

Securities Trading of Concepts (STOC)

[*Running title*: SECURITIES TRADING OF CONCEPTS (STOC)]

Ely Dahan^{*}, Adlar J. Kim^{**}, Andrew W. Lo^{***}, Tomaso Poggio^{*} and Nicholas Chan^{**}

September 9, 2009

* (*corresponding author*) UCLA, 110 Westwood Plaza B-514, Los Angeles, CA 90095-1481, edahan@ucla.edu

** Postdoctoral Associate, MIT Sloan School of Management, MIT, Cambridge, MA 02142

*** Harris & Harris Group Professor of Finance; Director, MIT Laboratory for Financial Engineering, MIT Sloan School of Management, Cambridge, MA 02142

• Eugene McDermott Professor in the Brain Sciences and Human Behavior, McGovern Institute, Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA 02139

** Two Sigma Investments, LLC, New York, NY 10012

This report describes research done within the UCLA Marketing Research Center, the Center for Biological and Computational Learning in the Department of Brain and Cognitive Sciences, the Artificial Intelligence Laboratory, and the Laboratory for Financial Engineering at the Massachusetts Institute of Technology. This research was sponsored by grants from: Office of Naval Research under contract No. N00014-93-1-3085, Office of Naval Research (DARPA) under contract No. N00014-00-1-0907, National Science Foundation (ITR) under contract No. IIS-0085836, National Science Foundation (KDI) under contract No. DMS-9872936, and National Science Foundation under contract No. IIS-9800032 This research was partially funded by the UCLA Marketing Research Center, the Center for e-Business (MIT), and the MIT Laboratory for Financial Engineering. Additional support was provided by Central Research Institute of Electric Power Industry, Eastman Kodak Company, DaimlerChrysler AG, Compaq, Honda R&D Co., Ltd., Komatsu Ltd., Merrill Lynch, NEC Fund, Nippon Telegraph & Telephone, Siemens Corporate Research, Inc., and The Whitaker Foundation. The authors also wish to thank Rob Hardy and Leonard Lee of MIT, and Limor Weisberg for their efforts in programming and designing many of the web sites that comprise this research. We thank Professors Hyun Shin, Robert Zeithammer, and Andrew Ainslie for their helpful modeling insights and Jeremy Dann and Craig Boreth for their editorial suggestions.

Securities Trading of Concepts (STOC)

Abstract

Identifying winning new product concepts requires insight into consumer preferences, which represent private information held by each consumer. Market prices are well known to efficiently collect and aggregate private information regarding the economic value of goods, services, and firms, particularly when trading financial securities. We apply the price discovery mechanism in pseudo-securities markets to measure consumer preferences for new product concepts. This is the first application of such markets to test potential new product concepts and to compare such an approach against stated-choice, conjoint and longitudinal revealed preference data. We address the challenge of validating simulated market results in which actual outcomes cannot be observed. A securities-trading approach may yield significant advantages over traditional methods - such as surveys, focus groups, concept tests, and conjoint analysis studies - for measuring consumer preferences. These traditional methodologies can be more costly to implement, more time-consuming, and susceptible to potential bias. Our approach differs from prior research on prediction markets and experimental economics in that we do not require any exogenous, objective “truth” such as election outcomes or movie box office receipts on which to base our securities market. We also differ by demonstrating that in this context, metrics summarizing all prior trades are more informative than closing prices alone. In fact, STOC markets are shown to resemble traditional market research more than they resemble prediction markets. We show that STOC trading reveals heterogeneous preferences at the individual level.

As a measure of internal validity, four product categories are tested in eleven independent STOC markets. In the context of new product development, exogenous truth may not be available as the majority of potential new product concepts are never launched, and actual demand may never be revealed for discarded concepts. To address the need for external validity we compare STOC trading results against preferences measured through: (1) virtual concept testing (of bicycle pumps and crossover vehicles), (2) stated-choices (of actual crossover vehicles and Wii video game concepts) and (3) actual sales of the subset of product concepts that *are* launched in a simulated store (laptop bags) and in the real marketplace (crossover vehicles), (4) surveys of individuals’ expectations of others’ preferences and (5) full-profile conjoint analysis (of bike pumps and Wii video games). These experiments reveal that the market prices of securities designed to represent product concepts are remarkably efficient, accurate, and internally consistent measures of expected market share based on group preferences, even when conducted with relatively few traders. We also note that while STOC prices measure preferences reasonably well, they do not necessarily predict actual sales. Because the number of stocks tested can scale in the number of traders, the STOC method is particularly efficient at screening promising new products and services from a large universe of possibilities. For new product development (NPD) teams deciding where to invest product-development resources, this scalability may be especially important in the Web 2.0 world in which customers interact with firms and with each other in suggesting numerous product design possibilities.

1 Introduction

The need for efficient and reliable methods of screening new product concepts for their marketability continues to grow in the context of Web 2.0's voluminous interactions between firms and customers. Now that customers are suggesting possible product concepts, the number of concepts to be market tested far exceeds firms' capacity to test them. Markets are well-known to be an efficient tool for collecting and aggregating diverse information regarding the value of commodities and assets (Hayek 1945), particularly in the domain of financial securities.

In this paper, we explore a novel application of the price-discovery mechanism of financial markets to measuring preferences using securities trading of concepts (STOC) to collect consumer preferences on product concepts. Our research explores four key issues about the STOC method: (1) How well does it measure aggregate preferences for new product concepts? (2) How does it compare to other concept testing methods on accuracy and cost? (3) How and why does the method work? How does STOC relate to prior theories of finance and consumer preference? (4) Is it limited to aggregate measurement, or can individual preference heterogeneity be detected? If so, how do individual participants make their trading decisions?

Briefly, our answers are (1) STOC measures aggregate preferences extremely well; (2) it is more cost efficient than most competing methods, especially regarding respondent recruiting and compensation; (3) the method works by using the pricing mechanism to aggregate information about traders' expectations of others preferences, which are heavily influenced by traders' self preferences. Stock prices are seen to be stationary (unlike "real" financial markets) and there are elements of both finance theory and consumer preference at work here; and (4) trader heterogeneity is evident in the data and can reveal individual preferences to some extent.

This application is motivated by the need for reliable, accurate, fast and economical means to gauge consumer preferences during new product development. It relies on the belief that markets are efficient in aggregating privately held information such as individual preferences and expectations of others' preferences. It exploits the incentive-compatible nature of markets, i.e. the fact that over- or undervaluing securities reduces the participants' rewards, and the fact that most participants prefer competing in games to responding to surveys.

Regarding the cost efficiency of any concept testing method, consider the total costs of recruiting and then compensating a sample of N respondents willing to share their preferences about M concepts. The firm must recruit more than N subjects as response rates are likely to be below 100%, and the costs per person for recruiting and compensation may vary based on the "attractiveness" of the preference measurement method being employed. The total costs can be summarized as follows (derived in Appendix 1):

$$(1) \quad TC = \left(M \cdot \frac{N_{sample}}{q_{respondent}} \right) \times \left(\frac{c_{recruit}}{r\%} + c_{respondent} \right).$$

The five key factors influencing the total respondent cost of measuring preferences for M given concepts, and the challenge underlying each of these factors is summarized in Table 1.

Table 1: Five Routes to Lower Respondent Costs in Concept Testing

Symbol	Description	Challenge	To Lower Cost	STOC's potential Benefit
N_{sample}	Number of required Respondents	Statistical power	Reduce N_{sample}	Lower due to <i>interactions</i> with others and <i>multiple answers</i> per respondent
$q_{respondent}$	question capacity per respondent	Bounded rationality	Increase $q_{respondent}$	Higher due to <i>motivation</i> and ability to self-select questions to be traded
$c_{recruit}$	cost to recruit people	People avoid surveys	Reduce $c_{recruit}$	Lower because recruits are <i>attracted</i> to playing the game, even multiple times
$c_{respondent}$	compensation for respondents	People value their time	Reduce $c_{respondent}$	Lower due to the intrinsic pleasure of playing the <i>game</i> itself
$r\%$	response rate	Many people opt out	Increase $r\%$	Higher due to <i>intrinsic</i> pleasure of game, desire to play again, <i>competitiveness</i>

All five of STOC's potential benefits are realized in our testing. Traditional market research typically requires respondent sample sizes of 50–300, or more, and most concept tests measure preferences for 5–12 concepts at a time, so 4–50 respondents per concept tested is typical, whereas our STOC research demonstrates that 1–2 respondents per concept stock is sufficient. This is due to the fact that each respondent rates each concept multiple times (i.e. through multiple trades), traders focus on more concepts due to the competitive nature of the game, and learning amongst traders takes place. Since respondents prefer playing STOC over completing surveys, the amount required to recruit potential respondents and compensate those completing the exercise can typically be reduced by half or more. Thus, STOC could potentially reduce the total respondent costs of a concept preference study by more than 75%. Since the infrastructure costs of running the STOC trading system are essentially fixed, we would expect long term cost advantages of this method.

To determine if the STOC methodology is as accurate as other approaches, we present results for eleven market experiments in four product categories—bicycle pumps, laptop computer messenger bags, crossover vehicles and Wii video games. The resulting securities prices, normalized to sum to 100 in each experiment, are compared against other methods where preferences are also normalized to sum to 100. The other methods are applied to the same product concepts within each product category using stated-choice and revealed preference approaches: rank-ordered choice, conjoint analysis, the virtual concept testing (VCT) methodology developed by Dahan and Srinivasan (2000), actual purchase decisions under controlled conditions and in actual product markets, and constant sum surveys. We find that results across different market experiments in the four product categories are largely reliable across repeated tests and predictive of stated-choice and revealed preference results, but less so

of longitudinal sales data. To gain a better understanding of how the STOC method may be achieving these results, we relate our market experiments with some classic examples from the experimental economics literature.

