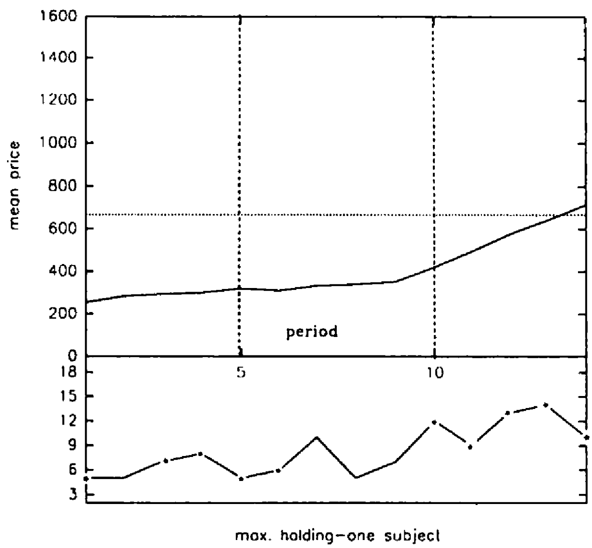


FIGURE 10 Time series of mean prices. Session 9m: mechanism-experienced subjects.

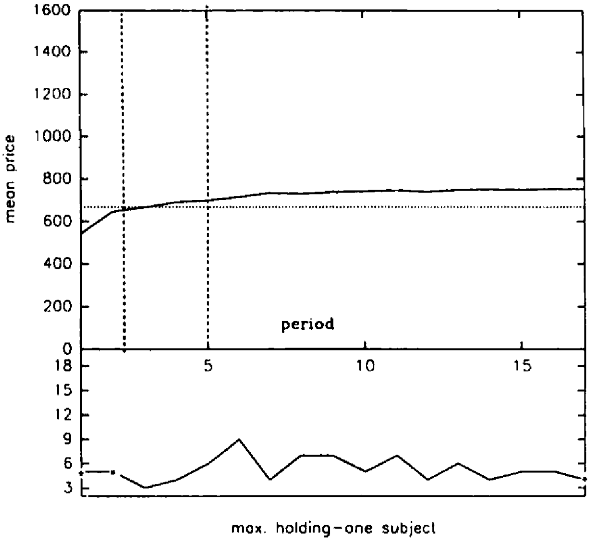


FIGURE 11 Time series of mean prices. Session 10c: content-experienced subjects.

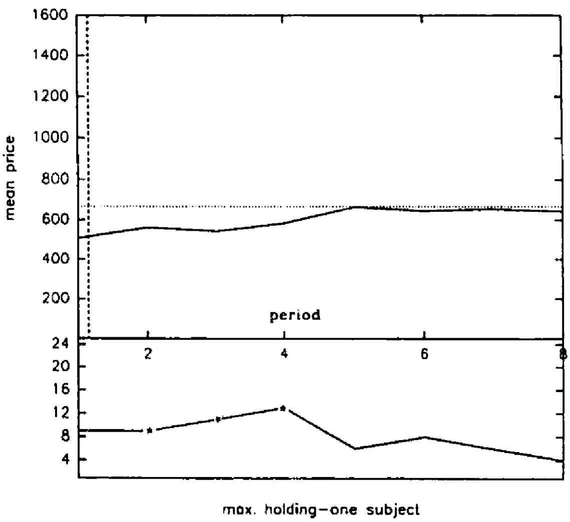


FIGURE 12 Time series of mean prices. Session 11k: inexperienced subjects.

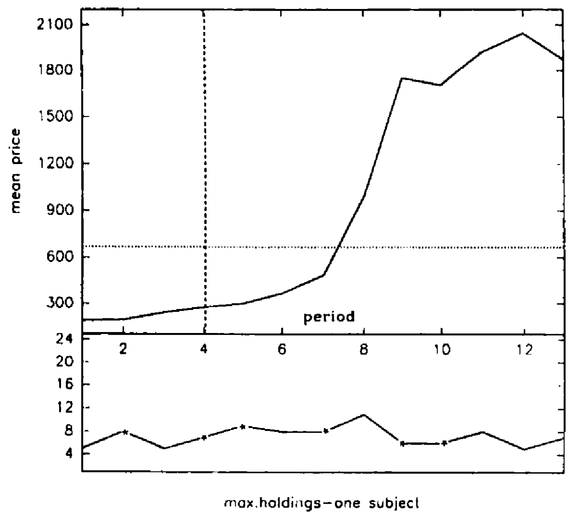


FIGURE 13 Time series of mean prices. Session 12k: inexperienced subjects.

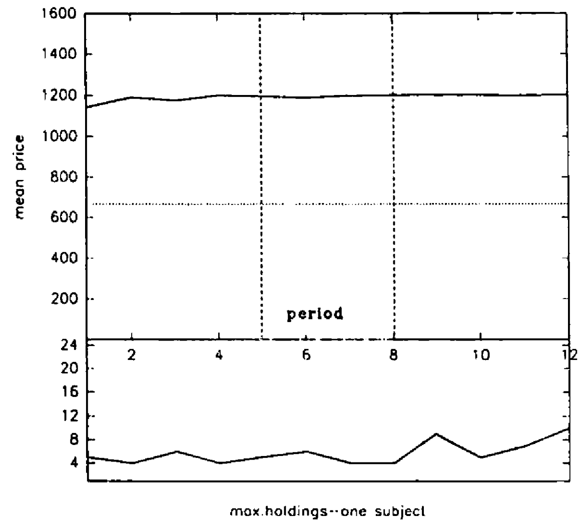


FIGURE 14 Time series of mean prices. Session 13r: inexperienced subjects.

TABLE 3 A. Efficiencies, Number of Trades, and Mean Prices[1]

Session	1	2	3	4	5	6	7	8	9	10
1										
efficiency-bs	.40	.80	.60	.60	.40	1.0	.90	.90	.80	.90
efficiency-bp		.88	.83	.78	.73	1.0	.97	.96	.93	.96
no. of trades	5	9[2]	15	12	12	14	10[2]	20	19	17
mean price	96	97	95	94	98	98	98	98	100	99
2										
efficiency-bs	.50	1.0	.75	1.0	.38	.50	.75	.38	.50	.75
efficiency-bp		.88	1.0	.92	.83	.84	.95	.86	.84	.88
no. of trades	7	17	17	15	14	13	14	16	11	12
mean price	91	74	97	99	115	107	143	131	165	168
3										
efficiency-bs	.63	.50	-1.1	.13	-1.1	-.25	-1.4	-.13	-1.6	.13
efficiency-bp		.79	-.06	.22	.57	.53	.50	.21	.22	.00
no. of trades	7	10	7	2	5[2]	7	6	3	5	0
mean price	77	146	236	390	248	286	346	360	425	587
4										
efficiency-bs	.33	-.50	.22	.83	-1.3	-.66	.11	-.55	.11	.66
efficiency-bp		.04	.42	.67	-.24	.06	-3.0	-.27	-.07	.40
no. of trades	4	3	4	5	2	2[2]	4	4	5	6
mean price	171	183	248	264	400	475	445	447	458	458
5										
efficiency-bs	.00	.17	.33	-.08	-.25	.25	.58	.58	.33	.75
efficiency-bp		.28	.29	.24	.34	.17	.33	.48	.37	.37
no. of trades	8	11	12	10	13	12[2]	9	15	12[2]	10
mean price	128	140	146	148	127	124	136	136	147	129
6										
efficiency-bs	.33	.00	.56	-.11	-.55	.77	-1.0	-1.7	.11	-1.6
efficiency-bp		.36	.86	.41	.03	.78	-.17	-.18	-.14	.15
no. of trades	4	6[2]	6	7	8[2]	6	7	4	1	2
mean price	163	153	163	182	183	189	202	221	230	300
7										
efficiency-bs	.00	-.11	.00	.00	.11	-.22	.11	-.77	-.22	.00
efficiency-bp		.20	.55	.10	.20	.29	.20	.20	.00	.18
no. of trades	6	2	7	6[2]	4	6	2	3	0	2[2]
mean price	215	338	477	650	718	780	825	893	850	855
8										
efficiency-bs	.75	.42	.42	.67	-.08	.42	.58	.67	.58	.67
efficiency-bp	.67	.53	.81	.38	.46	.75	.78	.72	.80	.79
no. of trades	5	7	8	7	6	7	6	7	10	8[2]
mean price	202	121	130	198	152	156	157	156	142	152

TABLE 3 A. Efficiencies, Number of Trades, and Mean Prices (continued)

11	12	13	14	15	16	17	18	19	20	21
-.20	1.0	-.80	.89	-1.7	.22	-.80	.30	-.50	.30	-1.0
.58	1.0	.41	.90	.08	.38	.33	.30	.21	.46	.13
11	10	8	4	6	3	10[2]	4	5	4	5
101	106	110	207	352	550	283	550	820	836	1252
-.50	.75	-1.3	.63	-1.1	.38	-1.1	.13	-1.5	.25	-.75
.50	.67	.40	.00	.00	.00	-.13	-.14	-.06	.00	.00
9	8[2]	2	2	2	2	3	1	2	3	1
196	209	262	350	553	475	900	1650	1975	2133	2200
-1.6	.13	.38	-.50	.38	-.63	.75	-.50	.13	-.75	-.38
.29	.19	.71	.08	.38	.32	.71	.33	.13	.43	.36
1	12	6[2]	6	5	3	5	4	2	5[2]	6
550	542	529	554	600	600	611	621	613	665	680
.00	.22	.39	-.22	.22	.22	-.17	.22	.22	.00	
.47	.44	.52	.27	.30	.26	.16	.42	.46	.38	
5	4	9	5	3	5[2]	4	3	8	6	
438	416	425	418	406	459	420	425	426	405	
.54	.58	.67	.42	.42	.67	.54	.25	.67	.58	.00
.46	.43	.40	.50	.37	.36	.52	.40	.30	.44	.28
15	10	11	13	13	12	14	13	9	13	12
136	141	145	147	155	146	149	151	145	154	152
-1.4	.56	-.11	.00	.33	-.55	-.55	-.11	-.66	-1.2	.00
.08	.29	.55	-.22	.33	.22	.07	.00	.12	.00	.18
3	3[2]	2	3	4	2[2]	6	4	1	0	2
305	345	245	277	301	328	331	364	395	375	450
.00	.11	.00	.11	.22	-.44	.22	-.44	-.33	.00	.11
.55	.00	.00	.00	.43	.18	.08	.30	.12	.17	.31
1	2	1	4	1	3	3	1	2	2	3[2]
850	850	855	944	860	833	897	900	825	830	838
.58	.67	.67	.25	.75	.75	.58	.83	.67	.67	.83
.73	.81	.63	.73	.86	.82	.88	.83	.83	.91	
5	8	7	8	7	9	9	9	8	8	5
146	151	144	153	151	148	151	153	148	151	154

TABLE 3 A. Efficiencies, Number of Trades, and Mean Prices (continued)

Session	1	2	3	4	5	6	7	8	9	10
9										
efficiency-bs	.58	-.25	-.41	.08	-.75	-1.1	.17	-.17	-1.3	.08
efficiency-bp		.38	.19	.50	.38	-.04	.33	.55	-.23	.15
no. of trades	6	10	5	7	7[2]	12	8	9	8	11[2]
mean price	254	283	293	300	320	310	333	340	352	420
10										
efficiency-bs	.58	-.25	.58	.58	-.16	.08	.50	.25	.92	.83
efficiency-bp		.38	.80	.72	.42	.56	.76	.57	.96	.93
no. of trades	8	7[2]	5	7	8[2]	10	9	7	13	12
mean price	543	645	666	690	698	715	733	729	737	741
11										
efficiency-bs	-.75	.75	.19	.06	.13	.50	.50	.44		
efficiency-bp		.84	.53	.23	.47	.77	.73	.68		
no. of trades	12	15[2]	13	13	12	14	15	15		
mean price	517	562	543	583	665	645	656	642		
12										
efficiency-bs	.50	.88	.56	.31	.38	.19	-1.2	.38	.00	-.31
efficiency-bp	.92	.83	.65	.63	.62	.12	.47	.30	.05	.19
no. of trades	9	15	14	16	15[2]	13	12	17	20	9
mean price	189	194	243	276	299	367	487	992	1775	1700
13										
efficiency-bs	.31	.31	.56	.75	.75	.38	.56	.63	.31	.81
efficiency-bp	.09	.63	.72	.58	.68	.66	.55	.66	.72	.66
no. of trades	11	21	14	17	14	9[2]	16	12	10[2]	14
mean price	1141	1190	1175	1200	1195	1189	1198	1200	1201	1201

[1] efficiency-bs = (Total dividends earned (TD) – dividends earned if traders kept endowed units(TDbs))/(Maximum possible dividends(MD) – TDbs).
efficiency-bp = (TD – dividends earned if traders kept units they owned at the beginning of the period(TDbp))/(MD – TDbp).
Minimum efficiency-bs: –2.33 (Experiment 5-7); –2.5 (Experiments 8-13r); –2.0 (Experiments 1-3); and –1.66 (Experiment 4).

[2] = new market—Note: For periods with no trades, the mean price is equal to the highest bid.