The STOC methodology centers around virtual stock markets trading hypothetical securities, each associated with an underlying product or service concept. Upon entering a concept market, each participant receives an initial portfolio of cash (virtual or real) and virtual stocks. Participants are also provided with detailed information on the products that includes specifications, images, and multimedia illustrations. A typical objective of the STOC game might be for each participant to maximize the value of his or her portfolio, evaluated at the closing stock prices or the volume-weighted-average prices. Markets are typically open for 20 to 30 minutes and end at random times. If participants play with real money, they will have the opportunity to profit from trading and will conversely bear the risk of losing money. The financial stakes in the game provide incentives for participants to reveal true preferences, process information and conduct research. If fictitious money is used, prizes can be awarded according to individuals' performances. One can also reward all participants simply for their service.

As in real financial markets, stock prices are determined by supply and demand, which depend on participants' evaluation of their own and others' preferences for the underlying products. Thus, at the market equilibrium, prices should fully reflect all participants' aggregate preference of the products. Traders make trading decisions just as they would in a financial stock market: they assess the values of the stocks, sell overvalued ones and buy undervalued ones, essentially voting on the worth of the underlying products. In this way, a stock's price becomes a convenient index of a product's consumer value. While in our STOC tests all trades require a specific buyer to purchase shares from a specific seller, both of whom have placed a

limit order, we could have employed market making software that enables one-sided trades in “thin” markets with few traders as Hanson (2003, 2005, 2009) proposes.

There are, of course, several well-established methods for estimating consumer preferences, e.g., surveys (cf. Burchill and Brodie 1997), voice-of-the-customer methods (cf. Griffin and Hauser 1993), conjoint analysis (cf. Srinivasan and Shocker 1973, Green and Wind 1981, Green and Srinivasan 1990), concept tests (cf. Urban, Hauser and Roberts 1990, Dahan and Srinivasan 2001, Dahan and Hauser 2002), and focus groups (cf. Mahajan and Wind 1992, Calder 1977, Fern 1982). However, concept markets may be a useful alternative to these methods for several reasons:

1. **Accuracy:** In order to win the game, participants have the incentive to trade according to the best, most up-to-date knowledge because of their financial stake in the market. STOC also captures, continuously, the ever changing “pulse of the market” for all participants since they can express their opinions multiple times during the course of the market rather than responding only once to a survey question. Moreover, the continuous nature of financial trading allow participants to express intensity of preference in ways that surveys simply cannot capture, e.g., by committing large amounts of capital to specific trades that represent strong preferences.
2. **Interactive Learning:** A STOC market participant not only evaluates concepts on his or her own behalf, but also considers the opinions of the public at large. Furthermore, participants can observe others’ valuations of the virtual products and update and adjust their own valuations dynamically in the market environment. In short, learning is a crucial element in these markets.
3. **Scalability:** Unlike surveys, in which the number of questions asked is limited by the capacity of each respondent to answer, markets are intrinsically scalable due to the fact that each trader need only evaluate a small subset of the universe of securities. In fact, the efficiency of the market, and therefore the quality of data collected, improves with the number of participants. This extends to the number of product concepts that may be evaluated—since there is no requirement that each respondent trade every security, the bounded rationality of the traders does not limit the number of concepts that can be evaluated in a STOC market.
4. **Integrated Product Concepts:** The STOC method is particularly useful relative to conjoint methods when a product cannot be easily quantified, delineated or represented by a set of attributes (for example, a movie script, fashion item, car body style or piece of art). Market participants evaluate the concepts directly and market prices effectively reflect the overall viability of the concepts, including the ability of a concept to fulfill unarticulated needs. All that is required is a thorough depiction of each concept.

Of course, market-based methods for eliciting information also have certain limitations. Unlike typical marketing research techniques in which information is collected from individuals and aggregated in subsequent analysis, the market method focuses on aggregate beliefs and preferences. While individual heterogeneity is not easily observed from market prices, it does enter the trading process in the form of differences in security valuation. Later we show that some of this heterogeneity may be captured by observing the trading behavior of each individual and the biases these trades reveal. Virtual concepts markets may be vulnerable to price manipulations and speculative bubbles because the values of virtual securities hinge on the aggregate beliefs, which are endogenously determined within the same market. Traders may form false beliefs that could cause prices to deviate from their fundamentals. And all of the behavioral critiques that have been leveled against the Efficient Markets Hypothesis in the financial economics literature (see, for example, Shefrin, 2005) apply to concepts markets as well. For these reasons, the market method must be applied with caution, and the consistency of the results must be checked through repeated STOC markets or other means of validation. The greatest level of vulnerability may occur when traders have a poor sense of their own preferences or of those of other people. This might occur, for example, when the product category is too new for traders to grasp, or when the stimuli shown prior to trading are unclear or confusing (as we demonstrate in one instance shortly). Another drawback is that STOC publicly exposes the tested concepts to a group of respondents, which may be problematic in the case of sensitive, proprietary intellectual property. While other concept testing methods have similar problems, the interactive nature of trading may make it more difficult to contain the problem of information leaks.

A number of practical issues arise in attempting to infer consumer preferences via the STOC method. For example:

- How many traders are needed?
- How knowledgeable does each participant need to be of the product category and concepts being studied? Of the target market?
- Do they need to be experienced at trading securities?
- What strategy do traders adopt in order to “win” the game?
- Are traders driven more by objectivity or personal biases?
- For how long must trading proceed in order to collect sufficient data?
- What, exactly, is being measured by STOC?

The present research attempts to answer many of these questions by positioning STOC in the context of prior experimental economics and prediction markets research, and by evaluating the results of empirical experiments in three product categories.

In Section 2, we outline the market research alternatives to STOC which also measure preferences for product concepts. These include virtual concept tests of bike pumps, a simulated store selling laptop bags and stated choice surveys, longitudinal revealed preference sales data for crossover vehicles, constant sum surveys of purchase intent for Wii video game concepts, and conjoint analysis for both bike pumps and Wii video games. We then summarize relevant research from the prediction markets and experimental economics literature. Section 3 introduces the four product categories that are tested in this research, the designs of the securities representing product concepts in those three categories, and the market mechanism used to trade these securities. In Section 3.1 we conjecture on how STOC works by considering alternative strategies that traders might employ, and show how stock prices capture consensus preferences. Section 4 presents aggregate and individual results from multiple STOC experiments, develops a taxonomy of internal and external validity testing, introduces the trader bias metric, and

compares individual heterogeneity in trading strategies against individual preference. We conclude in Section 5 with a brief discussion, possible extensions and limitations.

2 Background

To validate the STOC method, we compare it against alternative methods of measuring preferences for new product concepts. We also position STOC in the context of prior work on prediction markets and experimental economics.

2.1 Prior methods of measuring product concept preferences

Concept testing enables new product development teams to identify those concepts most preferred by consumers. Dahan and Srinivasan (2000) present a virtual concept testing (VCT) methodology employing the Internet with virtual prototypes in place of real, physical ones and compare these against conjoint analysis. Their studies identify the most preferred of nine bicycle bump concepts versus the two commercially available products depicted in Figure 2.

The authors find that static images of the bike pumps on the Internet produce market share predictions that closely resemble those for real physical prototypes examined in person. We employ their physical prototype, conjoint, and static web virtual concept test results for bicycle pumps in hopes of validating the STOC method. Both bicycle pump STOC tests, conducted on the other side of the country and six years after Dahan and Srinivasan collected their data, were conducted with the *same* group of traders as a method of confirming test-to-test reproducibility as well as external validity.

Dahan and Hauser (2002) add multiple web-based market research methods to the mix, applying them to the eight existing and yet-to-be-released crossover vehicles depicted in Figure 3. They also demonstrate a high degree of correlation between web-based concept testing and

respondents' self-stated-choices as measured by simple surveys. We test STOC in four independent crossover vehicle experiments and compare our results against self-stated data in three of them and against virtual concept testing (VCT) in all four. We estimate VCT preferences at the individual level in two ways: including both product preferences and vehicle prices to determine each trader's utility score, and utilizing product preferences alone, excluding the effect of vehicle prices. In each case, we then aggregate individual preferences to generate market share estimates.

Toubia, et. al. (2003) develop a new polyhedral adaptive approach to conjoint analysis, and test it against existing adaptive and static conjoint methods using the example of customizable laptop PC messenger bags sold for real money through a simulated store. Their work demonstrates the effectiveness of their method, but more importantly for the present research offers an excellent data set for validating STOC. We focus on eight randomly chosen bags, representing a range of popularity (market share) actually sold to 43% of the respondents in their research. Two STOC tests were run to measure preferences for the same eight bags, but utilizing two different forms of stimuli: the table shown in Figure 4 and the individual images shown in Figure 5.

Additionally, in the six years following the crossover vehicle STOC tests, that is from 2001–2006, we also collected unit sales data for each of the eight vehicles from *Ward's Automotive News*. These data are used as a test of external validity and the predictive power of the STOC method.

We are grateful for the cooperation of the aforementioned researchers who enabled us to adopt the identical product concept illustrations in our STOC tests. Thus, we are able to compare

results for identical market research problems using STOC versus each of the prior methods. We attempt to validate our method in the eight STOC trading tests conducted from 2000 to 2009, as summarized in Table 2. Traders in nine of the tests were MBA students, but additional tests included management level attendees from the MIT Center for Electronic Business conference (crossover vehicle test 3) and senior executives attending executive education classes (crossover vehicle test 4). All eleven tests were run under controlled conditions in a business school trading laboratory or classroom.

Table 2: Data Collected for each of Four Product Categories

Method <i>Product type</i>	Experiment	STOC Method	Conjoint Analysis	Virtual Concept Test	Self-Styled Choices	Simulated Store	Longitudinal Sales Data
<i>Bike Pump Concepts</i>	Tests 1 & 2 $n = 28$	9 Pumps; Same traders tested twice	Rank 18 full profiles, est. 10 parameters w/LINMAP $n = 141$	Dahan and Srinivasan '00 VCT Physical, VCT Web; $n = 102, 87$			
<i>Actual Laptop Bags</i>	Test 1 $n = 50$	Table of 8 Laptop Bags				Toubia, et. al. 2003 unit shares for 8 bags sold in the simulated store; $n = 143$	
	Test 2 $n = 62$	Images of 8 Laptop Bags					
<i>Actual Crossover Vehicles</i>	Test 1 $n = 49$	8 vehicles <i>No Prices</i>		VCT with and without Prices	Top 3 of 8 with prices		Cumulative units sold for each of 8 vehicles from 2001-2006 per <i>Ward's Automotive News</i>
	Test 2 $n = 43$	8 vehicles <i>No Prices</i>		VCT with and without Prices	Top 3 of 8 with prices		
	Test 3 $n = 42$	8 vehicles <i>With Prices</i>		VCT with and without Prices	Top 3 of 8 with prices		
	Test 4 $n = 16$	8 vehicles <i>No Prices</i>		VCT with and without Prices			
<i>Wii Video Game Concepts</i>	Test 1 $n = 35$	8 Own Wii Video Games	Rank 16 full profiles, est. 10 parameters w/ LINMAP $n = 35 \& 65$	Constant Sum Allocation of 100 Points across 8 or 11 Wii Games in (4) Surveys:			
	Test 2 $n = 55$	8 Others' Wii Video Games		<ul style="list-style-type: none"> • SELF Preferences • E[Others' Preferences] • E[STOC prices] • E[Actual Share] after STOC game 			
	Test 3 $n = 58$	11 Own Wii Video Games					

For the first eight STOC experiments, traders were told that their final portfolio valuations would be based on the closing prices of each stock plus the cash they had left on hand. For the final three experiments testing Wii video game concepts, traders were told that final portfolio valuations would be based *not* on closing prices, but rather on *mean* prices throughout the trading

period (i.e. using volume-weighted-average-prices). For each of the eleven (11) tests above, Figure 9's STOC user interface was employed, and each test ran in under one hour including instructions and wrap up.

2.2 Prediction Markets

Non-financial "prediction markets" have been established for political elections, movie box office estimation, and other real world outcomes. The Iowa Electronic Markets (IEM)¹ pioneered prediction markets for the purpose of forecasting election results (Forsythe, Nelson, Neumann & Wright 1993). The IEM was founded for research and educational purposes. Trading profits from the market provide incentives for traders to collect and process information about future events. The IEM features real-money futures markets in which contract payoffs depend on the outcome of political and economic events. IEM's predictions have outperformed most national polls.² Similarly, the Hollywood Stock Exchange, HSX.com, has provided accurate predictions of movie box office results (Spann and Skiera 2003). The Foresight Exchange (FX)³ predicts the probability of future events occurring such as changes in the environment, scientific breakthroughs, the collapse of companies, or political and news outcomes. Companies such as Hewlett Packard, Microsoft, Best Buy and Google have employed internal prediction markets to forecast printer sales, software release dates, consumer electronics sales, and software take-up rates.

Prediction markets share with STOC the benefits of information aggregation, the joy of competitive play, the ability to learn from others, and the incentive to be accurate. Prediction

¹ The Iowa Electronic Markets, <http://www.biz.uiowa.edu/iem/>

² *BusinessWeek*, 11/11/96

³ The Foresight Exchange, <http://www.ideosphere.com/fx>

markets focus on actual outcomes, operate for weeks, months, and sometimes years, and incorporate private information and news as it happens. STOC markets, in contrast, focus on concepts that may never come into existence, and therefore may never have actual outcomes, run for 10–60 minutes typically, and are not influenced by outside news. In fact, the only information available to STOC traders is the personal preferences they hold, their expectations of others' preferences, and whatever they learn by observing price movements during trading.

2.3 Rational Expectations (RE) Models and Experimental Markets

Our trading experiments are closely related to the literatures in rational expectations (RE) models with asymmetric information and experimental markets. In a standard asymmetric information RE model (Grossman, 1981), heterogeneous agents with diverse information trade with each other and, under certain conditions, the market will converge to an equilibrium in which prices fully reveal all relevant information. The most important criterion for convergence is that agents condition their beliefs on market information. In particular, agents make inferences from market prices and quantities about other agents' private information.

The RE model has received considerable attention in the study of experimental markets (Plott and Sunder, 1982, 1988; Forsythe and Lundholm, 1990; Davis and Holt, 1993). Studies of the informational efficiency of a market relative to the RE benchmark fall into two categories: markets with fully informed agents (“insiders”) and uninformed agents, and markets with many partially informed agents. In various experimental markets with human subjects, the results for both market structures are the same: markets eventually converge to the RE equilibrium, i.e., information aggregation and dissemination occur successfully.

STOC trading shares some characteristics with such experimental economics markets, and information aggregation and dissemination provide compelling explanation for the success of our experiments. For example, traders who possess superior information about the products, or have high confidence in their beliefs, can be considered “insiders.” On the other hand, traders who have little knowledge or opinion of the products can be regarded as the “uninformed.” The interaction between the insider and uninformed constitutes information dissemination. What is intriguing about this scenario is that even when a subset of traders ignores the underlying product information and only focuses on market information, the market still converges to efficient prices that aggregate all the relevant information and beliefs. A striking example of just how informationally efficient financial markets can be is provided by Maloney and Mulherin (2003), who document the fact that in the wake of the Space Shuttle Challenger’s explosion in 1986, the stock price of Morton Thiokol—the vendor who was ultimately held responsible for the infamous failed O-ring in the booster rocket—dropped precipitously within minutes after the explosion, and much more so than any other vendors’ stock prices during the same period.

Alternatively, individual traders may form their own beliefs about the products, acknowledging that market prices will depend on aggregate beliefs. This is similar to the information aggregation scenario in which there are no “insiders”, but where all traders are partially informed. Even in this case, where no single trader has full information, an RE equilibrium will be reached under very general conditions (Grossman, 1981; Davis and Holt, 1993, Chapter 7).

However, there is one important difference between STOC markets and the other exchanges in the experimental markets literature. In a typical experimental market, subjects’ preferences and their information set are fixed and assigned by the researchers. Therefore, even

before trading begins, theoretical equilibrium prices can be calculated. In contrast, in a STOC market, neither the subjects’ preferences nor their information sets are known—in fact, these are what STOC market trading experiments are meant to discover. This suggests an important practical consideration in implementing STOC markets: the composition of traders should match the population of target consumers as closely as possible, or at least include traders with insight into the preferences of these consumers. For example, if the target population for a particular product is teenage female consumers, a STOC market consisting of middle-age males may not yield particularly useful preference rankings for that product. However, if the cross section of traders in a STOC market is representative of the target population, the force of market rationality will ensure that the price-discovery mechanism will provide an accurate measure of aggregate preferences.

Table 3: Five Key Steps to Designing and Executing a STOC Experiment

Step	Description	Challenges	Key Considerations
1	Choose STOC Concepts	Narrowing from many options	Stocks should clearly and concisely depict multiple product concepts that differ from each other. Not every trader has to see every stock.
2	Define STOC Prices	Open-ended vs. precise definition	Traders need to understand the definition of each stock, e.g. “the % of people who prefer this concept” or “market share of this product”
3	Define & Teach Trading Method	Programming & User Interface	The user interface should be easy-to-use, informative about the trading activity for each security and trader performance
4	Trading & Data Collection	Need Simultaneous trading; Trader Errors	Transaction details between any two traders needs to be recorded: security name, volume, price, timing. Traders should be able to review, edit and cancel open orders that have not cleared.
5	Data Analysis	Choosing a metric; What is measured?	The metric should include all information such as the number of shares traded and at which price (not just closing prices)

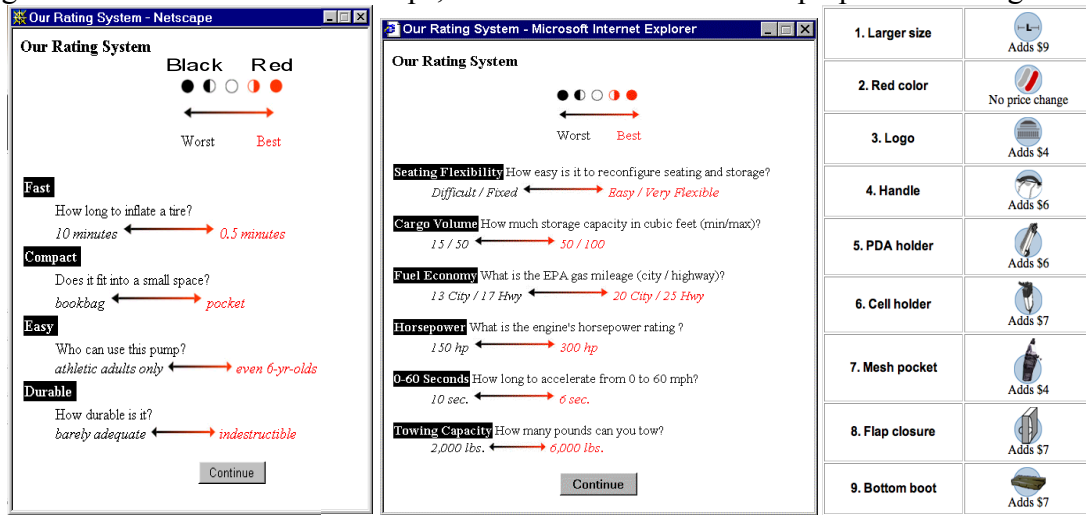
3 Design of Markets and Securities

There are many considerations when designing a STOC experiment to measure preferences for product concepts, five of which are summarized in Table 3 above.