TABLE 3 A. Efficiencies, Number of Trades, and Mean Prices (continued)

11	12	13	14	15	16	17	18	19	20	21
-1.1	-1.8	-.25	-.42							
.11	-.03	-.36	.51							
15	9	3	7							
493	573	639	715							
.75	.58	.92	.83	.83	.83	.67				
.88	.82	.96	.93	.92	.92	.85				
11	11	10	10	9	10	9				
746	739	748	750	748	752	752				
-.38	-.69	-.44								
.18	-.10									
9	9	7	9							
1921	2045	1873								
.75	.13									
.52										
15	15									
1198	1201									

TABLE 3 B. Efficiencies, Number of Trades, and Mean Prices[1]

Session	22	23	24	25	26	27	28	29	30	31
1										
efficiency-bs	.50	-1.2	.40	.13	.13	-.50				
efficiency-bp	.50	.12	.25	.55	.36	.07				
no. of trades	4	2	2	5[2]	4	3				
mean price	1558	1625	1550	875	869	887				
2										
efficiency-bs	.25	-.75	.00	-.63	.88	-1.1	.38	-.13	.25	
efficiency-bp	.00	.00	.00	.09	.90	.38	.00	.23	.38	
no. of trades	2	3	0	5[2]	8	6	3	6	3	
mean price	1823	1775	1877	1000	1003	1070	1117	1100	1073	
3										
efficiency-bs	-.88	-.25	-.50	-.50	-.63	.38	-.25	.50	-.38	
efficiency-bp	.73	.20	.20	-.08	.22	-.14	.00	.33	.21	
no. of trades	4	4	3	4	2	5[2]	3	4	5	
mean price	680	699	700	699	713	675	680	690	665	
6										
efficiency-bs	-1.1	-1.9	-.11	-1.1						
efficiency-bp	.00	-.13	-.25	.17						
no. of trades	1	3	1	2						
mean price	475	484	490	473						
7										
efficiency-bs	.22	.22	-.11	.00	.22	.00	-.22	-.33	.11	
efficiency-bp	.42	.36	.38	.00	.42	.36	.21	-.09	.29	
no. of trades	3	2	3	0	3	5	4	3	2	
mean price	717	778	824	800	780	742	772	725	760	
8										
efficiency-bs	.75	.67	.83	.73	.73	.83	.83	.58	.92	.58
efficiency-bp	.90	.83	.90	.90	.87	.91	.93	.82	.95	.83
no. of trades	12	10	10	11	9[2]	10	12[2]	8	10	12
mean price	150	147	156	152	152	152	153	154	159	170

[1] efficiency-bs = (Total dividends earned (TD) – dividends earned if traders kept endowed units(TDbs))/(Maximum possible dividends(MD) – TDbs).
efficiency-bp = (TD – dividends earned if traders kept units they owned at the beginning of the period(TDbp))/(MD – TDbp).
Minimum efficiency-bs: –2.33 (Experiment 5-7); –2.5 (Experiments 8-13r); –2.0 (Experiments 1-3); and –1.66 (Experiment 4).

[2] = new market—Note: For periods with no trades, the mean price is equal to the highest bid.

TABLE 3 B. Efficiencies, Number of Trades, and Mean Prices (continued)

32	33	34	35	36	37	38	39	40
.67	.08	-.25	-.16	-.36	-2.1	-.45		
.81	.54	.50	.33	.29	.00	.53		
10	9	6	7	3	1	1		
165	186	282	294	320	385	490		

equilibrium, rational expectations, expected utility maximization, risk neutrality, and accurate probability judgment. Relaxing each assumption yields several competing theories, which were spelled out above. The data from inexperienced subjects reject the *joint* hypothesis that all the assumptions are true, but no single competing theory fits much better:

Risk aversion predicts equilibrium prices will be below the risk-neutral discounted dividend value, but prices were often above. If subjects violate expected utility or *misjudge expiration probabilities*, then prices should always be either too high or too low; but prices were not systematically high or low. If subjects believe expiration draws are correlated (as in "gambler's fallacy" or "hot hand" beliefs; see Camerer[4]), then subjective probabilities should change after asset expirations; but prices did not change substantially after expirations. If expectations about future prices are myopic (subjects act as if the probability of expiration is one), then prices should equal 100, but prices often start higher than 100. If their expectations are naive (they act as if others have the same dividends they do), prices should be between 250 and 300, depending on the design, but equilibrium prices were rarely at this level.

There are several empirical regularities in the data. We discuss three at length:

1. Initial prices are low for long periods of time.
2. Price changes are often driven by the behavior of speculators who buy much of the supply and hold it for a few periods.
3. Allocative efficiencies are low. Speculation reduces efficiency.

1. WHY INITIAL PRICES ARE LOW

Perhaps the major economic fact discovered by experimental economists is that convergence to competitive equilibrium is rapid and reliable in double auctions which are not explicitly Walrasian. Our data suggest a boundary of this fact: In double auctions for more complex asset markets, convergence can be slow and unreliable. Prices remained low, between 100 and 300, for 5–15 periods in most of the sessions with inexperienced subjects. In session 5 prices never converged at all. One puzzle is why prices in early periods stayed below the naive expectations prices that subjects should pay to buy and hold assets forever (259 in most sessions). Put differently, if we substituted the prices subjects forecasted into the equilibrium condition (1), the resulting discounted forecasted dividend stream is still higher than the actual price assets traded at.

In auctions for one-period goods, competitive equilibrium usually occurs within five trading periods (e.g., Plott,[30] p. 1493). Convergence in market experiments depends on the existence of two factors: an information-rich exchange institution (in which bids and offers, along with trade prices, signal private information about valuations) and a small number of infra-marginal traders. In most commodity experiments downward-sloping demand and upward-sloping supply curves imply that some (infra-marginal) traders simply cannot trade unless prices are near the equilibrium. Competition between these traders and extra-marginal traders forces prices into a narrow tunnel near the competitive equilibrium. As a result, convergence can occur in these markets even if traders do not know much about the values of others, and exhibit very little rationality in trading strategies (see Easley and Ledyard, and Gode and Sunder, both in this volume).

Compared to commodity markets, markets for multi-period assets lack information—because prices in future periods are unknown—and infra-marginal traders do not create a narrow tunnel—so convergence is slower (see Forsythe, Palfrey, and Plott[15]; cf. Friedman, Harrison, and Salmon[17]). In markets for 2-period assets, for instance, the entire 2-period life is repeated several times. Convergence occurs by experiential backward induction (sometimes called "swingback"): In the first repetition of the asset's entire 2-period life, prices are far too low in the first period, but prices are roughly correct in the second (and last) period. Observing prices in that second period, traders form an expectation of second-period prices which they use in the next repetition to determine how much to pay in the first period of asset's life. Observations of asset resale value in period 2 "swings back" into period 1 prices, which converge to the rational expectations level (see also Porter and Smith[32] for similar results in 15-period asset markets).[9] In our sessions, subjects cannot learn by swingback. Since the asset does not live a fixed number of periods, there is no final period for them to swing back from. The only hope for convergence is that traders are well informed and rational enough to calculate the discounted dividend value, then bid and offer near it.

In multi-period asset markets (e.g., Forsythe, Palfrey, Plott[15]), when spot markets and futures markets are open simultaneously, the futures market provides useful information about future prices and speeds convergence in the spot markets. There is no futures market in our experiments. However, if there was a futures market, it would not provide much information: Prices in a period t futures market for period $t+1$ assets should be the same as spot market prices in period t, because the stream of marginal dividends in period $t+1$ is the same as in period t (conditional on existence of the asset in $t+1$). Therefore, prices in the spot market serve a dual role—they equate current demand and supply and they serve the informational role of futures prices. A trader who realizes this can learn by "bootstrapping"—using current spot market prices to form expectations about future prices—instead of learning by swingback. Bootstrapping is more cognitively difficult than swingback because it requires traders to understand that the current-period price is also a forecast of the next-period price (this insight is not required by swingback learning).

In sum, quick convergence in one-period markets occurs when: the exchange institution provides lots of information (the double auction is perhaps the richest);

[9] People also seem to backward-induct in experiments on sophisticated voting[11] and sequential bargaining.[21] When they cannot observe future periods in sequential bargaining, perfect equilibria based on backward induction predict poorly.[40,29]

infra-marginal traders create a tunnel forcing prices toward the competitive equilibrium; and traders have some minimal degree of rationality. In multi-period asset markets, convergence is slower (unless there are futures markets) because traders lack information about prices in future periods. In our markets for stochastically lived assets, convergence is even slower because information about future period prices is not provided by repetition of the entire asset life which enables learning by swingback. The minimal rationality which is sufficient for convergence in simpler markets is insufficient in our markets.

2. SPECULATION AND CONVERGENCE

Since subjects in our asset markets did not seem to calculate the discounted dividend value and could not learn it by experiential backward induction (as in other asset experiments), how did prices converge? In many sessions, prices converged toward 667 when one entrepreneurial subject bought a large fraction of the supply of assets and held it for several periods (creating a kind of intertemporal market power; cf. Holt, Langan, and Villamil[22]). The behavior of these speculators is illustrated by the bottom graphs in Figures 2–14, which show the maximum number of units held by a single subject (i.e., the time series $\max_i N_{it}$). An asterisk indicates a period in which the subject holding the most units is the same as in the previous period (i.e., argmax $N_{it} =$ argmax N_{it-1}).

Looking through the figures, it appears that many movements of prices toward discounted dividend value are preceded by an increase in the maximum number of units held. In sessions 1–2, for example, prices rose after the entrepreneur began buying midway through the session. In session 3, which is identical in design to session 2, an entrepreneur began buying earlier than in session 2; prices rose earlier. In session 5 a buyer held nine units in periods 5–6 but prices did not change. Then the asset expired; after re-endowment, no buyer held many units and prices never rose. In session 8 several different buyers held many units in periods 21–27 (shown by the lack of asterisks), but prices did not increase. Then beginning in period 33, one buyer held around nine units for three periods and prices did increase.

Based on these observations, we conjecture that convergence in markets for stochastically lived assets occurs when one speculator buys several marginal units (which may earn her zero dividends) and holds them for several periods. Through her actions, a single highly rational speculator (who calculated the discounted dividend value or bootstrapped it) can produce convergence in the entire market. However, it is crucial that other traders respond to speculation by raising their bids.

We have two ideas about why other traders respond to speculation by bidding higher. One explanation is simple-minded: When a speculator buys several units and holds them, the supply curve shown in Figure 1 shifts inward. If the speculator holds a large enough supply permanently (or others think she will[10]), the market-clearing one-period dividend rises. Even if traders are myopic, prices should then increase (and the equilibrium price rises, too). Initial increases, from 100 to 300, may be enough to get convergence started.

Another explanation is more sophisticated. Suppose most subjects do not know the discounted dividend value but the speculator does (for reasons beyond our experimental control, like uncontrolled variation in the education of subjects or their rationality). Then the subjects play a complicated incomplete information game. Buying units and holding them signals that the speculator knows the price is too low. The signal is credible because it costs the speculator something (viz., expected losses due to the possibility of expiration). Once others observe the signal—offers begin to thin, indicating a shortage of supply—they raise their bids until the speculator sells her units. Prices rise only incrementally, rather than jumping immediately to the discounted dividend level, because bidders fear adverse selection (the speculator will only sell for the discounted dividend value or more). In further work, we hope to formalize this model (cf. Sobel and Takahashi[38]) and test it.

To check the speculator theory empirically, we created two measures. First define "the speculator" to be the one subject holding the most units at the end of a period; his or her holdings which pay marginal (or submarginal) dividends, divided by the total number of units in the session, forms the "speculation index." (Taking holdings by the *two* subjects holding the most units made little difference.) The idea that price movements are caused by supply restriction (due to speculation) can then be tested by regressing price changes against the number of unsatisfied high-dividend units, assuming the units held by the speculator in period $t-1$ are held off the market throughout period t. The theory that price movements are caused by information signaling can be tested by regressing price movements against the speculation index.

The difference in the supply-restriction and information-signaling approaches is subtle. Holding only a few units might signal that prices are low, without shifting the supply curve enough to create unsatisfied high-dividend demand. In fact, the measure of supply restriction was not a good predictor of price changes. Therefore, we only report results using the speculation index.

Denote mean and closing prices in period t by P_t and C_t; then price changes from period to period are measured by $P_t - C_{t-1}$. Denote the speculation measure by S_t. Then the speculation theory predicts a positive b in the regression of price changes against the speculation index,

$$P_t - C_{t-1} = a + bS_{t-1} + e_t \tag{6}$$

[10] This is not an entirely rational explanation. Since the purpose of entrepreneurial speculation is profit upon resale, the implicit threat to withhold supply permanently is not credible. But some incomplete information about true motives gives the entrepreneur some reputation-building incentive to withhold supply which makes the threat credible.

TABLE 4 Tests of Price Formation Models[1]

		(a) $P_t - C_{t-1} = a + b(B_{t-1} - O_{t-1}) + u_t$					(b) $P_t - C_{t-1} = a + bS_{t-1} + u_t$				
Exp.	Sample Size	a (t-stat)	b (t-stat)	Adj. R^2	F-stat	D-W	a (t-stat)	b (t-stat)	Adj. R^2	F-Stat	D-W
Inexperienced Subjects											
1	26	−10.75	.38	−.02	.43	2.31	1.70	−11.16	−.03	.09	2.26
		−.65	.60				.06	−.15			
2	29	13.67	.80	.00	1.17	2.07	−51.93	153.55	.07	3.09*	1.64
		−1.07	.92				−1.75	1.66*			
3	29	31.16	−4.25	.00	1.08	2.58	−56.14	295.86	.01	1.43	2.75
		1.18	1.18				−1.16	1.32*			
5	20	−1.90	−1.77	.46	17.51***	1.70	−13.64	98.88	.13	4.01*	2.26
		−.63	−4.17				−2.13	1.98**			
6	24	−12.75	1.47	.03	1.54	2.57	−2.02	11.39	−.06	−.14	2.53
		1.05	1.40*				−.20	.50			
8	37	−1.38	−.03	−.02	.12	1.56	−8.24	52.31	.02	1.99	1.75
		−.32	−.04				−1.31	1.45*			
11k	7	13.53	.18	−.19	.02	2.45	−40.11	179.04	.28	3.33	2.89
		.50	.16				−1.20	1.84*			
12k	12	48.05	−.26	.33	2.80	.90	−152.80	840.92	−.23	−1.10	1.57
		1.07	−.16				−1.20	1.55*			
13r	11	6.92	.36	−.02	.79	1.81	10.87	−40.09	−.04	.14	2.46
		1.23	.90				.72	−.39			
Mechanism Experienced											
4m	19	12.45	1.12	.12	3.53*	1.67	6.15	33.45	.07	2.41	1.72
		1.29	1.19				.29	.65			
9m	13	6.62	1.61*	.25	2.60	.84	−4.54	92.08	−.20	−1.04	1.23
		.48	1.36*				−.31	2.06**			

Context Experienced											
7c	29	13.84	4.10	.37	17.03***	1.94	48.70	−125.36	.12	4.76**	1.10
		1.09	3.38***				2.30	−.85			
10c	16	17.39	2.02	.60	23.95***	1.94	30.36	−118.99	.14	3.54*	1.19
		3.89	4.74***				2.71	−1.69			
Pooled Results											
5,8,9m,	97	−.83	−.49	.00	.73	1.77					
10c,13r		−.35	−1.35								
2,5,6,8,	150						−8.67	34.38	.03	5.02**	2.08
9m,10c,13r							−2.21	2.32**			

[1] Both regressions are weighted by the reciprocal of the average price for the period. Asterisks indicate the level of the t-statistic's p-value in a one-tailed test: * < .10; ** < .05; and *** < .01.

where e_t is an error term. Intuitively, when the speculation measure is high in period 7, the change in price between the close of period 7 and the average of period 8 should be high.