In answer to the design questions raised earlier in Section 1, we find that approximately 1–2 traders per concept are sufficient for a successful STOC experiment, but additional traders improve accuracy and stock price convergence. While participants need not have strong personal motivations to buy within the product category, it does improve results if they at least have insights about how others in the target market make product choices. Participants need not have prior familiarity with securities trading as the instructions are relatively simple, but prior experience tends to improve individual performance in the game. We shall empirically show that traders base much of their trading strategy on expectations of others that they form prior to the start of trading, and that these individual expectations are highly correlated to each person’s personal preferences. In that sense, individual traders are driven more by subjectivity than objectivity—as they should be, given our ultimate objectives—which is different than in financial markets where outside news and facts are plentiful. Trading must proceed long enough to generate a statistically significant number of trades for each security, which we have observed to be between 15 to 45 minutes. What appears to be measured by STOC is the aggregate expectation of others’ preferences across all traders in that experiment. Since expectations of others are strongly correlated with *self* preferences, aggregate expectations of others’ preferences nicely summarize aggregate *individual* preferences. STOC trading behavior can also be shown to capture individual preference heterogeneity, although not as richly as more individualized instruments such as conjoint analysis or surveys.

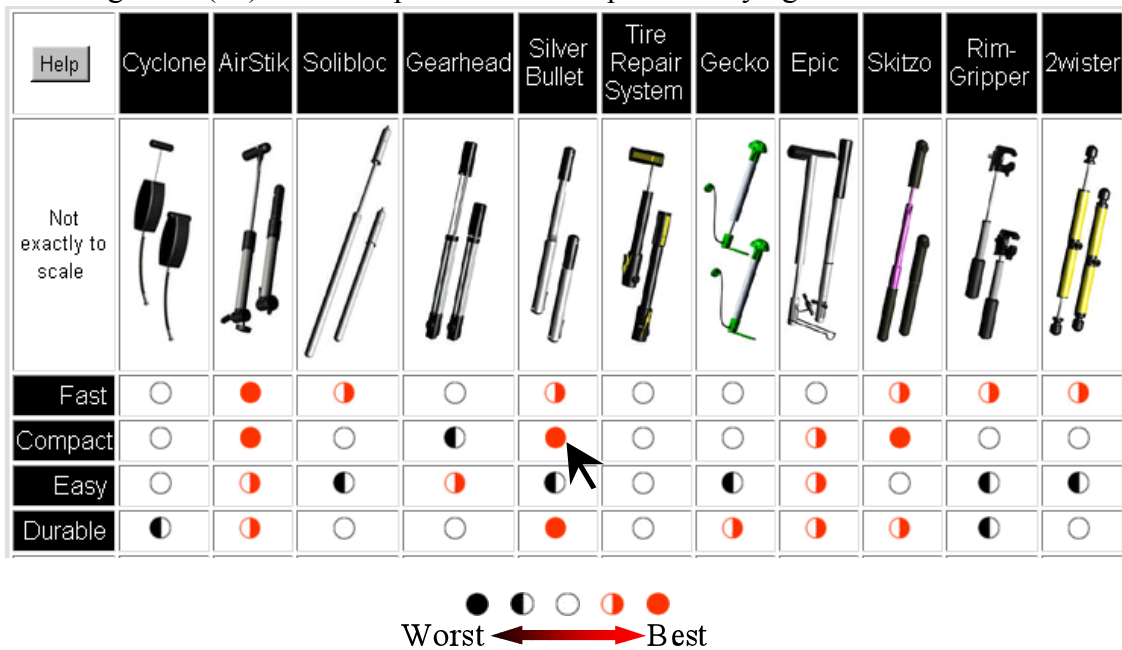
While presenting full product concepts, we also educate traders about the product attributes and attribute levels using Consumer Reports-style ratings as in Figure 1.

Figure 1: Attributes for Bike Pumps, Crossover Vehicles and Laptop PC Messenger Bags



In our eleven tests, the STOC method is applied to the product concepts described in Dahan and Srinivasan 2000 (Figure 2), Dahan and Hauser 2002 (Figure 3) and Toubia, et. al. 2003 (Figure 4 and Figure 5) and the Wii Video Game concepts developed by MBA teams in 2009 (Figure 6 and Figure 7).

























































Figure 2: (11) Bike Pump Product Concepts Underlying the STOC Securities



In order to anchor the value of the fictitious currency in the case of bike pumps, one of the eleven securities—Cyclone—has its price fixed at \$10 and is not traded. Thus, Cyclone serves as a reference price or numeraire security. For example, if a trader thinks that TRS is worth twice as much as Cyclone, he or she would pay up to \$20 for one share of TRS. The stocks of the ten freely traded concepts may be priced at any level, depending on the supply and demand in the market, i.e. the willingness of at least one trader to buy at the price at which another trader is willing to sell.









The eight crossover vehicles in Figure 3 were tested in 2000 and 2001 and consisted of three already released at the time (Lexus, Mercedes and BMW) and five yet-to-be-released vehicles (Pontiac, Acura, Buick, Audi and Toyota).

Figure 3: (8) Eight Crossover Vehicles

	Pontiac Aztek	Mercedes-Benz ML320	Acura MD-X	Buick Rendezvous	Lexus RX-300	BMW X-5	Audi All-Road	Toyota Highlander
								
Seats	5	5 (7 opt.)	7	7	5	5	5 (7 opt.)	5
Seating Flexibility								
Cargo Volume								
Fuel Economy								
Horsepower								
0-60 acceleration								
Towing Capacity								

The eight laptop PC messenger bags shown in Figure 4 and Figure 5 were part of the controlled study described in Toubia, et. al. (2003) in which 330 first year MBA students were provided cash towards the purchase of a customized laptop PC messenger bag. These eight bags include designs that ranged from low- to medium- to high-popularity amongst the 143 respondents to the original study who bought them from a simulated store with actual cash.

Figure 4: (8) Eight Laptop PC Messenger Bags in a Tabular Format

	Bag 3	Bag 4	Bag 8	Bag 9	Bag 10	Bag 13	Bag 15	Bag 16
								
Price	\$89	\$88	\$99	\$80	\$95	\$79	\$78	\$87
Size	Medium	Large	Large	Medium	Large	Medium	Medium	Large
Appearance	Black	Red & Black	Black	Black	Red & Black	Red & Black	Red & Black	Black
Logo	No	Yes	No	Yes	No	No	No	Yes
Handle	Yes	Yes	Yes	No	No	No	Yes	Yes
PDA Holder	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Cell Phone Holder	No	No	Yes	Yes	Yes	No	No	No
Mesh Pocket	No	Yes	Yes	Yes	No	Yes	Yes	No
Closure for Sleeve	Full Flap	Velcro Tab	Velcro Tab	Velcro Tab	Full Flap	Velcro Tab	Full Flap	Velcro Tab
Boot	Yes	No	Yes	No	Yes	Yes	No	Yes

In the first laptop bag STOC test, traders saw the eight laptop bags in the table shown in Figure 4. In test 2, the eight laptop bags were depicted as simpler images, four of which are reproduced in Figure 5, leaving out the table of product attributes and simply showing nine product attributes visually rather than verbally. The eight laptop PC messenger bags depicted in the two types of experiments are identical.

Figure 5: Laptop PC Messenger Bags Depicted Visually without a Tabular Format



The reason that a new set of stimuli were created to represent the same eight bags, frankly, was that the tabular form of presenting the bags was not well-received nor understood by the traders. In section 5, this will become more apparent when the trading results of the two STOC tests are analyzed.

Our final three STOC games involved (19) nineteen Nintendo Wii video game concepts developed by student teams competing with each other in two different MBA courses. Each video game concept included game play software as well as at least one piece of original hardware, i.e. a peripheral device that would enhance game play and work with the Wii console's motion detecting, force sensing and/or WiFi features.

Each Wii video game concept was also defined by its feature levels for six product attributes that were studied in a conjoint study completed by each of the 90 students as well as outside respondents. The six attributes were: (1) Genre (Perform, Learn, Work Out or Experience), (2) Price (\$69, \$149 or \$299), (3) Number of Simultaneous Players (1 or 2-4), (4) Metacritic score (52, 73, or 94, which measure game play quality on a scale of 0-100), (5) On-line access for downloading content and playing over the web (no or yes), and Energy Level required (Relaxed or Energetic).