A competing benchmark theory is the Smith-Suchanek-Williams[37] hypothesis that excess bidding—the number of bids B_t minus the number of offers O_t—predicts price changes. Note that the excess bidding and speculation theories overlap substantially, because speculation will typically dry up supply and shrink O_t, raising S_t and lowering $B_t - O_t$ at the same time.

Excess bidding predicts a positive b in the regression

$$P_t - C_{t-1} = a + b(B_{t-1} - O_{t-1}) + e_t. \tag{7}$$

Generalized least-squares estimates of Eqs. (6) and (7) are summarized in Table 4.[11] We report separate estimates for each session and estimates for sessions that could be pooled together (according to Bartlett's test for homogeneity of variance and the Chow test for equality of coefficients).

Both theories are modestly successful, but neither predicts price changes extremely well. Excess bidding (see Table 4(a)) is significant with the correct sign ($b > 0$) in four sessions (6, 9m, 7c and 10c) and is insignificant when five sessions are pooled.[12] The coefficient in the speculation regression (Table 4(b)) has the correct sign ($b > 0$) in nine out of thirteen sessions. Coefficient estimates are weakly significant (at $p < .10$, by a one-tailed test) in seven of those sessions, and more significant ($p < .05$) when seven sessions are pooled. Either speculation effects are inherently weak (or variable across sessions), or we have not found the best econometric measurement of those effects, despite trying several methods.

3. VOLUME AND EFFICIENCY

As in Plott and Smith,[31] we can define allocative efficiency in each period by calculating the total amount of dividends earned by all traders (TD). It is helpful to normalize TD by comparing it with the maximum amount of dividends that could be earned (MD), and with various benchmarks. One benchmark, TDbs, is the amount of dividends traders would earn if they kept the units they were endowed with at the *b*eginning of the *s*ession. Another benchmark, TDbp, is the amount of dividends traders would earn if they kept the units they owned at the *b*eginning of

[11] We ran generalized least squares (GLS) to correct for apparent heteroskedasticity in the errors which reduced power. Estimates using White's[46] heteroskedasticity-consistent variance-covariance matrix improved power slightly, but his method is only appropriate when the error variance depends only on the independent variable—in this case, the number of excess bids or the speculation index—which has no appealing interpretation. Since error variance appeared to increase with increases in price level, we used GLS in which all observations were divided by P_t. The differences between GLS and OLS estimates were not large or systematic.

[12] Excess bidding did not predict well in Smith et al.'s[37] experiments when prices were stable or continually increasing (their Table IV) or in Daniels and Plott's[9] inflation experiments, so it may not be surprising that it predicts poorly here.

the period. Using these figures, we construct two ratios, (TD-TDbs)/(MD-TDbs) and (TD-TDbp)/(MD-TDbp). The two ratios give two different measures of the amount of extra dividends that were gained, as a fraction of what could have been gained. (An alternative measure subtracts the *minimum* total dividends instead of the maximum dividends MD. This measure is an affine transformation of ours which always lies between zero and one.)

If there is no trading in a period the ratio (TD − TDbp)/(MD − TDbp) will be 0; if trading is perfectly efficient, the ratio will be 1. The ratio can be negative (and often is) if trading causes subjects to collect fewer total dividends than if they had not traded at all.

Table 3 reports both kinds of efficiencies in each period, along with trading volume and mean prices. The third column of Table 5 shows average efficiencies using the beginning-of-session measure, (TD−TDbs)/(MD−TDbs). The efficiencies are low (less than 50%). Many are negative, and most are much lower than those reported in experiments with goods or one-period assets (e.g., Smith[36]). In other experiments with multi-period assets, allocation of assets is usually efficient even when prices are far from equilibrium.[15] In our sessions, capital-gains expectations or knowledge of the asset's discounted dividend value, rather than current-period dividends, seem to be important determinants of who buys and who sells. Efficiency suffers as a result.

A glance through Table 3 suggests that efficiency falls as prices rise, and when speculators buy units and hold them. To look for determinants of efficiency more carefully, we regressed efficiencies in each period on three variables: The index of speculation by entrepreneurs (S_{t-1}), trading volume, and the gap between discounted dividend value and prices. We also included a lagged efficiency term to pick up any "seasonality" due to the 2- or 3-period dividend cycles.

Results are reported in Table 5. The lagged efficiency term explains around a third of the variance in efficiency, suggesting that some subjects always buy when their dividend level is high and others do not. The other three variables add a little explanatory power to lagged efficiency, but not much. (The hypothesis that the three variables add nothing is rejected by an F-test, shown in the rightmost column, in nine sessions and in eight sessions that could be pooled.) Of the three independent variables in Table 6 (other than lagged efficiency), the speculation index has the strongest effect. Speculation reduces efficiency in *every* session, significantly so in six out of thirteen sessions, and its effect is highly significant ($t = 7.59$) in eight pooled sessions. The number of trades improves efficiency when results are pooled ($t = 3.67$), but the deviation of prices from the equilibrium prediction ($667 - P_t$) has no discernible effect.

The negative effect of speculation on efficiency points out an important irony. We think speculation is important to the process of forcing prices to the discounted

TABLE 5 Tests of Efficiency Models[1]

			$\text{Eff}_t = a + b(\text{Eff}_{t-n}) + u_t$				
Exp.	Sample Size	Mean Efficiency	a (t-stat)	b (t-stat)	Adj. R^2	F-Stat	D-W
Inexperienced Subjects							
1	25	.17	-.02	.77	.57	32.98***	2.09
			-.17	5.74***			
2	28	.05	-.04	.79	.67	54.23***	2.50
			-.47	7.36***			
3	28	-.31	-.19	.54	.28	11.74***	2.06
			-1.70	2.06			
5	18	.38	.25	.43	.13	3.59*	1.20
			2.37	1.89*			
6	22	-.44	-.32	.62	.33	11.34***	1.30
			-2.17	3.37***			
8	35	.45	-.33	1.36	.44	27.53***	1.47
			-2.02*	5.25***			
11k	5	.23	.32	.10	-.25	.19	1.10
			2.94	.43			
12k	10	.10	.33	.33	.02	1.20	2.28
			1.35	1.10			
13r	9	.52	.40	.32	-.07	.45	1.17
			1.56	.67			
Mechanism Experienced							
4m	17	.02	.02	.49	.21	5.25**	1.32
				.19	2.29**		
9m	11	-.47	-.32	.74	.41	8.02**	1.58
			-1.75	2.83**			
Context Experienced							
7c	27	-.05	-.05	-.12	-.03	.35	2.11
			-1.09	-.59			
10c	14	.55	.37	.46	.25	5.30**	1.09
				2.92	2.30**		
Pooled Results							
2,4,5,8,	126	.31	.06	.76	.51	129.44***	1.50
10c,11k,13r			1.47	11.38***			

[1] Lags are $n = 2$ for Exp. 1–3 and $n = 3$ for Exp. 4–13. Asterisks indicate the level of the t-statistic's p-value in a one-tailed test: * < .10; ** < .05; and *** < .01.

TABLE 6 Tests of Efficiency Models[1]

$$\text{Eff}_t = a + b(\text{Eff}_{t-n}) + c(\text{Spec}_t) + d(\text{Trades}_t) + e(667 - P_t) + u_t$$

Exp.	Sample Size	a (t-stat)	b (t-stat)	c (t-stat)	d (t-stat)	e (t-stat)	Adj. R^2	F-Stat	D-W
Inexperienced Subjects									
1	25	.52	.73	-1.59	.01	.00	.65	12.11***	2.41
		1.30	5.46***	-2.37**	.42	.42			
2	28	.19	.69	-1.31	.03	.00	.72	18.31***	2.76
		.53	5.63**	-1.79**	1.55	-1.02			
3	28	-.08	.52	-.21	.02	-.00	.32	4.11**	2.38
		-.19	3.37***	-.15	.33	1.94			
5	18	-3.62	.50	-1.13	-.00	.01	.01	1.04	1.49
		-.74	1.93*	-.96	-.02	.79			
6	22	.86	.57	-2.10	-.07	.00	.72	14.30***	1.97
		1.99	4.41***	-5.33***	-1.26	.25			
8	35	-1.14	.52	-1.07	.04	.00	.72	21.90***	2.71
		-2.16	1.94*	-2.33**	1.24	2.04*			
11k[2]	5								
12k	10	-.33	.17	-2.43	.08	.00	.03	1.06	3.10
		-.29	.40	-.81	1.53	.44			
13r	9	-9.26	-.04	-3.49	.03	-.02	.70	5.68*	2.89
		-1.19	-.15	-3.94***	1.25	-1.26			
Mechanism Experienced									
4m	17	-.94	.32	-.55	.10	.00	.40	3.65**	1.76
		-1.46	1.65	-.96	1.62	1.32			
9m	11	1.55	.79	-2.11	-.05	-.00	.50	3.44*	1.96
		1.67	2.98**	-1.66*	-1.10	-1.27			
Context Experienced									
7c	27	.27	-.06	-1.13	-.02	.00	.15	2.21	2.62
		1.46	-.31	-2.64***	-.65	1.13			
10c	14	-.03	.23	-.28	-.03	.01	.42	3.38*	1.41
		-.10	1.14	-.34	-.57	-2.22*			
Pooled Results									
2,4,5,6,7c,	170	.14	.55	-1.14	.03	.00	.69	96.93***	1.92
8,10c,13r		1.92	9.48***	-7.59***	3.67***	-.43			

[1] Lags are $n = 2$ for Exp. 1–3 and $n = 3$ for Exp. 4–13. Asterisks indicate the level of the t-statistic's p-value in a one-tailed test: * < .10; ** < .05; and *** < .01.

[2] Omitted due to insufficient sample size.

dividend level (as the Table 4b estimates suggest). But speculation reduces the index of allocative efficiency, by 4–5% for each unit held for speculation.[13] So the long-run benefits of speculation in establishing accurate prices come at a short-run expense.

This kind of tradeoff between short-run allocative efficiency and long-run pricing accuracy may be common in many kinds of economic development. Inhibiting speculation by requiring owners of assets to derive utility from them—forcing real estate speculators to rent houses they own, or live in the houses themselves, for instance—would improve short-run efficiency but harm long-run convergence.

OTHER OBSERVATIONS

We make three other observations.

First, in our companion paper on bubbles,[6] we study the rationality of the forecasts subjects made in sessions 4m–13. Rational forecasts should be uncorrelated with observable information, including previous forecast errors $P_{t-1} - F_{t-1}$ and forecast levels F_t. (Otherwise the observable information could be used to reduce errors.) In fact, subjects' forecast errors were not rational in this sense. Errors were typically negatively correlated with forecasts—when forecasts were low, they were too low; when forecasts were high, they were too high—and positively autocorrelated (around $r = .3$).

We also tested whether forecasts were rational or adaptive (i.e., $F_t - F_{t-1} = a + b(P_{t-1} - F_{t-1}] + e_t$). They were generally adaptive, with an adaptation coefficient of $b = .8$ or so. Our results corroborate most other studies of forecasts in experimental markets. (For example, previous estimates of b include .62,[7] .86,[47] .82,[37] and .70[45]; cf. Daniels and Plott[9]).

Second, when prices overshot the discounted dividend value, they were very slow to converge downward. The downward stickiness may be related to "disposition effects"—people who paid P for assets are reluctant to sell for less than P.[12,35] Disposition effects appear to be a manifestation of aversion to recognizing losses.[43]

Third, our data are different than those of Smith, Suchanek, and Williams.[37] In their 15-period markets, when traders are inexperienced prices typically start too low, then quickly converge and overshoot the discounted dividend value in speculative bubbles. There are at least two important differences between their results and ours. First, all their subjects earned the same dividend (effectively making each trader an infra-marginal trader). Second, their subjects were informed of the sequence of (expected) discounted dividend values in a chart handed out before the experiment. The gap between traders' dividends in our experiments, and their ignorance about the discounted dividend value, probably accounts for the slow convergence from below that we saw.

[13] Adding one unit to the speculator's holdings increases the speculation index by 1/18 or 1/24 (depending on the design), because the index is normalized by the total number of units available. Since the coefficient of the speculation index on efficiency is around -1, this implies that adding a unit to holdings reduces the level of efficiency by 1/24 to 1/18, or 4.2–5.6%.

CONCLUSION

Subjects in our sessions traded an asset that pays a dividend each period and lived from period to period with a known probability. This probability was formally equivalent to a discount rate in an equilibrium condition relating current prices to expectations of future prices, so we could test whether prices converged to the discounted dividend value. Probabilistic discounting seemed to work well as a method for mimicking infinitely lived assets and inducing discount rates.

Without "swingback" learning, as in earlier experimental double auctions for multi-period assets, convergence to discounted dividend value was very slow. Convergence seemed to be driven by the actions of a single rational speculator, who bought and held much of the market supply for several periods and forced the price up. We think that prices rose either because the supply restriction raised the equilibrium price, or because it signaled the speculator's superior information about the asset's true discounted dividend value to others and caused others to bid more for the assets.

The kind of speculation we observed is roughly analogous to episodes in natural markets in which prices rise because better-informed speculators recognize undervaluation of assets, buy the assets and hold them, and sell them after others bid the assets up. Commercial real estate speculation in borderline neighborhoods and the discovery of creative talent (e.g., writers and musicians) are examples. Tender offers for shares of undervalued firms are another example: even if the offer is unsuccessful, shares tend to trade at the tender offer price rather than at the lower pre-tender price, which suggests that the tender offer permanently signals to the market that the shares were undervalued.

Despite slow convergence, in almost all our sessions inexperienced subjects move toward the risk-neutral discounted dividend value. Experienced subjects trade at prices close to it. This result is remarkable because there are many ways in which the risk-neutral discounted dividend value prediction could go wrong (risk aversion, violations of expected utility, subject doubts about asset life, etc.).

The double auction does not fail completely at generating convergence to competitive equilibrium prices for stochastically lived assets—convergence occurs with experienced traders—but the auction mechanism performs much more slowly and erratically than in simpler settings. Furthermore, the process of convergence in this setting is fundamentally different than processes others have documented, which require little trader rationality. Since the information-rich double auction does not answer the most basic question traders have—how much will assets be worth next period?—the actions of speculators who are willing to bet that they know the answer become crucial in generating convergence.

ACKNOWLEDGMENTS

This research was supported by the Richard Paget Research Chair at Northwestern University and the Strategic Research Management Fund at New York University. We thank several anonymous referees, Andy Daughety, Dave Grether, Ron King, Marc Knez, Tom Palfrey, Charles Plott, Vernon Smith, Martin Weber, Arlington Williams, seminar participants at Northwestern, NYU, Caltech, Penn, and the Public Choice Society/Economic Science Association meetings, the Santa Fe Double Auction Markets conference, and especially Jack Farley, Peter Knez, and Marcy Lynch, for comments and help.