Figure 6: Wii Video Game Concepts Developed by 35 students on (8) competing teams

(8) Eight
Student-Developed
Wii Video Games

<p>Wii Wiitches A heart-pumping fantasy game for the thinking man</p> <p>Features:</p> <ul style="list-style-type: none"> 94 Score on Metacritic Broom stick and balance board peripheral Bundle of games including: <ul style="list-style-type: none"> Flying Mind Control Word dashes Priced at \$79 <p>Leave your world behind!</p> <p>\$79.00 94</p>	<p>Strike-a-Pose</p> <p>2+ Players · Perform · Online</p> <p>\$69.00 91</p>	<p>MUSICXPLOSION</p> <p>Experience your favorite music like you've always wanted by transforming your musical energy into dazzling images on screen!</p> <ul style="list-style-type: none"> Hear the song you love while creating an amazing visual display Perform the way you want to Play alone or with your friends Game includes 15 songs and you can upload your own CDs & mp3s Four modes of game play: Abstract, Analogue, Creation and Destruction! <p>\$79.99 92</p>
<p>Wii Mill</p> <p>The WII MILL, taking real gaming experience to the next step, literally!</p> <p>Walk into dark caverns filled with snakes, spiders, and scorpions.</p> <p>Run through caves chased by crocodiles and cannibals.</p> <p>Punch and shoot your way out of death traps and bad guys.</p> <p>Exercise your body and mind through the adventures of Indiana Jones.</p> <p>\$179.00 95</p>	<p>WHUP ASS</p> <p>PERSONALIZE your fight – Upload your friends' & enemies' photos, or sync with Facebook & MySpace</p> <p>New Wii-enabled KICKERS give you extra control of your fighting moves</p> <p>You pick the style of the fight</p> <p>Includes 2 KICKERS</p> <p>\$99.99 80</p>	<p>THE AMAZING RACE THE GAME</p> <p>Travel the world like a millionaire or a backpacker</p> <p>3D, Hi-Def, motion-sensing "GPS" has you the perspective of actually being there</p> <p>Explore the land, learn the native language and dance to the local beats</p> <p>Travel at your own pace or use the multiplayer option to compete against new and old friends alike</p> <p>\$149.99 99</p>
<p>LUSH</p> <p>92 Metacritic</p> <p>Mature Rating</p> <p>Experience the world's most exclusive bars & nightclubs</p> <p>Feel the excitement of performing fair tricks in front of huge crowds</p> <p>Learn how to make real cocktails</p> <p>2 Player Co-op (Local & Online)</p> <p>Up to 4 player Social Interactive Drinking Games (Local & Online)</p> <p>Flip Cup, Beer Pong, Quarters, and more</p> <p>\$69.99 92</p>	<p>Yoga Your Way</p> <p>Virtual Yoga Instructor</p> <p>Personalized & Customizable</p> <p>Audio & Visual Feedback</p> <p>Effectively learn proper form</p> <p>Progress Tracking</p> <p>Strength, Flexibility, or Refutation</p> <p>Single Session & Programs</p> <p>Customizable</p> <p>See if you can pull your shoulders back & more!</p> <p>\$69.99 86</p>	

The concepts proposed by the 35 students competing on 8 teams in an elective course on *New Product Development* are depicted in Figure 6 while those proposed by 55 students comprising 11 teams in a *Product Innovation and Marketing* course are shown in Figure 7.

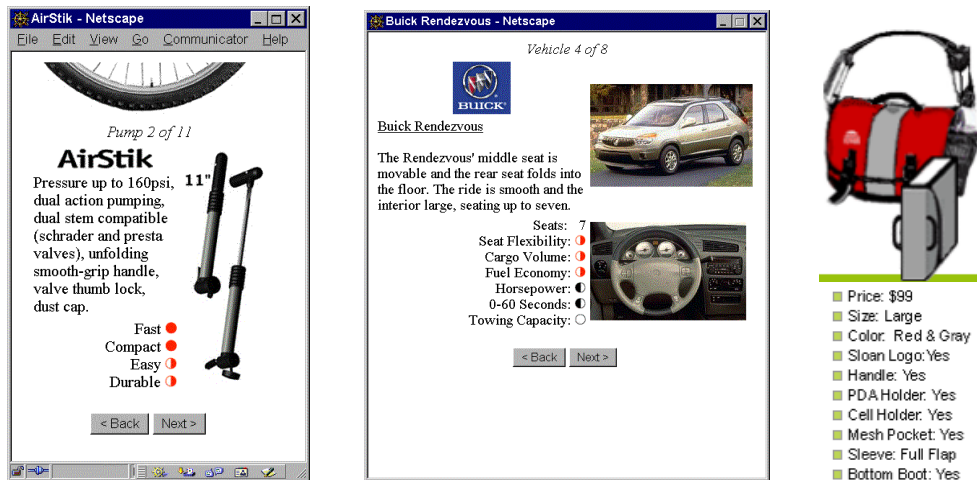
Each stock security represents a particular bike pump, crossover vehicle, laptop PC messenger bag, or Wii Video Game concept. The objective of traders in the STOC game is to maximize the value of one's portfolio. The value of a portfolio is calculated as the sum of the cash on hand plus the total market value of the stocks, as determined by closing market prices or, for the Wii video game concepts, by volume-weighted-average-prices (VWAP).

Figure 7: Wii Video Game Concepts Developed by 55 students on (11) competing teams



Participants strive to maximize their profits by trading the stocks by buying low and selling high. Beyond the stimuli they are shown, the only available information on which to base trading decisions is: (1) their own preferences for the product concepts; (2) their perceptions of others' preferences; and any (3) information they can glean from the trading dynamics of the game. Fictitious money is used in the markets, but top players may be rewarded with actual prizes and recognition by their trading peers. This provides the participants an incentive to perform well in the experiments. In these eleven tests, peer recognition was the primary incentive.

Figure 8: Typical Product Information for Bike Pumps, Crossover Vehicles, and Laptop PC Messenger Bags

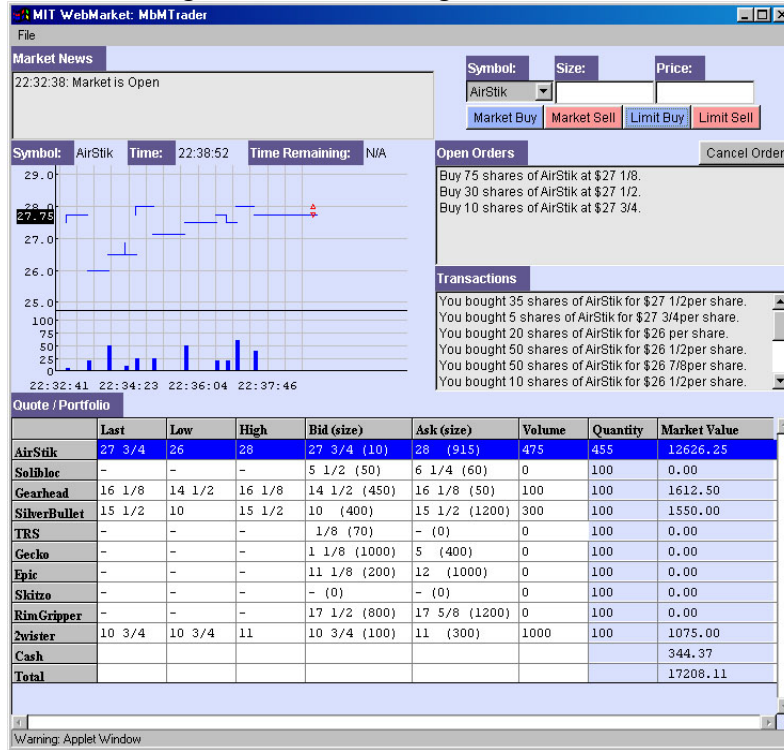


Each trader is provided with an identical portfolio that consists of \$10,000 of cash and 100 shares of each security. No borrowing or short-selling is permitted in the market. Participants trade the securities through the graphical user interface shown in Figure 9.

Market information available to the traders includes the last transaction price and size, current bid/ask prices and sizes, and a historical price and volume chart for each security. A trader can submit either a limit or market order to trade, or cancel an outstanding order that has not been executed. The markets are typical double auction markets with no market-makers. A

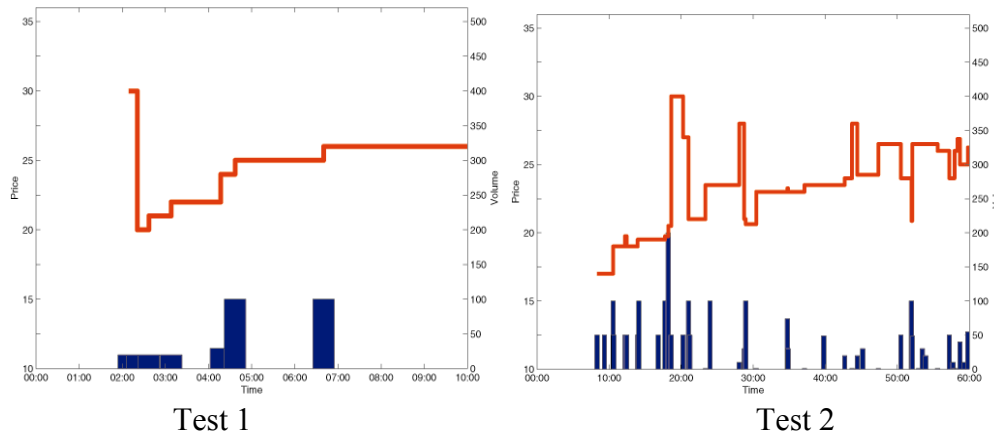
transaction occurs when a market or limit order matches with a limit order on the opposite side of the market. All prices are specified in sixteenths of a dollar.