APPENDIX: DERIVATION OF EQUILIBRIUM CONDITION (1)

Subjects are endowed with wealth W_0, and an initial supply of two assets, denoted $S_0 = 2$. Since francs earned in the experiment are converted directly into dollars at a known rate and cannot be consumed until then, the subject's problem is to maximize expected utility of terminal wealth. Suppose for simplicity that the experiment ends after the asset expires the first time. (We relax this assumption below.)

Denote the wealth of subject i after period t by W_{it}. Denote subject i's holdings of assets at the end of period t by S_{it}, and denote prices in period t by P_t. Call the dividend of the jth unit of the S_{it} units, D_{itj}. Taking prices as given (as in a Walrasian competitive equilibrium), subject i's problem is to pick S_{it} to maximize expected utility of terminal wealth. The problem is:

$$\max_{S_{it}} pU(W_{i1}) + (1-p)pU(W_{i2}) \ldots (1-p)^{n-1}pU(W_{in}) + \ldots . \qquad (A.1)$$

The budget constraints are

$$W_{i1} = W_{i0} + \sum_{j=1}^{S_{i1}} D_{i1j} + (S_{i0} - S_{i1})P_1; \qquad (A.2)$$

$$W_{it+1} = W_{it} + \sum_{j=1}^{S_{it}} D_{itj} + (S_{it} - S_{it+1})P_{t+1}, \qquad t \geq 2. \qquad (A.3)$$

Now suppose the subject is risk-neutral (e.g., $U(W) = W$), an assumption we relax below. Then substituting Eqs. (A.2) and (A.3) into Eq. (A.1), the subject's problem reduces to

$$max_{S_{it}} \sum_{j=1}^{\infty} p(1-p)^j W_{i0} + \sum_{m=1}^{\infty}\sum_{j=1}^{S_{im}} D_{imj}p(1-p)^m + \sum_{m=1}^{\infty}(S_{im-1} - S_{im})P_m p(1-p)^m . \qquad (A.4)$$

Suppose that $D_{ikj} = D$ (an assumption which is also relaxed below.) Then Eq. (A.4) is easily differentiated with respect to S_{it}, for $t = 1, 2, \ldots$, to yield

$$\frac{\partial(\text{A.4})}{\partial S_{i1}} = \sum_{k=0}^{\infty} (D - P_1 + (1-p)P_2)(1-p)^k = 0, \qquad (A.5)$$

$$\frac{\partial(\text{A.4})}{\partial S_{in}} = \sum_{k=n-1}^{\infty} (D - P_n + (1-p)P_{n+1})(1-p)^k = 0, \qquad (A.6)$$

which simplify to

$$D - P_1 + (1-p)P_2 = 0, \qquad (A.7)$$

$$D - P_n + (1-p)P_{n+1} = 0. \qquad (A.8)$$

So far, we have suppressed uncertainty about prices. In a Walrasian world, current prices are known when trades are executed but next period's prices may not be. Assuming risk neutrality, we can substitute $E(P_n)$ for P_n in Eq. (A.8) to get equilibrium condition (1) in the text. (Alternatively, if there is no uncertainty prices will obey Eq. (A.8) rather than Eq. (1) and deterministic bubbles can occur.) Q.E.D.

We made three simplifying assumptions in deriving Eq. (A.8). We now relax these:

i. *Risk Tastes*: If $U(X)$ is not linear in X, then the summations in Eqs. (A.5) and (A.6) include marginal utility terms. They take the form

$$\sum_{k=0}^{\infty} p(1-p)^k U'(W_{ik+1}).$$

If $U(\cdot)$ reflects risk aversion (seeking), $U'(W_{ik+1})$ is declining (increasing) in W_{k+1}, i.e., $U''(\cdot)$ is negative (positive). It is easy to show that the prices implied by Eq. (A.6) under risk aversion (seeking) are *lower* (*higher*) than D/p. (For instance, express the marginal utilities $U'(W_{ik})$ with $k > 2$ as ratios, less than one, of $U'(W_{i1})$. Then the convergent sums over $p(1-p)^k$ in Eqs. (A.5)–(A.6) are still convergent, but the sums from $k = 0$ will be proportionally larger than the sum from $k = 1$. Since the $(1-p)$ terms in Eqs. (A.7)–(A.8) represent the ratio of the sum from $k = 1$ to the sum from $k = 0$, under risk aversion (seeking) this ratio will be smaller (larger) than $1-p$, which directly implies equilibrium prices below (above) D/p.)

ii. *Differing Dividends*: In our experiments, the assumption $D_{ik+1j} = D$ is not satisfied. Typically, for $i = 1, 2, 3$, $D_{ik+1j} = 300$ for $j = 1, 2, 3$ and $D_{ik+1j} = 100$ for $j \geq 4$; $i = 4, 5, 6$ have $D_{ik+11} = 300$ and $D_{ik+12} = 100$; and $i = 7, 8, 9$ have $D_{ik+1j} = 0$ for all j. Furthermore, these dividend schedules rotate systematically (in period $k+1$ the ith subject inherits the $\max(i-6, i+3)$ subject's period k schedule). It is straightforward but tedious to show that Eq. (A.8) determines the competitive equilibrium. Subjects will have one of three optimal levels of S_{it} (depending on their dividend schedules) so the additional aggregation constraint $S_{1t} + S_{2t} + \ldots + S_{nt} = 2_n$ is needed to establish a version of Eq. (A.8) with the marginal ($(2n)$th highest) dividend substituting for D. The price (marginal) D/p still clears the market (absent bubbles) because there is excess demand (supply) at any lower (higher) price.

iii. *Single Expiration*: We assumed above that the experiment would end after the asset expired the first time. Instead, suppose there is a chance d that the experiment will end after any given expiration ($d = 1$ in the derivation above). If the experiment does not end after period t, subjects are re-endowed with assets ($S_{it+1} = 2$) and their wealth continues to accumulate. Therefore, from period $t+1$ on, their maximization problem is exactly the same as in period 1 (ignoring changes in marginal utility of wealth). Denote the solution to the maximization problem beginning in period $t+1$ by M_{t+1}. Then their period 1 maximization problem (with $D_{ik+1j} = D$, suppressing i) is

$$\max_{S_t} pdW_1 + p(1-d)M_2 + (1-p)pdW_2 + (1-p)p(1-d)M_3 + (1-p)^2 pdW_3 + (1-p)^2 p(1-d)M_4 + \dots \tag{A.9}$$

We know that Eq. (A.8) is the first-order condition for the maxima M_i (by definition of maxima). If Eq. (A.8) is also the first-order condition for maximizing the non-maxima terms in Eq. (A.9), then it is the first-order condition for Eq. (A.9) in total. But the non-maxima terms are simply the terms in the original maximization problem (A.1) (assuming $U(W) = W$) times d. Since the common multiple d will drop out in the first-order condition, Eq. (A.8) maximizes the non-maxima terms in Eq. (A.9) so it maximizes all of Eq. (A.9). Therefore, Eq. (A.8) is the equilibrium condition even when the end of the experiment is stochastic.

REFERENCES

1. Bar-Hillel, Maya. "On the Subjective Probability of Compound Events." *Org. Behav. & Human Perf.* **9** (1973): 396–406.
2. Berg, Joyce E., Lane A. Daley, John W. Dickhaut, and John R. O'Brien. "Controlling Preferences for Lotteries on Units of Experimental Exchange." *Quart. J. Econ.* **101** (1986): 281–306.
3. Camerer, Colin F. "Bubbles and Fads in Asset Prices." *J. Econ. Surveys* **3** (1989a): 3–41.
4. Camerer, Colin F. "Does the Basketball Market Believe in the 'Hot Hand'?" *Am. Econ. Rev.* **79** (1989b): 1257–1261.
5. Camerer, Colin F., and Keith Weigelt. "A Test of a Probabilistic Mechanism for Inducing Stochastic Horizons in Experiments." In *Research in Experimental Economics*, Volume 6, edited by R. M. Isaac. Greenwich, CT: JAI Press, in press.
6. Camerer, Colin F., and Keith Weigelt. "Bubbles and Forecast Rationality in Experimental Markets for Stochastically-Lived Assets." Working Paper, Department of Management, University of Pennsylvania, 1992.
7. Carlson, John. "The Stability of an Experimental Market with a Supply Response Lag." *So. Econ. J.* **33** (1967): 305–321.
8. Cox, James, Vernon L. Smith, and James Walker. "Experimental Development of Sealed-Bid Auction Theory: Calibrating Controls for Risk Aversion." *Am. Econ. Rev.* **75** (1985): 160–165.
9. Daniels, Brian, and Charles R. Plott. "Inflation and Expectations in Experimental Markets." In *Bounded Rational Behavior in Experimental Games and Markets*, edited by R. Tietz, W. Albers, and R. Selten, 198–218. Berlin: Springer-Verlag, 1988.
10. Easley, David, and John Ledyard. "Theories of Price Formation and Exchange in Double Oral Auctions." This volume.
11. Eckel, Catherine, and Charles A. Holt. "Strategic Voting in Agenda-Controlled Committee Experiments." *Am. Econ. Rev.* **79** (1989): 763–773.
12. Ferris, Stephen P., Robert A. Haugen, and Anil K. Makhija. "Predicting Contemporary Volume with Historic Volume at Differential Price Levels: Evidence Supporting the Disposition Effect." *J. Finance* **43** (1988): 677–697.
13. Fisher, Irving. *The Theory of Interest.* New York: Macmillan, 1930.
14. Forsythe, Robert, and Russell Lundholm. "Information Aggregation in an Experimental Market." *Econometrica* **58** (1990): 309–347.
15. Forsythe, Robert, Thomas R. Palfrey, and Charles R. Plott. "Asset Valuation in an Experimental Market." *Econometrica* **50** (1982): 537–567.
16. Friedman, Daniel. "On the Efficiency of Experimental Double Auction Markets." *Am. Econ. Rev.* **74** (1984): 60–72.
17. Friedman, Daniel, Glenn W. Harrison, and Jon Salmon. "The Informational Efficiency of Experimental Asset Markets." *J. Pol. Econ.* **92** (1984): 349–408.
18. Garber, Peter M. "Tulipmania." *J. Pol. Econ.* **97** (1989): 535–560.

19. Garber, Peter M. "Famous First Bubbles." *J. Econ. Perspectives* 4 (1990): 35–54.
20. Gode, Dan, and Shyam Sunder. "Lower Bounds for Efficiency of Surplus Extraction in Double Auctions." This volume.
21. Harrison, Glenn, and Kevin McCabe. "Testing Bargaining Theory in Experiments." In *Research in Experimental Economics*, Volume 5, edited by R. M. Isaac. Greenwich: JAI Press, in press.
22. Holt, Charles A., Loren W. Langan, and Anne P. Villamil. "Market Power in Oral Double Auctions." *Econ. Inquiry* **24** (1986): 107–123.
23. Joyce, Patrick. "The Walrasian Tatonnement Mechanism." *Rand J. Econ.* **15** (1984): 416–425.
24. Knez, Peter, Vernon Smith, and Arlington Williams. "Individual Rationality, Market Rationality, and Value Estimation." *Am. Econ. Rev.* **75** (1985): 386–390.
25. Leroy, Stephen F. "Efficient Capital Markets and Martingales." *J. Econ. Lit.* **27** (1989): 1583–1621.
26. Lucas, Robert F. "Adaptive Behavior and Economic Theory." *J. Business* **59** (1986): S401–S426.
27. Machina, Mark J. "'Expected Utility' Analysis Without the Independence Axiom." *Econometrica* **50** (1982): 277–323.
28. Machina, Mark J. "Choice Under Uncertainty: Problems Solved and Unsolved." *J. Econ. Perspectives* **1** (1987): 121–154.
29. Ochs, Jack, and Alvin E. Roth. "An Experimental Study of Sequential Bargaining." *Am. Econ. Rev.* **79** (1989): 355–384.
30. Plott, Charles R. "Industrial Organization Theory and Experimental Economics." *J. Econ. Lit.* **20** (1982): 1485–1527.
31. Plott, Charles R., and Vernon Smith. "An Experimental Examination of Two Exchange Institutions." *Rev. Econ. Studies* **45** (1978): 133–153.
32. Porter and Smith. "Stock Market Bubbles in the Laboratory." Working Paper, University of Arizona, Department of Economics, 1990.
33. Roth, Alvin E., and M. K. Malouf. "Game-Theoretic Models and the Role of Information in Bargaining." *Psych. Rev.* **86** (1979): 574–594.
34. Sargent, Thomas. *Macroeconomic Theory.* New York: Academic Press, 1979.
35. Shefrin, Hersh, and Meir Statman. "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence." *J. Finance* **40** (1985): 777–790.
36. Smith, Vernon L. "Microeconomic Systems as an Experimental Science." *Am. Econ. Rev.* **72** (1982): 923–955.
37. Smith, Vernon L., Gerry L. Suchanek, and Arlington Williams. "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica* **56** (1988): 1119–1151.
38. Sobel, Joel, and Ichiro Takahashi. "A Multistage Model of Bargaining," *Rev. Econ. Studies* **50** (1983): 411–426.

39. Sunder, Shyam. "Experimental Asset Markets: A Survey." In *Handbook of Experimental Economics*, edited by A. E. Roth and J. Kagel. Princeton: Princeton University Press, 1992.
40. Thaler, Richard. "Anomalies: The Ultimatum Game." *J. Econ. Perspectives* **2** (1988): 195–206.
41. Tirole, Jean. "On the Possibility of Speculation Under Rational Expectations." *Econometrica* **50** (1982): 1163–1181.
42. Tirole, Jean. "Asset Bubbles and Overlapping Generations." *Econometrica* **53** (1985): 1071–1100. Reprinted with corrections on pages 1499–1528.
43. Tversky, Amos, and Daniel Kahneman. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *Quart. J. Econ.* **106** (1991): 1039–1061.
44. Walker, James M., Vernon L. Smith, and James C. Cox. "Inducing Risk-Related Preferences: An Examination in a Controlled Market Environment." *J. Risk and Uncertainty*, in press.
45. Wellford, Charissa. P. "A Laboratory Analysis of Price Dynamics and Expectations in the Cobweb Model." Discussion Paper 89-15, Department of Economics, University of Arizona, September 1989.
46. White, Halbert. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* **48** (1980): 817–837.
47. Williams, Arlington W. "The Formation of Price Forecasts in Experimental Markets." *J. Money, Credit, & Banking* **19** (1987): 1–18.
48. Wilson, Robert. "Incentive Efficiency of Double Auctions." *Econometrica* **53** (1985): 1101–1115.

John R. O'Brien and Sanjay Srivastava
Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh, PA 15213

FOURTEEN
Liquidity and Persistence of Arbitrage in Experimental Options Markets

1. INTRODUCTION

In this chapter, we continue our study of the relationship between arbitrage and information aggregation in experimental stock and options markets. In previous research,[2] we examined the informational implications of zero arbitrage. We showed that if the set of securities is sufficiently rich, then the only prices for which no trader can perceive an arbitrage opportunity based on his/her private information are the full information prices.

The theory identified a set of portfolios, called separating portfolios, which must trade at a zero price in an arbitrage-free equilibrium. If they did not have a zero price, then it was shown that some trader would perceive an arbitrage opportunity *based only on his/her private information.* The theory thus suggests a dynamic story about convergence to rational expectations. One might expect that over time, arbitrage opportunities disappear, in which case we must converge to rational expectations.

In the previous experiments, we found that while there were generally no unexploited arbitrage opportunities in the separating portfolios, we did not always obtain convergence to a rational expectations equilibrium (REE). We did find, however, that in many cases where the market failed to converge to REE, the arbitrage opportunities persisted over time. What was surprising about the persistence was that it tended to occur in later periods rather than early in the experiments. In some cases, we found lack of convergence and also a wide bid-ask spread in the separating portfolios.

These observations lead to two questions. First, how does the bid-ask spread affect the argument relating arbitrage to efficiency? Second, and this is the main focus of this paper, is there some theoretical reason why arbitrage opportunities can persist?

In Section 2, we examine the effects of the bid-ask spread on information aggregation, and show that the spread in the separating portfolios bounds the deviation from the REE. This relationship and the rules of the double auction market then motivate a new experiment in which traders are given (indirect) incentives to reduce the bid-ask spread. We then show that given the structure of our experimental markets, persistent arbitrage opportunities are quite possible and need not imply any "irrationality" on the part of traders. We show that based on the theory, we can classify the data into three categories. The first of these is a situation relating to wide bid-ask spreads. The second is when arbitrage opportunities can persist. The last category is one which should not occur; it involves some trader offering prices to the market while knowing full well that he/she is providing an arbitrage opportunity to the market.

The experimental framework is described in Section 3. In Section 4, we reexamine some data from our old experiments and also the data from the new experiment in light of the results of Section 2. We show that we rarely find prices in the last category described above. In the older experiments, the data falls into either the first category or the second. In the new experiment, where we expect to find narrow bid-ask spreads, we find that prices either converged to an REE or fell into the second category.

Our entire discussion will concentrate on the implications of arbitrage. We deliberately ignore other factors which may or may not lead to information aggregation. This is not to say that these factors are unimportant; they may be very significant, but our interest in this line of research is to examine the extent to which arbitrage arguments alone can explain behavior in experimental asset markets.

2. LIQUIDITY, ARBITRAGE, AND EFFICIENCY

We start with a summary of the relationship between arbitrage and efficiency, taken from O'Brien and Srivastava.[2] The presentation here is a variation which applies to the experimental markets analyzed in this paper. The analysis is simplified by the fact that for the data analyzed in this paper, there is a complete set of markets. The argument requires unlimited ability to borrow and short sell, both of which are present in our experiments.

2.1 ARBITRAGE AND EFFICIENCY

Consider a market with n assets, $x_1, \ldots, x_n$, and let $S = \{s_1, \ldots, s_m\}$ denote the set of states. Let $x_j(s)$ denote the payoff from asset j in state s, and assume that $x_j(s) \geq 0$ for all j and s and $x_j(s) > 0$ for some s. If state s is realized, we denote by $E^i(s) \subseteq S$ the set of states agent i knows has occurred. $E^i(s)$ is called the *private information* of agent i in state s. We assume that $s \in E^i(s)$ for all s and that $\bigcap_i E^i(s) = \{s\}$, so that information always includes the truth and there is no aggregate uncertainty. We assume that there is a risk-free asset which has the same strictly positive payoff in every state and, without loss of generality, that the risk-free rate is zero. Let p_j denote the price of asset j.

A portfolio is an n-vector, $w = (w_1, \ldots, w_n)$, where w_j denotes the weight of asset j in the portfolio. $p(w) = \sum w_j p_j$ is the price of portfolio w, and $x_w(s)$ denotes the payoff from portfolio w in state s.

There is no arbitrage at prices p given $E^1(s), \ldots, E^I(s)$ if there is no agent i and portfolio w such that $p(w) < 0$ and $x_w(t) \geq 0$ for all $t \in E^i(s)$.

In the experimental framework, there are three states,

$$S = \{r, s, t\},$$

and two types of agents. If state s occurs, the first (type of) agent has information $E^1(s) = \{r, s\}$, while $E^2(s) = \{s, t\}$. There is a complete set of markets, so let $w(r)$, $w(s)$, and $w(t)$ denote the portfolios which reproduce the state-contingent claims for the three states, each paying 1 in the relevant state. Finally (dropping the subscript), let x be the security we are focusing on, and suppose without loss of generality that $x(r) < x(s) < x(t)$, and let p denote the price of x. If state s occurs, then the rational expectations equilibrium price of x is $x(s)$ since the market as a whole has perfect information.

The relationship between arbitrage and efficiency is straightforward when markets are complete.[1] Suppose state s occurs. Then, agent 1 knows that $w(t)$ is worth zero, so no arbitrage implies $p(w(t)) = 0$. Note that agent 1 can price $w(t)$ based only on his/her private information. Similarly, the information of agent 2 implies that $p(w(r)) = 0$. If $p > x(s)$, then agent 1, for example, can sell one unit of x and sell $[x(s) - x(r)]$ of $w(r)$. This yields revenue of p, and implies a future liability of $x(s)$ for every state in $E^1(s)$. Since the interest rate is zero, agent 1 has an arbitrage opportunity. Lack of arbitrage then implies that $p \leq x(s)$. If $p < x(s)$, then the same agent can buy one unit of x and buy $[x(s) - x(r)]$ of $w(r)$. Therefore, we must have $p = x(s)$.

[1]The general case is analyzed in O'Brien and Srivastava.[2]

The state contingent claims here are special cases of what are called separating portfolios in O'Brien and Srivastava,[2] and from now on, we will call them separating portfolios in this paper.

The fact that arbitrage implies information aggregation is illustrated by the following example (which is a special case of the trading environment described in the next section). Suppose a stock pays a dividend of either 0, 25, or 45, depending on the state, the actual state is such that the dividend will be zero, and there are two types of traders, T1 and T2, whose information events in terms of stock payoffs are:

Trader Type	Possible Dividends
T1	0,45
T2	0,25

Suppose that besides the stock, we have a put option and a call option, both with a strike price of 30. Consider the portfolio $S - 3C$, i.e., long one stock and short three calls. Then, this portfolio pays zero to T1 whether the dividend is 0 or 45. Lack of arbitrage says this must trade at a zero price, i.e., $S - 3C = 0$. Trader T2 knows that the call is worthless (recall that the strike price is 30), so we get $C = 0$. Together, these two restrictions imply $S = 0$, $C = 0$. The parity relationship (which in our setting with a zero interest rate reduces to $S - C + P = 30$) implies that $P = 30$, where P is the price of the put on stock 1. Thus, lack of arbitrage implies full information aggregation.

2.2 THE EFFECT OF LIQUIDITY

To examine the effects of the bid-ask spread on the arbitrage-based argument, first suppose that there is a zero bid-ask spread in the market for security x, i.e., $p^a = p^b = p$. For a portfolio the "bid" price is the cost of buying one unit of the portfolio. Similarly, the "ask" price is the revenue that is generated by selling one unit. Assuming again that state s occurs, lack of arbitrage implies

$$p^b(w(r)) \leq 0 \leq p^a(w(r)) \text{ and } p^b(w(t)) \leq 0 \leq p^a(w(t)).$$

Consider again the argument employed above for the case $p > x(s)$: agent 1 sells one unit of x and sells $[x(s) - x(r)]$ of $w(r)$. The portfolio $w(r)$ can only be sold at the price $p(w(r))$, so this strategy is not profitable for agent 1 as long as

$$p + p(w(r))[x(s) - x(r)] \leq x(s)\,.$$

We know that no arbitrage implies $p^b(w(r)) \leq 0$; if it is strictly negative, then this inequality can hold even though $p > x(s)$ since $[x(s) - x(r)]$ is positive. In the case that $p < x(s)$, there is no arbitrage as long as

$$p + p^a(w(r))[x(s) - x(r)] \geq x(s)\,.$$

Again, this can hold with $p^a(w(r)) > 0$ since the bracketed term is positive. Taken together, we find that lack of arbitrage now only yields

$$p^b(w(r))[x(s) - x(r)] \leq x(s) - p \leq p^a(w(r))[x(s) - x(r)],$$

so the deviation from the efficient market price is bounded by the spread in the separating portfolios. A similar bound results from the second portfolio.

Finally, we complete the argument by relaxing the assumption that $p^b = p^a$. This leads to the following two inequalities for the two sides of the arbitrage argument:

$$x(s) - p^b \geq p^b(w(r))[x(s) - x(r)],$$
$$x(s) - p^a \leq p^a(w(r))[x(s) - x(r)].$$

Again, since we must have $p^a \geq p^b$ if there is no spread in separating portfolios, then the market must aggregate all information. Conversely, if there is a wide spread in the separating portfolios, it is possible for there to be no arbitrage at prices quite different from the rational expectations equilibrium.

This can be illustrated by the above example. We stated that $S - 3C = 0$; otherwise, trader T1 would see an arbitrage opportunity. At the level of bid-ask spreads, lack of arbitrage implies[2]:

1. $S^a - 3C^b \geq 0$, so trader T1 cannot buy this portfolio for a negative price,
2. $S^b - 3C^a \leq 0$, so T1 cannot sell this portfolio for a positive price,
3. $C^b \leq 0$, so T2 cannot sell calls for a positive price, and
4. $C^a \geq 0$, so T2 cannot buy calls for a negative price.

Since we restrict bids/asks to be nonnegative, implication 3 reduces to $C^b = 0$, and implication 4 is not a binding constraint. Now, consider the following bids/asks:

$S^b = 24$	$C^b = 0$	$P^b = 4$
$S^a = 26$	$C^a = 8$	$P^a = 5$

It is easy to check that implications 1–4 are satisfied, and in addition, the parity bounds are satisfied (these are $S^a - C^b + P^a \geq 30$, $S^b - C^a + P^b \leq 30$). Here, the bids and asks across the market do not permit any arbitrage, yet the prices are quite far from the rational expectations equilibrium (which has a stock price of 0, call of zero, and a put price equal to 30). In fact, these prices would seem to indicate convergence to 25 and could easily be interpreted as an indication of "false" convergence. Note that the spread is 26 for the first portfolio, and 8 for the second, even though it is quite narrow in the stock and put markets. Note that the spread in the separating portfolio is the sum of the spreads in the securities which make up the portfolio weighted by the absolute value of their portfolio weights.

[2] Lack of arbitrage also requires asks to be greater than or equal to bids. Also, trader T1 should not bid more than 45 for the stock or 15 for the call. Similarly, T2 should not bid more than 25 for the stock, and so on.

The example indicates that arbitrage arguments are very sensitive to the spreads *across* markets, and if we rely only on arbitrage arguments, then it is possible for the market to settle at an arbitrage-free set of bids and asks with a wide spread in at least one of the markets.

Based on this development, we define "liquidity" in terms of the bid-ask spread in the separating portfolios. A perfectly liquid market is one in which a trader can buy or sell any quantity of an asset with no transactions costs. Hence, a liquid market for us is one in which a trader can buy or sell securities without facing a wide bid-ask spread.

The fact that the arbitrage argument is sensitive to bid-ask spreads motivated the manipulation which provided incentives for attempting to narrow the spread. One feature of the double auction institution is that if bids and asks exist, then any subsequent bids and asks *must* narrow the bid-ask spread. Thus, if we wish to narrow the spread, we simply have to induce traders to enter bids and asks more frequently. However, since we do not want to reward "spurious" bids and asks (i.e. very low bids and very high asks), we should only reward them on the basis of bids and asks which were actually accepted. Our new experiment consists of exactly this manipulation i.e., we rewarded traders on the basis of the total volume of bids and asks they submitted and which were accepted. The rewards were computed separately for the stock and option markets, as described in the next section. Of course, if traders are not competing actively to be the market makers, we may not observe narrow spreads, since under the rules of the double auction, there is no minimum bid or maximum ask once the market clears.

2.3 PERSISTENCE OF ARBITRAGE

Our theory provides a classification of possible outcomes from an experiment. In particular, if there are no arbitrage opportunities, then failure to attain the rational expectations equilibrium can only be due to lack of liquidity. As we will see, in some trials arbitrage opportunities existed, and interestingly, they persisted over time.

The main question, then, is whether in a market with "rational" traders arbitrage opportunities can persist. Recall that the theory only says that if the market is perfectly liquid *and* there are *no* arbitrage opportunities, we must be at the REE. However, the theory does not say that arbitrage opportunities cannot persist, even with rational traders. As we now show, it is possible for the market to converge to a *false* equilibrium in a manner perfectly consistent with the information of all traders and with the following behavioral assumptions:

1. If trader i can determine that there is an arbitrage opportunity *based only on the trader's private information set* $E^i(s)$, then the trader exploits the opportunity.
2. The trader never knowingly gives up an arbitrage opportunity: i.e., he never submits bids and asks in the various markets such that if he were faced with these prices, he would know there is an arbitrage opportunity *based only on* $E^i(s)$.

The fact that arbitrage opportunities can persist is illustrated quite dramatically by the following example, which has one stock, a put option, and a call option with strike prices of 30. Suppose the stock can pay 0, 35, or 60, the true payoff is 0, and there are two types of traders. Type 1 (T1) knows the dividend is either 0 or 35, and type 2 (T2) knows it is 0 or 60. Then, the separating portfolios are: $S - 7C = 0$ and $S - 2C = 0$. Of course, these together imply $S = 0$, $C = 0$.

Consider T1. Being a rational trader, T1 is willing to submit any stock bid and a call ask pair such that

$$S^b - 7C^a \leq 0. \qquad (T1)$$

Submitting these involves offering to buy the stock and sell the calls. Note that T1 is *making* the market at these prices.

Trader T2 is willing to accept this, i.e. sell the stock to T1 and buy the calls from T1, as long as

$$S^b - 2C^a \geq 0. \qquad (T2)$$

Trader T2 here is *taking* the market.