Figure 9: STOC Trading User Interface



The trading prices (lines) and volumes (bars) for a sample security in the two bike pump tests are shown in Figure 10. We note that price converges to approximately \$25 in both tests.

Figure 10: Price and Volume History of AirStik



For analysis, we focus on the trading data, which consists of the time series of trading prices and volumes $(p_{1,i}, V_{1,i}), (p_{2,i}, V_{2,i}), \dots, (p_{T_i,i}, V_{T_i,i})$, where i is the index for the i^{th} product and T_i is the total number of the cleared trades for the i^{th} product. Our hypothesis is that prices reveal market consensus of relative preferences for each product concept. In particular, we propose that a product's market share can be predicted by its relative STOC price. In addition to the closing price, we consider other metrics that take into account *all* transactions during the session: the high, low, mean, median and volume weighted average prices. The high, low, mean and median prices are calculated from the time series of trade prices $p_{1,i}, p_{2,i}, \dots, p_{T_i,i}$; the volume-weighted average price (VWAP) is computed as follows:

$$\text{VWAP}_i = \frac{\sum_{t=1}^{T_i} p_{t,i} V_{t,i}}{\sum_{t=1}^{T_i} V_{t,i}}$$

The mean, high and low prices are sensitive to outliers—a small number of transactions that occur at extreme prices. All but VWAP ignore the volume in a transaction and treat all trades equally. Volume can be regarded as a measure of the amount of information in a transaction. A trade with higher volume may well be more informative than one with lower volume, since traders are risking more when they trade larger quantities of a stock. In our markets, volume is also related to how confident the traders are and their intensity of preferences at the corresponding transaction price. VWAP effectively summarizes the prices by considering the amount of information and confidence behind the trades. In practice, VWAP has been a widely accepted benchmark price in financial markets. It is a commonly used metric for the evaluation of trade executions. So we might expect VWAP to more fully capture the consensus preferences of the traders.

Given a price statistic \tilde{p}_i , which can be the high, low, closing, mean, median or volume-weighted average prices, we can arbitrarily compute predicted market share as the relative market capitalization, normalized to 100%:

$$MS_i = \frac{\tilde{p}_i E}{\sum_{j=1}^M \tilde{p}_j E} = \frac{\tilde{p}_i}{\sum_{j=1}^M \tilde{p}_j},$$

where M is the number of securities comprising the market and E is the total endowment of shares to all traders for a security, which is the same for each security (since initial endowments of each security are the same). Among the six price statistics, we expect VWAP to be particularly robust against potential price volatility since it includes all information in the market.

3.1 Possible Trading Strategies

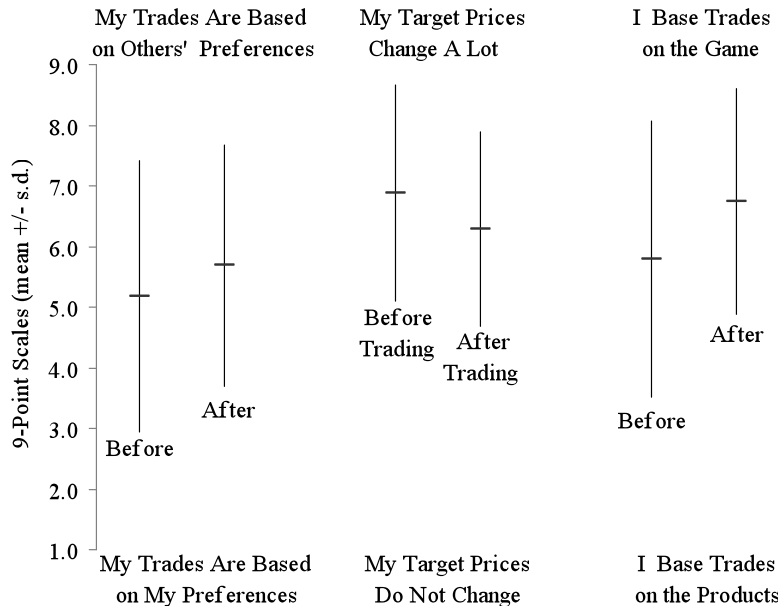
Our market tests are intended to aggregate diverse preferences or beliefs from all traders. We have evidence that the traders' individual preferences prior to the STOC game were quite heterogeneous as reflected in stated-choice surveys and virtual concept tests run prior to the start of each game.

One's beliefs and preferences, and the trading strategy based upon them, derive from:

1. **Product Information.** This is what a participant knows about the underlying products. All participants are provided with the same facts and specifications of the products, but they may have obtained extra product information from their personal experience outside the experiments.
2. **Personal Preferences.** This is what surveys and polls typically collect. Although the aggregate preference of the whole market is the object of interest, one's personal view and biases may contribute to trading decisions.
3. **Assessments of Others' Preferences.** A participant may form opinions and expectations of what others think so as to make profitable trading decisions. This adds a significant element of gaming and competition to the STOC method.

To get a sense of traders' strategies before and after playing the STOC game, we surveyed the 77 traders in two crossover vehicle STOC games, as summarized in Figure 11.

Figure 11: Crossover Vehicle Trader Attitudes Before and After Trading ($n = 77$)



We note that trader attitudes are quite heterogeneous for all three questions, narrowing slightly after trading, but not shifting in a statistically significant way. The picture that forms is that traders' strategies encompass both self preferences and expectations of others' preferences. Traders expect their target prices for buying and selling will vary considerably throughout the game, even though no new outside information is added after the start of trading. We can therefore infer that traders expect to learn from each other through the pricing mechanism. And traders focus slightly more on the gaming aspect of STOC than they do on the product concepts underlying the securities.

How are preferences aggregated in STOC markets? Not only do traders form their own assessment of value, but they also infer the stocks' market value from the market itself. In typical experimental economics markets, both the preferences of the traders and the state of

nature (for example, the probability distribution of a security payoff) are known to the researchers (Plott & Sunder 1982, Plott & Sunder 1988, Forsythe & Lundholm 1990, O'Brien & Srivastava 1991). Traders are assigned preferences that specify securities payoffs in various possible states. The theoretical equilibrium (rational expectations equilibrium) prices can be derived given full information of the markets. The main focus of these experiments is whether and under what conditions rational expectations equilibria can be attained in double auction markets. Some attempts have been made to understand the convergence of prices and how learning occurs in the market as a whole. But it is unclear how individual human traders learn and react to the market. Attempts to model the trading strategies of individual traders from the market data may be overly ambitious. Below we try to shed some light on some possible strategies employed by different types of traders.

The objective of the trading game is to predict the final or mean prices of the securities, trade accordingly, thereby maximizing one's final portfolio value. A trader may form an assessment of the fair values of the securities before trading begins. This assessment may naively take into account only her own preferences for the underlying products, or, if she is more sophisticated, what she perceives as the preferences of others. The trader then bases her trading decisions on her beliefs: she buys undervalued stocks and sells over-valued ones. During the course of trading, she may either maintain her valuations or update her beliefs in real time, conditioning on her observation of the market dynamics. Learning has taken place if the latter approach is adopted. But learning is a rather complex process because one's expectations of prices affect prices, prices are used to infer others' assessments, and the inference of others' assessments in turn affects both prices and expectations of prices.

Some traders may take a dramatically different approach by largely ignoring all fundamental information about the underlying products and focusing on stock market dynamics only. These traders play the roles of speculators or market-makers who try to gain from the market by taking advantage of price volatility, providing liquidity, or looking for arbitrage opportunities. Their presence may introduce mixed effects to the market. While they could enhance liquidity on one hand, they may also introduce speculative bubbles and excess volatility into the market.

In summary, STOC participants may include some combination of naïve traders, long-term investors, and predatory arbitrageurs. The dynamics of the interactions between different groups within a given population is quite complex (Farmer & Lo 1999, Farmer 2002), and are beyond the scope of our study, but the principal of information revelation via the price-discovery process is the key to the STOC market's ability to infer aggregate preferences for concepts.