If these inequalities are strict, then *both* parties make arbitrage profits. T1 buys at a price less than zero and the portfolio pays off zero in both states in his information set. T2 sells the portfolio to T1 at a positive price, and the portfolio is worth zero at every state in his information set.

Note that there is an imbalance in that the supply of calls exceeds the demand[3] (seven calls are offered for every two calls demanded), so one might expect that eventually prices respond to this imbalance. However, the rules of the double auction *do not* require market clearing in *offers* at any point in time. This is analogous to a "Ponzi" scheme, and we expect that given sufficient time the bubble would burst. However, the situation can persist over the time intervals of our typical experiment, particularly if the market is competitive in that several traders are competing to make the market and others are competing to accept the offers.

The arbitrage possibilities are depicted in Figure 1. The ray $0A$ represents the equation $S = 7C$, and $0B$ represents $S = 2C$. Consider the point marked Z on the graph. This represents $C^a = 6$, $S^b = 35$. At this point type T2 perceives an arbitrage opportunity since $S^b - 2C^a > 0$. By selling the stock at 35 and buying two calls at 6 each, T2 guaranteed a profit of 23. On the other hand, consider T1. If T1 is making both the markets, i.e. submitting the stock bid and the call ask, then T1 is guaranteed a profit of 7 by having this portfolio trade. If we observe this, we would conclude that the market converged to a false equilibrium. The region with the point marked Z in Figure 1, is the region in which mutually beneficial arbitrage is possible. Figure 1 also shows the regions in which T1 and T2 would make offers and the regions in which they would accept offers made by the other type. Lack of liquidity is indicated by points such as x.

[3]The arbitrage arguments obviously require variable quantities to be traded and also require short selling and borrowing.

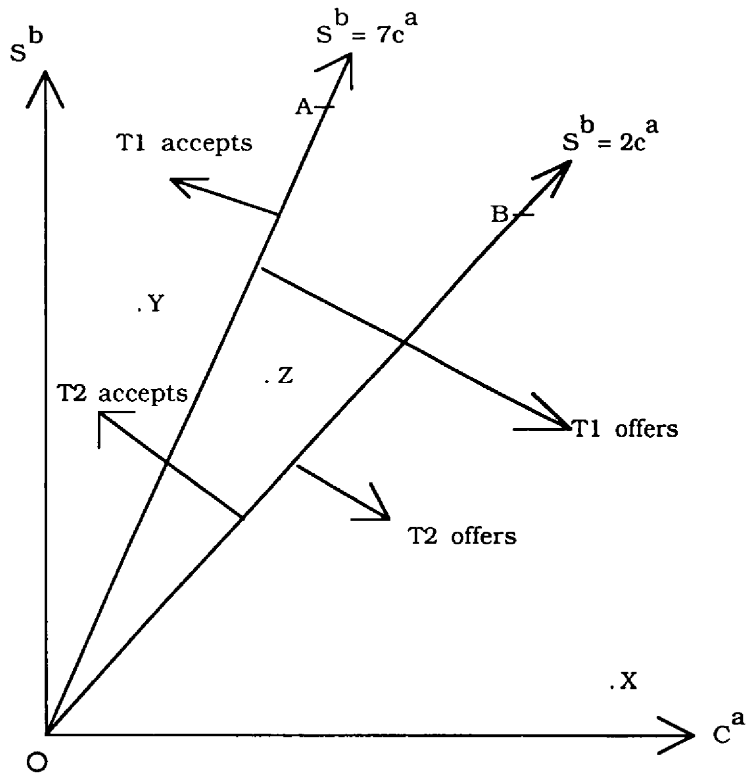


FIGURE 1

We are thus in a position in which *both* T1 and T2 are both making arbitrage profits.[4] While our focus is entirely on the extent to which we can explain observations using only arbitrage arguments, we should note that it may be difficult for a trader to make inferences about the information of the other trader types. This is because we cannot make a clear prediction about which trader type will make markets. For instance, we stated that trader type T1 is willing to offer the point Z in Figure 1. However, type T2 is also willing to make the stock market at that price, since offering $S^b + 2P^b \geq 210$ is an arbitrage-free offer for type T2. T2 is also willing to submit a call ask as long as $C^a + P^a \leq 30$. The fact that there is no clear prediction about who can make which market thus interferes with the ability

[4]Since we are permitting unlimited borrowing and short selling, this points out that we need to limit our liability in some other way. The binary lottery technique is an extremely convenient device in this regard. If we restrict borrowing and short selling, this could interfere with the arbitrage arguments.

of traders to make inferences about the types of other traders from market prices, since a given bid-ask pair is not uniquely associated with a trader type.

One might expect that competition among, say, the T1 traders reduces the arbitrage profits accruing to these traders. This would involve increasing the stock bid or reducing the call ask. Reducing the call ask goes in the direction of the REE, but increasing the stock bid goes in the wrong direction. In fact, the imbalance in the call market, together with some notion of supply-demand adjustment, would seem to make it more likely that the call ask would come down. Reducing the call ask would *increase* the profits to the T2 traders.

Once we see zero profits for the market makers, say, if T1 traders set $S^b = 35$, $C^a = 5$, the rules of the double auction can actually hinder information aggregation. From this point, a T1 trader would like to decrease the stock bid and increase the call ask in order to generate arbitrage profits. The rules of the double auction prevent any trader from doing this unilaterally if S^b and C^a exist. If there was a single market maker, then this trader could obviously do this, once one side of one market clears (either the stock bids clear or the call asks clear.) With competition among the T1 traders, this may prove difficult.[5] There is also the possibility of competition among the T2 traders. With these prices, the act of selling a stock and buying two calls yields a riskless profit of 25. Instead of being market takers, they could offer to say sell the stock at $S^a < 35$. However, if $S^b = 35$, then this involves knowingly giving up an arbitrage opportunity.[6]

Note that in this example, the larger is S^b or C^a the greater is the sum of the acceptable arbitrage possibilities for the two types of traders. However, if prices are very far away from the REE, then there are other arbitrage possibilities; in this example if $S^b > 35$, then the T1 traders can obtain a riskless gain by selling the stock.

We observe that the possibility of persistent arbitrage is created by the fact that different types of agents can play different market-making roles. This is actually independent of liquidity. To see this, note that if $S^a = S^b$ and $C^a = C^b$, then $S^b - 7C^a < 0$, (which is happily offered by T1 in the example), implies $S^a - 7C^b < 0$. This means that the stock ask and call bid cannot both be submitted by trader type T1.

Returning to Figure 1, we see that our rationality assumption does impose restrictions on (S^b, C^a) pairs we would *not* expect to see. If we observed a bid/ask pair such as point Y in Figure 1, then whoever is making the market is knowingly giving up an arbitrage opportunity whether or not there is a liquidity problem. This is a testable prediction.

Finally, we note that our ability to observe mutually beneficial arbitrage depends on observing bid/ask pairs in the triangular region. This actually requires

[5] Our liquidity-enhancing maneuver may produce more competition among the T1 traders, making aggregation less likely in the current situation.

[6] Note that in the arguments we have used up to this point, no trader type is knowingly giving up money.

that the bid side of the separating portfolio not be too far away from zero. For instance, the more negative is $S^b - 7C^a$, the less likely it is that $S^b - 2C^a$ is positive. This completes the motivation for our liquidity manipulation.

3. THE EXPERIMENTAL FRAMEWORK

We will analyze data from two of the experiments from O'Brien and Srivastava,[2] and also the results from a new experiment. The experiments conducted earlier were designed to study a broader and somewhat different set of issues than those of interest here. For control purposes, the new experiment was conducted under the same conditions as the previous ones.

3.1 THE MARKET STRUCTURE

There are 2 two-period lived stocks. The payoff ("dividend") from a stock depends on both a first-period and a second-period state realization. The states in either period are labelled x, y, and z, respectively, the dividend in period one is always zero, and second-period dividends from either stock are given in the following matrix.

$$
\begin{array}{ccc}
 & & \text{Period 2} \\
 & & \begin{array}{ccc} x & y & z \end{array} \\
\text{Period 1} & \begin{array}{c} x \\ y \\ z \end{array} & \begin{pmatrix} 0 & 20 & 40 \\ 0 & 25 & 45 \\ 0 & 35 & 60 \end{pmatrix}
\end{array}
$$

If the first-period state is y and the second-period state is z, then the stock will have a liquidating dividend of 45 at the end of period 2, and so on. All states are equally likely, and the two stocks are independent. Thus, while the two stocks are identical *ex ante*, they can differ *ex post*.

There is one American put option on stock one with an exercise price of 30, and one American call option on stock one with an exercise price of 30. Both these securities expire at the end of period 2.

Each run of a two-period market is called a trial. There were 16 traders in every market, and we ran ten trials. Each period lasted 240 seconds.

3.2 INFORMATION AND ENDOWMENTS

The information structure is as follows. In every period, the market as a whole has complete information about all periods. A typical trader receives information which eliminates one or more states from consideration. For example, if the true states are y in period 1 and z in period 2, one trader would receive the following information at the beginning of period 1: Not x in period 1, not y in period 2. At the beginning of period 2, all traders were told what the first-period state was, and the second-period information was repeated. Continuing with the example, at the beginning of period 2, the above trader would be told that the first-period state was y, and the second-period state was not y.

The private information given to a trader and the disclosure of the first-period state were printed on the computer screens of the traders. The distribution of information was chosen as follows. Given that a state occurs, there are 2 first-period states and 2 second-period states which did not occur. This means that in period 1, there are four possible clues of the form "Not x in period 1, not y in period 2." The four clues were given to the first four traders as follows. We used a random number generator to decide which of the four traders would receive the first clue. Similar randomization among the remaining three traders determined who got the second clue, and so on. This procedure was repeated with the next group of four traders, and so on. In all of the trials, we had 16 traders, so in period 1, exactly one-quarter of the market had a given clue. In the second period, there are two possible clues since the first-period state was disclosed. We used the same randomization process but among groups of two traders. In this case, exactly half the market had a given clue.

At the beginning of each trial, every trader was given an initial endowment of stocks and market cash. There were three endowment types. The first received 100 units of stock 1 and 2500 in market cash. The second type received 100 of stock 2 and 2500 in cash. The third type received 100 of each stock and no cash. The *ex ante* expected final value of all the endowment types is 5000. The traders rotated across the various endowment types in sequence as follows. In the first trial, the first trader was type 1, the second was type 2, the third type 3, the fourth type 1, and so on. In the second trial, the first trader was type 2, the second was type 3, the third was type 1, and so on.

3.3 THE TRADING RULES

The trading mechanism follows the rules of the double auction. Any trader can submit bids/asks and accept bids/asks for any of the four securities in the market. If there is an existing bid, the next bid has to be strictly greater, and the opposite holds for asks. Traders must specify the quantities they wish to trade. Finally, they can choose to exercise any option on which they have a long position. If an option is exercised, it is exercised against a randomly chosen trader who has a short position in the option. If a trader exercises one call (put) option, the trader pays (receives) the exercise price and gives up (gains) one unit of the stock. Unlimited borrowing (at a zero interest rate) and unlimited short selling is permitted.

3.4 THE PAYMENT SCHEME

At the end of period 2, dividends are paid. Any option which is in the money given final dividends is exercised, and the end of trial market cash of each trader is determined. If a trader ends up with a short position in a stock at the end of period 2, then the dividend payment is subtracted from the trader's market cash.

Traders are then compensated using the binary lottery technique (see Roth and Malouf[3] and Berg et al.[1]) where we induced risk neutrality as follows. We drew a random number between 0 and 9999. If a trader's end-of-trial market cash exceeded this number, then the trader won a (real) cash prize of \$2. Otherwise, the trader won \$0. Note that a trader who has final cash in excess of 10,000 is assured of the prize, while one with negative cash is assured of not winning. The *ex ante* expected value of 5000 translates into a 50% probability of winning the prize.

While our analysis is independent of the actual and induced risk preferences of the traders, there is a very important reason for using the lottery technique. The arbitrage arguments require us to allow unlimited borrowing and short selling. The binary lottery technique has the important feature that it permits us to allow unrestricted borrowing and short selling with experimental liability limited to the cash prize.

In addition to this reward scheme, we also compensated traders in the new experiment for making markets, as follows. In each two-period market, or trial, we calculated the total volume of bids and asks submitted by each trader which were accepted by some other trader. This computation was done separately for the two stocks and for the options. In each trial, we determined the top three market makers for stocks and the top three market makers for options among the traders who ended the trial with nonnegative market cash. These traders were paid a lump sum of \$.25. There was no incentive for market making in the experiments described in O'Brien and Srivastava.[2]

The traders were paid at the end of the experimental session of ten trials. The earnings they had cumulated in previous trials were known to them in that they noted them on a sheet of paper. Cumulative earnings were also displayed on their screen at the end of every trial.

3.5 THE TRADERS

In every experiment reported here, the traders were master's students at the business school at the authors' home institution. The students were very experienced in trading in our computerized markets. The experience level of the students was similar. They had all participated in a market with only two stocks (with a different payoff structure from the current experiment) for at least twenty 5-minute trials conducted over three weeks. These trials were conducted in exactly the same manner as the experiments: the students received information, rotated through endowment types, were compensated using the binary lottery technique, etc. They had two practice trials of the experimental market immediately before the experiments. The option markets, option payoffs, and the process of exercising an option were explained to the students one week before the practice and experimental runs. The reward for market making in the new experiment was explained immediately before the two practice options trials. Each of the three markets consisted of an independent group of traders.

3.6 THE EXPERIMENTAL DESIGN

We will restrict attention to the second-period data for stock 1 and the options, since there is a complete set of markets for stock 1 in period 2, and it is relatively simple to construct and illustrate the appropriate portfolios. We will ignore those parts of the data which do not contribute to the analysis of liquidity and persistence. These include the period 1 data and the data on stock 2. (There are no separating portfolios for stock 2, and it was originally introduced as a control against stock 1. The period 1 data served as a control against period 2 since, in most cases, there are an insufficient number of separating portfolios in period 1 to imply aggregation. See O'Brien and Srivastava.[2])

In the interests of comparability, our new experiment is an exact replication of the previous one with an independent group of traders and with the market-making incentive described earlier, even though the points of interest in this paper could have been studied in a simpler market.

It can be shown that in period 2, for every possible state realization, there is a complete set of markets. The state-contingent claims for every state realization and every possible information type are presented in Table 1. There are multiple

TABLE 1 Separating Portfolios

	T1		T2	
State	Info	Portfolio	Info	Portfolio
xx	0,20	$S+P=30$	0,40	$3S+4P=120$
xy	0,20	$S+P=30$	20,40	$S+2P=40$
xz	0,40	$S-4C=0$	20,40	$S-2C=20$
yx	0,25	$S+P=30$	0,45	$2S+3P=90$
yy	0,25	$S+P=30$	25,45	$S+4P=40$
yz	25,45	$3S-4C=75$	0,45	$S-3C=0$
zx	0,35	$S-7C=0$	0,60	$S-2C=0$
zy	0,35	$S-7C=0$	35,60	$S-C=30$
zz	0,60	$S-2C=0$	35,60	$S-C=30$

separating portfolios for each state. For the purposes of this paper, we have made the following choices. First, we choose separating portfolios such that both the separating portfolios for a given state involve the same option (see Table 1). This makes it possible to depict the arbitrage possibilities in a two-dimensional picture. Second, if possible, we choose portfolios involving the call option. This is because our experience has been that there is more trading in calls than in puts in the experimental markets.

4. EXPERIMENTAL RESULTS

In this description, markets A1 and A2 are those in O'Brien and Srivastava,[2] while A3 is the identical market with an independent group of traders and the market-making incentive. We eliminated all observations which involved a stock bid exceeding 65 and option bids exceeding 35. These were the same adjustments as made in our previous paper, and happened extremely rarely.

We show the aggregate results in Figures 2–7. We describe Figure 2 in detail. In both graphs, each trial is marked by the vertical lines, and the state is given at the bottom of the graph. The first graph on the figure has time on the horizontal axis, and the plotted points give information on bids which were submitted to the market. The bids are those relating to the separating portfolio. For example, in the first trial, the state is "yy," and Table 1 indicates that the relevant portfolios are $S + P = 30$ and $S + 4P = 45$. The bid side of the first portfolio is $S^b + P^b - 30$, while that of the second is $S^b + 4P^b - 45$. As discussed in the previous section, a T1 trader knowingly gives up arbitrage if $S^b + P^b - 30 > 0$, and similarly for the T2 trader.

In the first trial of Figure 2 we have plotted the points as follows:

1. Suppose $S^b + P^b - 30 > 0$ and $S^b + 4P^b - 45 < 0$. Then, we plot the point as zero; these bids can be submitted by a rational T2 trader and accepted by a rational T1 trader. If $S^b + P^b - 30 < 0$ and $S^b + 4P^b - 45 > 0$, then for the same reason we also plot it as a zero.
2. Suppose $S^b + P^b - 30 < 0$ and $S^b + 4P^b - 45 < 0$. Then, we plot the value which is closer to zero. In this situation, no trader type could be giving up arbitrage, and no trader type would accept these offers. The distance from zero gives us a measure of liquidity since, as discussed in Section 2.3, we measure liquidity by the spread in the separating portfolios. When there is no arbitrage, the stock prices are bounded by the spreads in *both* the separating portfolios, and the minimum of these distances is then the appropriate measure (recall from Section 2 that the deviation from the REE is bounded by the spreads in both separating portfolios).

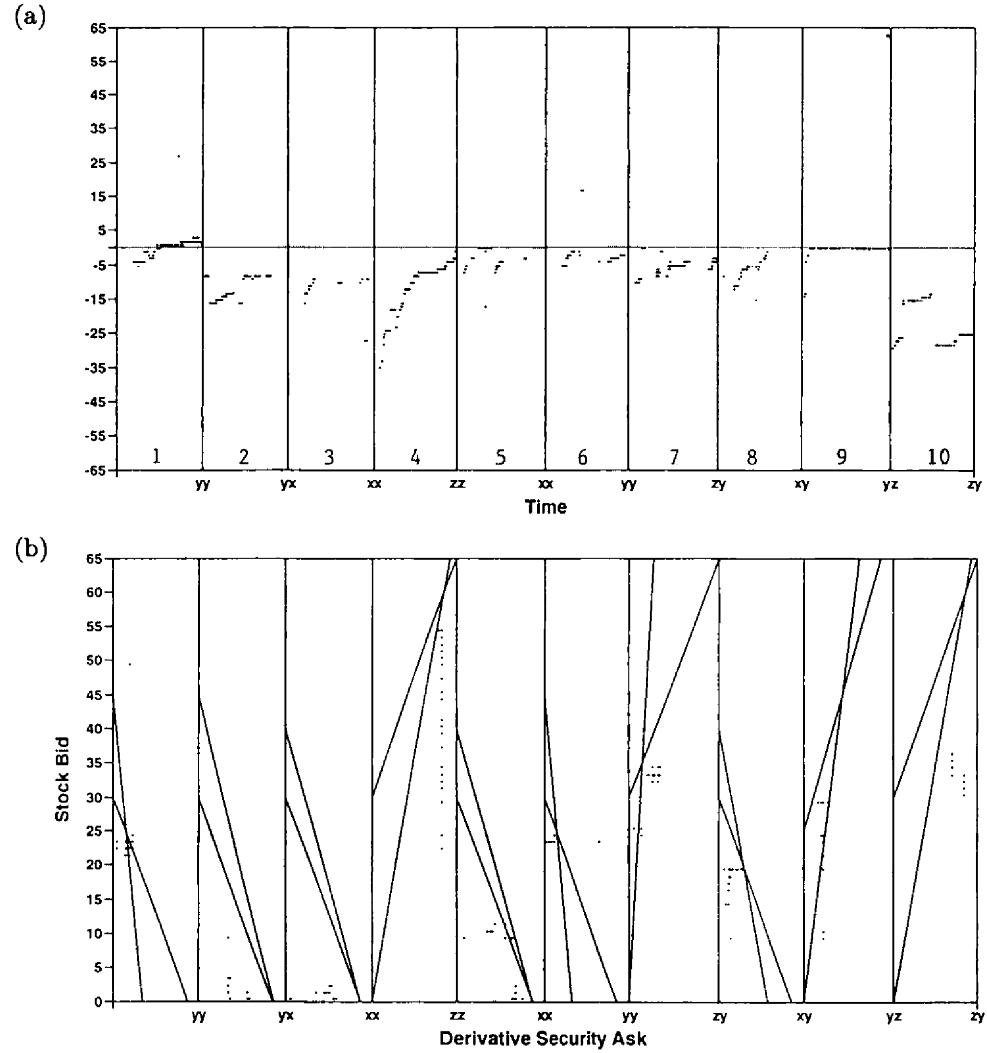


FIGURE 2 (a) Dynamic classification by region: Market A1, Bids.
(b) Arbitrage persistence regions: Market A1.

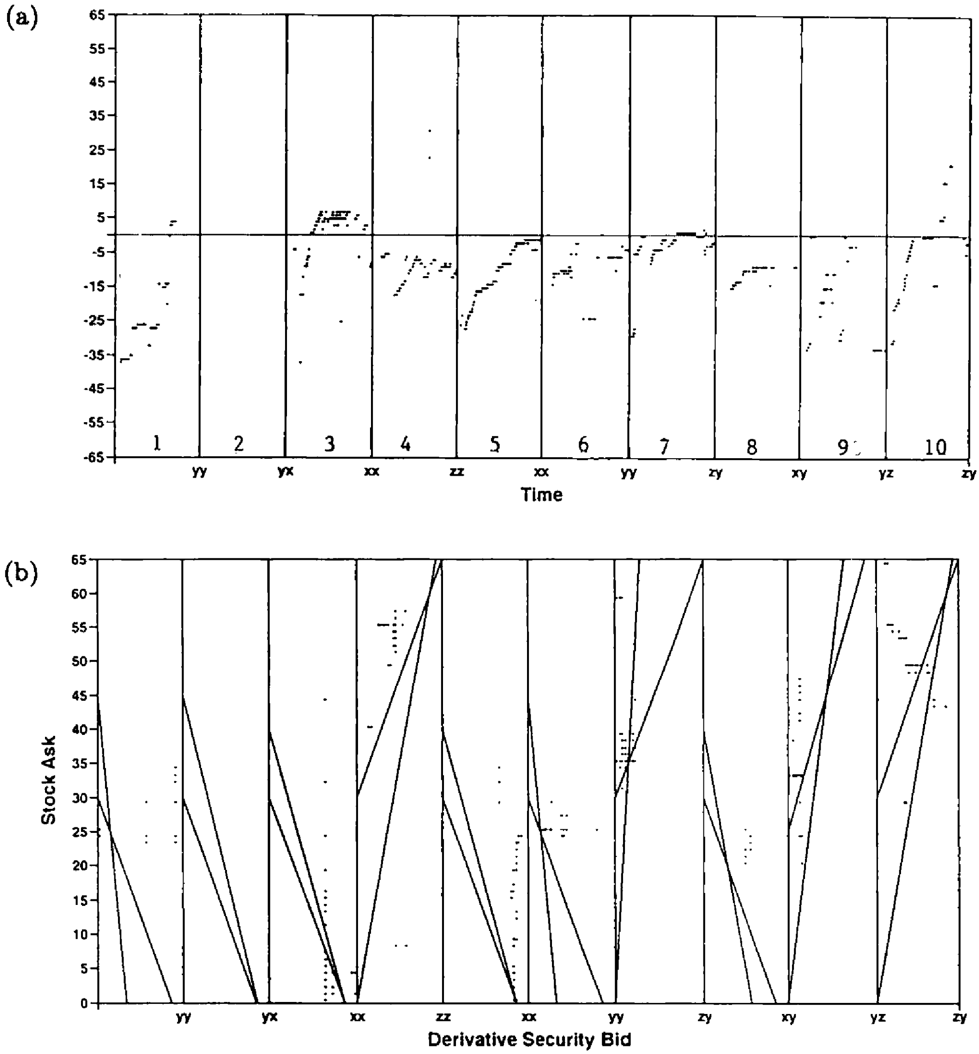


FIGURE 3 (a) Dynamic classification by region: Market A1, Asks.
(b) Arbitrage persistence regions: Market A1.

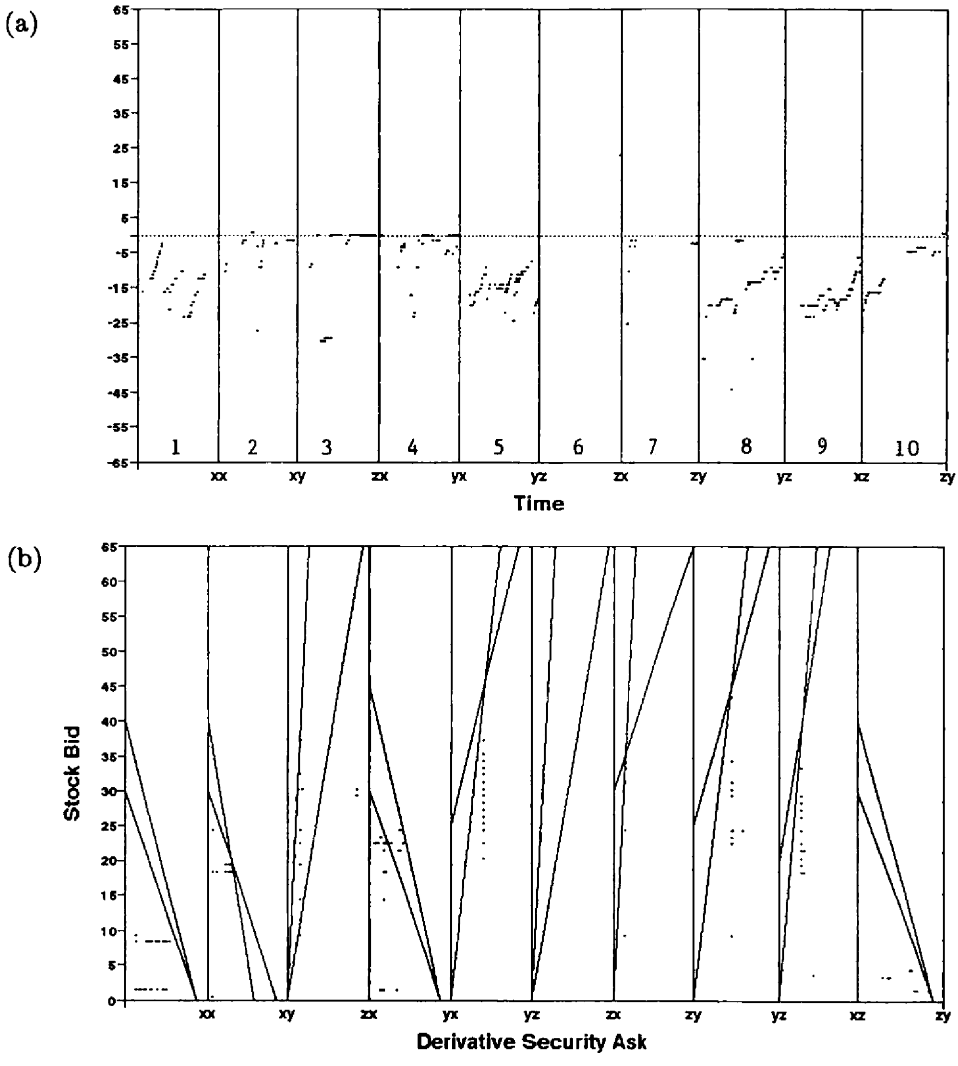


FIGURE 4 (a) Dynamic classification by region: Market A2, Bids.
(b) Arbitrage persistence regions: Market A2.