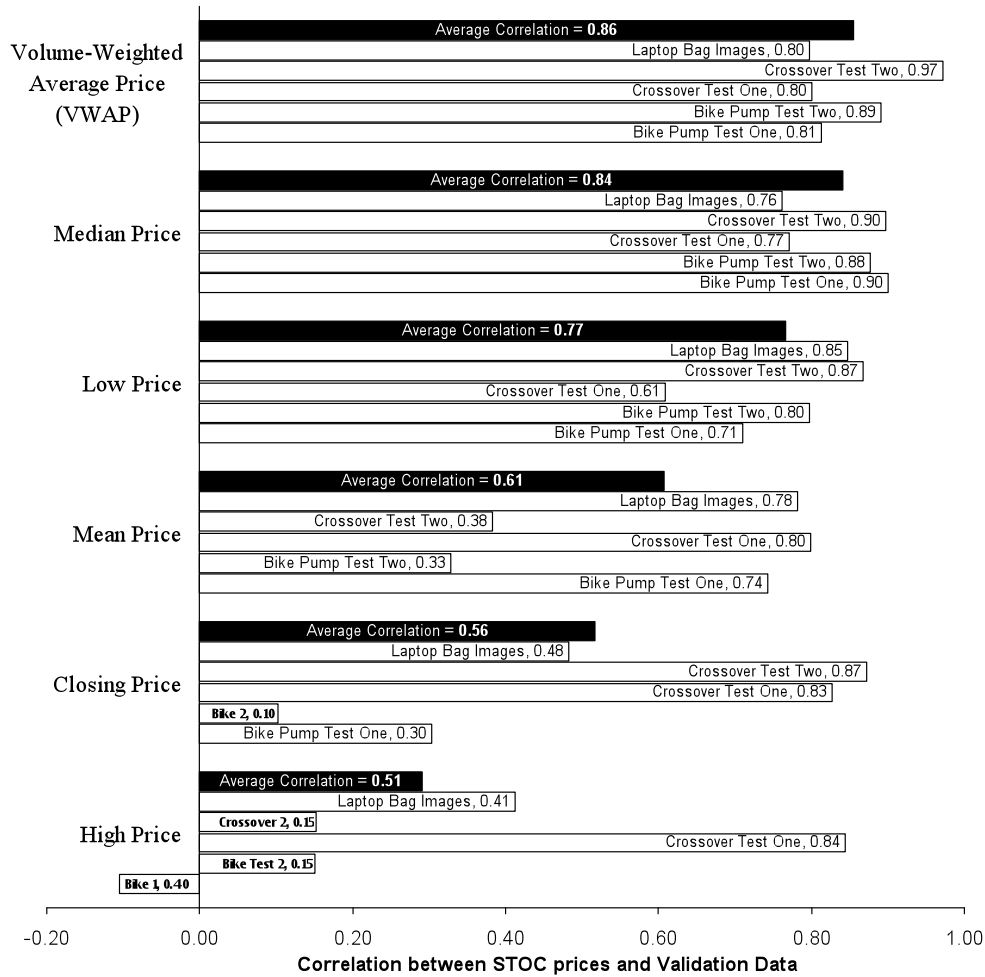
4 Results of STOC Tests

Our key findings focus on three areas: (1) which data metric should be employed? (2) how valid are STOC's results? and (3) what does trading reveal about preference heterogeneity?

4.1 STOC metrics

The outcomes of the first eight STOC tests for three product categories which are summarized in Table 2 lead to our first key result regarding which metric best summarizes trading. In Figure 12, we see that Volume-Weighted Average Prices (VWAP) fit the validation data better and more consistently than five alternative metrics.

Figure 12: Which STOC Metric Correlates Best to Validation Data?
 Performance of (6) Potential STOC Metrics in Five Empirical Tests



In subsequent results, therefore, we utilize VWAP outcomes to check internal and external validity. Recall that VWAP summarizes all of the trades made from start to finish during a trading game, weighting each trade price by the number of shares traded. Closing and High prices, which do not depend on all trades but rather depend exclusively on the final or highest trade for each security, had the worst fits with the validation data. One potential cause of the poor correlation is that in early tests we allowed market orders, in which a trader does not specify a price, and is therefore vulnerable to executing a trade at an unreasonable price. After seeing this occur in approximately 1% of trades in our two bicycle pump tests, we eliminated the option of placing market orders and required each trader to specify a price by placing limit

orders. We note that the VWAP and median metrics are less sensitive to a small number of trades at extreme prices.

Another explanation for the dramatic difference in correlation indicates that STOC markets, unlike prediction markets, behave more like traditional market research in which data is sampled from distributions with stationary means. In financial and prediction markets, this is not the case as security prices do not have stationary means due to the continuous arrival of new information. In our case, each STOC trade is similar to a survey response, and additional traders and greater trading time increase the sample size.

Shin and Dahan (2008) develop a statistical model to test whether STOC market data conform to stationary or non-stationary processes. They analyze the same trading data as in the present research, and employ unit-root tests to verify the stationarity or lack thereof of the mean prices for each security. Their model is:

$$(1) P_{i,t} = \alpha_i + \gamma_i \cdot DIFF_{i,t-1} + e_{i,t}, e_{i,t} = \rho \cdot e_{i,t-1} + v_{i,t}, v_{i,t} \sim w.n.$$

where α is an intercept term for each stock, $DIFF$ is the difference between the current stock price, P , and that stock's volume-weighted average price (VWAP) *up to* time t , and the errors, the e 's, are autoregressive and follow a white noise process (*w.n.*), i.e. a sequence with zero mean, constant variance, and no serial correlation (Enders 2004).

In their empirical analysis, the α_i 's for each stock are highly correlated to our VWAP values, and their results support the conclusion that an ideal STOC metric should include all trades. Importantly, applying Shin⁴ and Dahan's test of stationarity to all of our STOC data, we

⁴ We thank Professor Hyun Shin for his assistance with these analyses.

find that all of our trading experiments pass the stationarity test at the 95% confidence level. That is, they do not follow a random walk in which the next trade is equally likely to be above the current stock price or below it. In fact, their model correctly predicts whether the price of each stock will rise or fall in the next trade 72.1% of the time vs. a 50-50 guess. This stationarity occurs because there is no outside information entering the market in a STOC game, only the individual preferences and beliefs about others that exist at the time trading begins, and that are revealed through the price mechanism as the game unfolds. One way of interpreting this result is that STOC may work more like traditional market research in which the trade data collected is a random sample drawn from a stationary distribution.

4.2 STOC validity testing

We wanted to see whether traders were consistent from test 1 to test 2, and whether the preference “consensus” represented by the STOC results matched those of Dahan and Srinivasan (2000). To verify the validity of the market method, we ask two questions: (1) whether the results from the market method are consistent across different experiments, and (2) how close the results from the markets are to those from Dahan and Srinivasan (2000).

Table 4: Correlations Between (2) Bicycle Pump STOC Tests and Validation Data From Dahan and Srinivasan (2000)

	Conjoint Analysis	Web Static Images	Test 1 STOC	Test 2 STOC
Physical Prototypes	0.99****	0.99****	0.75**	0.82***
Conjoint Analysis		0.98****	0.75**	0.83***
Web Static Images			0.81***	0.89***
Test 1 STOC				0.86***

Significance level: **p<0.05, ***p<0.01, ****p<0.001

The results in Table 4 reveal a high degree of correlation between both STOC tests and the original validation data. Table 5 calculates the mean absolute errors between any two sets of market share predictions.

Table 5: Mean Absolute Error (MAE) Between (2) Bicycle Pump STOC Tests and Validation Data From Dahan and Srinivasan (2000)

	Conjoint Analysis	Web Static Images	Test 1 STOC	Test 2 STOC
Physical Prototypes	5.6%	1.7%	7.7%	7.2%
Conjoint Analysis		5.3%	5.7%	5.4%
Web Static Images			8.0%	7.3%
Test 1 STOC				1.4%

We find that the top three products (Skitzo, Silver Bullet and Epic), in terms of predicted market share and rankings, are the same in the two experiments, as well as in the validation data in the original Virtual Concept Test research. In a typical concept testing process, it is important to be able to identify the best designs so as to allocate resources to those opportunities with the greatest potential, and STOC seems to fulfill this role well.

For consistency across experiments, we calculate the Pearson correlation and mean absolute error between the nine market shares based on the VWAP's from the two tests. Comparing STOC with the original VCT web static data, correlations of 0.81 and 0.89 for tests 1 and 2, respectively, are slightly higher than those between STOC and the physical prototypes (0.75 and 0.82, respectively). This is consistent with the fact that the STOC traders were shown the web static images and not the physical prototypes. But MAE results for STOC show slightly

better agreement with physical prototypes than with web images. Also noteworthy is that test-to-test reproducibility is quite high, with a 0.86 correlation and a very low 1.4% MAE.

The results from these initial bike pump experiments show a remarkable agreement with those from the Dahan and Srinivasan study despite fundamental differences between the two methods, and wide separation in time and location. Differences include those in the data collection mechanism (a virtual security market versus a virtual shopping experience), the modeling of the predicted market share (the use of relative security prices versus individual level conjoint analysis), the questions asked (what *you* prefer versus what *the group* prefers), and lastly the subject population (MIT MBA students versus Stanford students of all types).

The two STOC tests of laptop PC bags tell a slightly different and remarkable story. The first laptop bag STOC test, in which the stimuli were presented in tabular form (Figure 4) with small images and feature details, failed to predict simulated store sales. On the other hand, the second STOC test, in which traders learned about the eight laptop bags by viewing full-sized images (Figure 5), performed quite well, yielding a correlation of 0.80 with the simulated store data, as seen in Table 6.

We attribute the dramatic difference in outcomes to the only factor that changed between the two tests, namely the stimuli. For the STOC method to perform well in capturing consensus preferences, traders must understand the product concepts reasonably well, so the quality of stimuli is crucial. While the feature table seems to have confounded or confused the traders, the full product images must have resonated with them. Extensive pre-testing of STOC stimuli is advised.