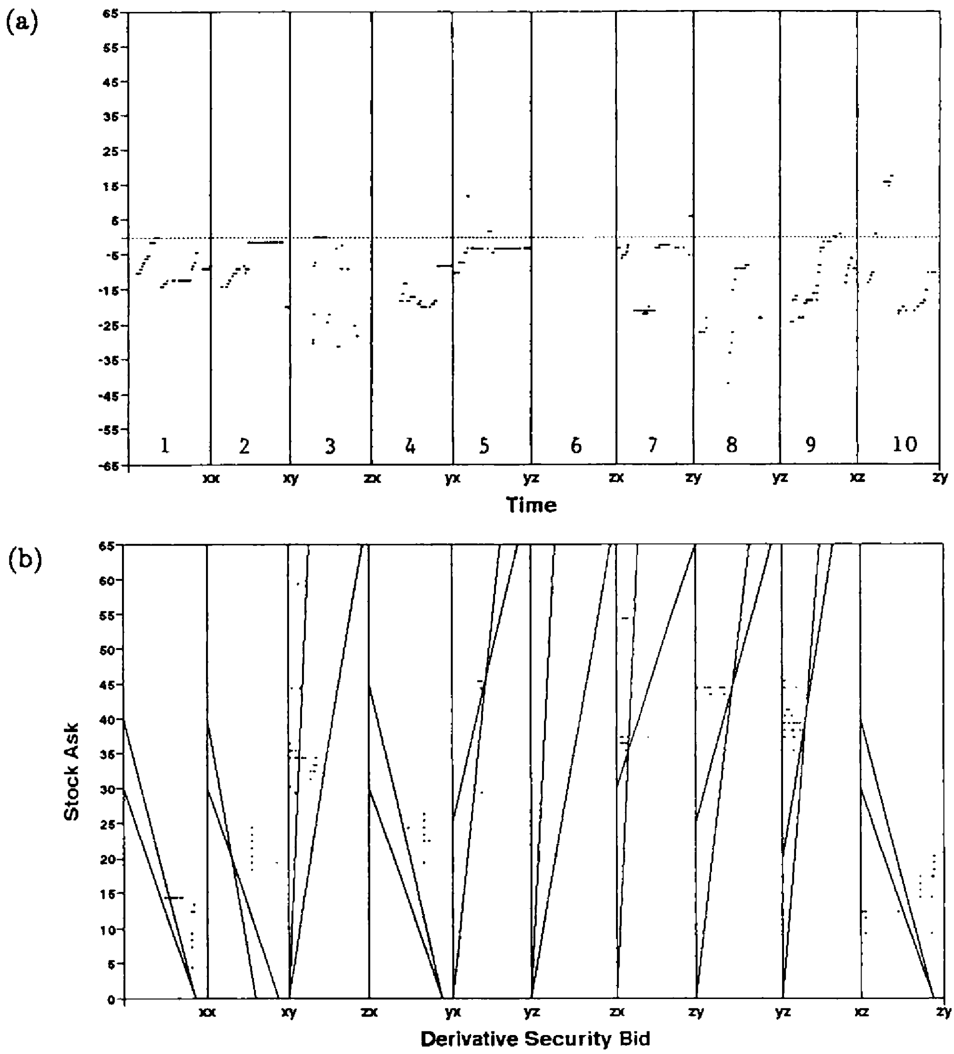


FIGURE 5 (a) Dynamic classification by region: Market A2, Asks.
(b) Arbitrage persistence regions: Market A2.

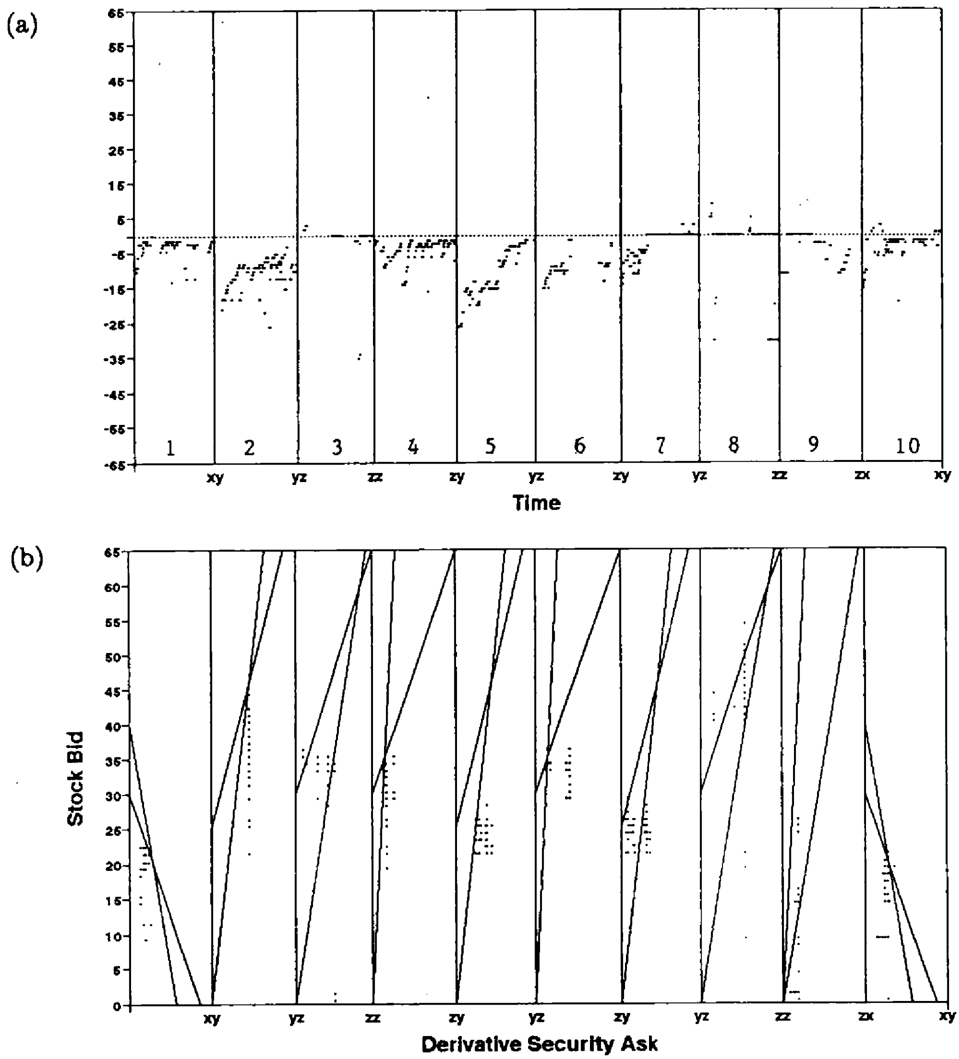


FIGURE 6 (a) Dynamic classification by region: Market A3, Bids.
(b) Arbitrage persistence regions: Market A3.

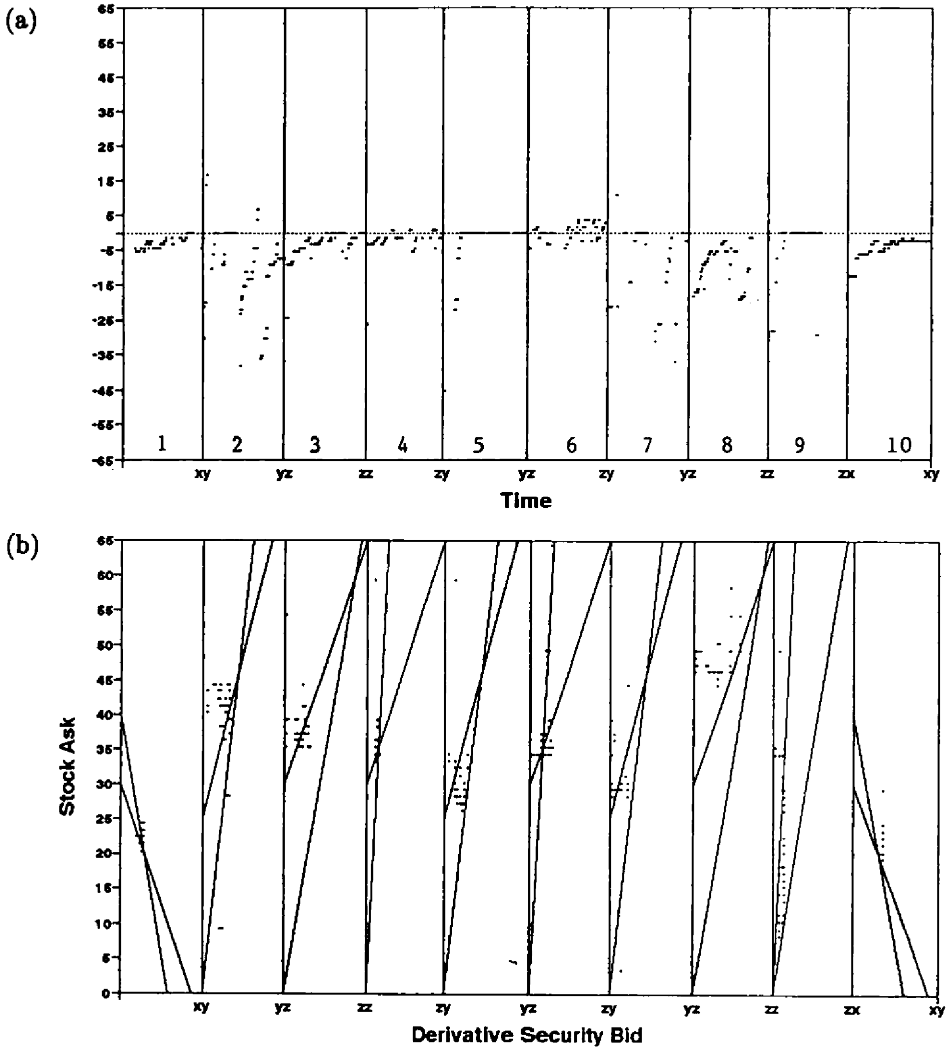


FIGURE 7 (a) Dynamic classification by region: Market A3, Asks.
(b) Arbitrage persistence: Market A3.

3. If $S^b + P^b - 30 > 0$ and $S^b + 4P^b - 45 > 0$, then no matter who is making the market, they are knowingly giving up arbitrage. In this case we plot the point closer to zero, which represents the least amount of arbitrage given up.

To summarize, points above zero indicate arbitrage, zero means mutually beneficial arbitrage, and negative points indicate nonexistence of any type of arbitrage opportunity. Since the points are plotted in real time every time some market activity takes place, we get a dynamic view of the persistence of arbitrage.

The second graph has the option price on the x-axis and the stock price on the y-axis, so it contains the same information as in Figure 1. In each trial, there are two lines, corresponding to the two trader types, and we plot the stock bid and corresponding option price (which could be a bid or ask, depending on the portfolio). This graph thus shows how the stock bid/option price relate to the different regions given by the arbitrage free lines. Note that this graph does *not* provide a dynamic picture; it is possible, for instance, for there to be persistent arbitrage but for only one point to be plotted in the graph.

The other figures have the same information for the ask side and across markets. We have normalized the first graph on the ask side so that it has the same orientation as the bid graph, i.e., a positive point implies someone is knowingly giving up arbitrage.

The first graph in each figure clearly shows that virtually every time the requisite bids and asks were present, there was some trader type for whom it was rational to make the offer. This is true in all six figures. Exceptions are trial 1 in A1 on the bid side, trial 3 in A1 on the ask side, and trial 6 in A3 on the ask side.

In all three markets, we note that the bid-ask spread narrowed over time. This is certainly more observable in A3, where we see many more points plotted. However, in both A1 and A2, the market ended with either the bid or the ask quite far from zero in several trials. (In A1 these included trials 2, 3, 7, and 10 on the bid side, and trials 1, 3, 4, 6, 8, and 9 on the ask side, with no observations on the ask side in trial 2. In A2, these included trials 1, 5, 8, and 9 on the bid side, and trials 1, 3, 4, 8, and 10 on the ask side.) By contrast, we have a liquidity problem in A3 only in trial 2 on both sides. Note also that the spread narrows early in the period in A3 relative to Al and A2.

Mutually beneficial and persistent arbitrage in markets Al and A2 occurred infrequently. If we look at whether we get a series of dots plotted at zero toward the end of the period, out of the 40 possibilities, we only observe persistence in three trials: A1 (bid) trial 9, A2 (bid) trials 3, 4. In market A3, which has our liquidity manipulation, we find evidence of persistence in trials 3, 7, and 8 on the bid side. In trial 9, there was persistence in the first half of the period, and then the market moved away from this region. On the ask side, we find persistence in trials 3, 5, and 9. The market heads for this region in trial 1, and there was frequent mutually beneficial arbitrage in the first half of the period in trials 2 and 7.

Finally, we examine the various bid-ask combinations in terms of the three categories described in the introduction. We classified behavior into three types: C (for convergence), P (for persistent arbitrage), and S (for stuck away from the REE without persistent arbitrage due to a lack of liquidity).

Consider market A1. We obtain C in all trials except 4, 9, and 10. Figures 2 and 3 show that trials 4 and 10 correspond to S, while trail 9 is clearly a P.

In market A2, we have convergence in all trials except 1, 3, 4, 5, 8, 9, and 10. Figures 2 and 3 show that the behavior corresponds to S in trials 1, 5, and 8, and to P in trial 3. It tends toward situation P in trial 4. Trial 9 is interesting in that it is in S but tends toward the REE with some arbitrage showing on the ask side toward the end of the trial.

In market A3, we do not observe S. Lack of convergence is observed in trials 3, 5, 7, 8, and 9. Every one of them corresponds to P. Note that we have many more trials with persistence in A3 than in either Al or A2.

5. SUMMARY AND CONCLUSIONS

Perhaps the main finding of this paper is that we rarely observe prices in which some trader type is giving up arbitrage, and when the market enters a persistence region, it has a tendency to stay there. As we noted earlier, this is perhaps the result of the rules of the double auction. We find this result very interesting in that it means that despite the theoretical relationship between arbitrage and efficiency, market dynamics *do not* appear to be characterized by arbitrage opportunities disappearing over the time interval of our typical experiment. On the other hand, we found that by increasing the incentive for making markets, the spread in the separating portfolios became narrow quite early on in every trial but that instead of leading to convergence, persistence of arbitrage actually increased.

The classification of data we have provided points to three immediate directions which could be fruitfully explored. First, the fact that we can have persistent arbitrage even in a complete market is due to the fact that different trader types can be making different sides of the market. This suggests that one way to eliminate informational inefficiency is to move to a market system in which a single trader type (such as a specialist) sets both the bids and the asks. Second, it is clear that liquidity across markets is critical, and we need to think carefully about the effect of alternative market institutions on spreads across markets. The notion of separating portfolios provides a direct way to study the effects of alternative institutions on liquidity. Finally, one could see whether in markets such as A3, persistence dies out when the time period of the experiment is increased, due to the imbalance between supply and demand in the persistence region.

ACKNOWLEDGMENTS

We thank the participants, especially Colin Camerer, Dan Friedman, Charlie Plott, John Rust, and Bob Wilson, for valuable discussions and comments.

REFERENCES

1. Berg, J. A., L. A. Daley, J. W. Dickaut, and J. R. O'Brien. "Controlling Preferences on Units of Experimental Exchange." *Quart. J. Econ.* **101** (1986): 281–306.
2. O'Brien, J. R., and S. Srivastava. "Arbitrage and Informational Efficiency: Theory and Experimental Evidence." Technical Paper, Carnegie-Mellon University, 1990.
3. Roth, A. E., and M. W. K. Malouf. "Game Theoretic Models and the Role of Information in Bargaining." *Psych. Rev.* **86** (1979): 574–594.

Index

D

E

J

K

L

M

T

U

V

W

Z