Table 6: Correlations Between (2) Laptop Bag STOC Tests and Simulated Store Unit Sales

	Test 2 STOC Image Format	Simulated Store Sales
Test 1 STOC Table Format	-0.14	-0.05
Test 2 STOC Image Format		0.80**

Significance Level: **p < 0.05

A more complete data set comes from the crossover vehicle case and the four STOC tests conducted using those eight stimuli. Key correlation results are summarized in **Error! Not a valid bookmark self-reference.**, with

Table 7: Correlations Between (4) Crossover Vehicle STOC Tests and Validation Data from Actual Unit Sales, Self-Stated Choices, and Virtual Concept Tests

	Test 1 Self- Stated	Test 1 VCT w/Prices	Test 1 VCT NO Prices	Test 1 STOC	Test 2 Self- Stated	Test 2 VCT w/Prices	Test 2 VCT NO Prices	Test 2 STOC	Test 3 Self- Stated	Test 3 VCT w/Prices	Test 3 VCT NO Prices	Test 3 STOC	Test 4 VCT w/Prices	Test 4 VCT NO Prices	Test 4 STOC
Actual Units Sold 2001-2006	0.44	0.58	-0.2	0.22	0.42	0.63*	-0.1	0.03	0.62*	0.48	-0.0	0.52	0.52	-0.4	-0.4
Test 1 Self- Stated		0.54	0.54	0.62*	0.89***	0.79**	0.55	0.64*	0.91***	0.37	0.63*	0.90***	0.74**	0.19	0.29
Test 1 VCT w/Prices			-0.1	-0.1	0.32	0.91***	-0.2	-0.1	0.55	0.97****	-0.0	0.50	0.94****	-0.3	-0.2
Test 1 VCT NO Prices				0.80**	0.44	0.14	0.96***	0.91***	0.51	-0.2	0.95****	0.62*	0.10	0.81**	0.72**
Test 1 STOC					0.58	0.20	0.85***	0.92****	0.61	-0.2	0.90***	0.69*	0.07	0.65*	0.62*
Test 2 Self- Stated						0.59	0.54	0.66*	0.84***	0.12	0.48	0.80**	0.49	0.00	0.07
Test 2 VCT w/Prices							0.08	0.20	0.81**	0.81**	0.27	0.79**	0.95****	-0.0	-0.0
Test 2 VCT NO Prices								0.97****	0.52	-0.4	0.93****	0.62*	0.00	0.76**	0.66*
Test 2 STOC									0.65*	-0.3	0.93****	0.73**	0.07	0.68*	0.57
Test 3 Self- Stated										0.35	0.61	0.97****	0.72**	0.12	0.07
Test 3 VCT w/Prices											-0.2	0.31	0.85****	-0.3	-0.2
Test 3 VCT NO Prices												0.72**	0.19	0.83***	0.73**
Test 3 STOC													0.67*	0.31	0.19
Test 4 VCT w/Prices														-0.1	-0.0
Test 4 VCT NO Prices															0.83**

Significance level: *p<0.10, **p<0.05, ***p<0.01, ****p<0.001

significant results in bold, and all are calculated within each group of traders. “Self-stated” survey data represent the normalized market shares for vehicles ranked by each individual in the top 3 out of eight choices. “VCT w/Prices” represent vehicle market shares based on scoring in the top three of eight vehicles using Dahan and Srinivasan’s 2000 methodology and accounting for the price of each vehicle when calculating utility. “VCT NO Prices” is the same calculation, but based only upon vehicle preferences *without accounting for vehicle prices* in the utility calculations. And, as before, “STOC” represents the normalized market shares based on the volume-weighted average prices of each of the eight securities. Several significant results are captured in A more complete data set comes from the crossover vehicle case and the four STOC tests conducted using those eight stimuli. Key correlation results are summarized in **Error! Not a valid bookmark self-reference.**, with

Table 7, including:

1. **STOC vs. Actual Sales:** The first row of the table, in which correlations to actual 2001–2006 unit sales of the eight vehicles were calculated for each method, reveals that all four STOC tests failed to predict actual sales. This result confirms our earlier analysis that STOC markets are not prediction markets, but rather measure underlying preferences among the traders as we shall see shortly.
2. **Self-Stated Choices and VCT w/Prices vs. Actual Share:** Self-stated choices and Virtual Concept Testing with pricing did predict actual unit sales (correlations in the 0.42 to 0.63 range), though not in the statistical significance sense. One explanation for the superiority of these two measures over STOC is that there is an important difference between what people *prefer*, and what they are *willing to pay*. STOC seems to zero in on preference rather than willingness-to-pay. Also, vehicle prices were not emphasized during STOC tests 1, 2 and 4, and were only featured prominently during STOC test 3, which was the only STOC test with some predictive value (0.52, but not significant).
3. **STOC Test-to-Test Reproducibility:** We saw with bike pumps that test-to-test reliability between STOC games using the same stimuli and same traders was quite good. In the crossover case, we can go further and measure test-to-test reliability across different groups of traders. Five of the six pairings of STOC tests reveal reasonably strong correlations between 0.57 and 0.92. But STOC tests 3 and 4 were not in agreement at all (correlation of 0.19), possibly because

in STOC Test 3, where vehicle prices were emphasized, the higher-priced Audi, Mercedes and BMW vehicles garnered only 33% share. Test 4 had only 16 traders and vehicle prices were not emphasized, and the three highest-priced vehicles garnered a much higher 64% share.

4. **Stated Choice and VCT Test-Test Reproducibility:** We note that the four VCT w/Price tests correlated amazingly well with each other (0.81 to 0.97) as did the VCT NO Price Tests (0.76 to 0.96). Similarly, aggregate Self-Stated data were highly correlated (0.84 to 0.91). So even though individual preferences were heterogeneous, and group sample sizes were small ($n = 16$ to 49), aggregate preferences across groups were consistent across groups.
5. **STOC vs. VCT with and Without Prices:** STOC correlates remarkably well with the virtual concept test results when vehicle prices are not factored in (correlations of 0.80, 0.97, 0.72 and 0.83, respectively, for STOC tests 1 through 4). There was no correlation between the STOC tests and VCT *with* vehicle prices (-0.10 to 0.31). In short, STOC traders seem to focus on the vehicles when trading, but neither on the prices of those vehicles nor on the willingness-to-pay for those vehicles.

We consider the degree of correlation within- and across multiple tests and measures remarkable considering that most of the vehicles studied had not even entered the market and that the individuals comprising the trading groups were heterogeneous in their preferences and backgrounds.

A final data set including additional individual-level measures comes from three tests of Wii Video Game concepts conducted in 2009. Test 1 involved 35 MBA students on eight teams that developed the Wii Video Games depicted in Figure 6 on page 21. After presenting their concepts to each other at the final session of a course on New Product Development, the students used constant-sum voting to complete three surveys: (1) SELF Preferences for all concepts except their own (which determined project grades); (2) Expectations of Others' Preferences for all eight concepts (with recognition for the "best guessers"); and (3) expectations of the eight average STOC Prices just prior to trading. After trading was completed, the students completed a fourth constant sum "Post-STOC" survey utilizing what they had learned from trading to re-estimate the mean preferences for all eight Wii Video Game concepts.

Test 2 was identical to Test 1, except that the 55 Executive Education students completing the four surveys and playing the STOC game were judging the eight concepts from the first class. That is, they had no involvement in the development of those eight concepts. Key results for tests 1 and 2 are summarized in Table 8.

Table 8: STOC Prices for (8) Wii Game Concepts vs. Constant Sum Surveys of SELF Preferences, E[Others' Preferences], E[STOC Prices], and Post-STOC Estimates in Two Experiments

	Test 1 E[Others]	Test 1 E[STOC]	Test 1 Post	Test 1 STOC VWAP	Test 2 SELF Pref	Test 2 E[Others]	Test 2 E[STOC]	Test 2 Post	Test 2 STOC VWAP
Test 1 (n=35) SELF Preferences	0.94****	0.92***	0.90***	0.89***	0.02	0.09	0.24	0.37	0.39
Test 1 E[Mean(Others)]		0.98****	0.90***	0.91***	0.28	0.35	0.47	0.57	0.59
Test 1 E[STOC Prices]			0.92****	0.93****	0.29	0.40	0.55	0.64*	0.67*
Test 1 Post-STOC				0.98****	-0.0	0.14	0.35	0.51	0.56
Test 1 STOC VWAP					0.00	0.15	0.33	0.50	0.55
Test 2 (n=51) SELF Preferences						0.94****	0.84***	0.72**	0.69*
Test 2 E[Mean(Others)]							0.93****	0.81**	0.79**
Test 2 E[STOC Prices]								0.95****	0.94****
Test 2 Post-STOC									0.99****

Significance level: *p<0.10, **p<0.05, ***p<0.01, ****p<0.001

Test 3, the results of which are summarized in Table 9 below, was similar to Test 1, only testing the (11) eleven Wii Video game concepts depicted in Figure 7 on page 22. Here again, the 58 students who developed the concepts were the ones voting on their relative preferences and trading in the STOC game. Comparisons to conjoint analysis for all 3 tests are in Table 10.

Table 9: STOC Prices for (11) Wii Video Game Concepts vs. Constant Sum Surveys of SELF Preferences, E[Others' Preferences], E[STOC Prices], and Post-STOC Estimates

	Test 3 E[Others]	Test 3 E[STOC]	Test 3 Post	Test 3 STOC VWAP
Test 3 ($n=56$) SELF Preferences	0.88****	0.92****	0.47	0.32
Test 3 E[Mean(Others)]		0.97****	0.75****	0.66**
Test 3 E[STOC Prices]			0.74****	0.63**
Test 3 Post-STOC				0.96****

Significance level: ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Table 10: Predicting Aggregate SELF preferences: STOC vs. Conjoint Analysis
r-Squared, and Rank Correlation for 3 Tests (**winners in bold**)

	Test 1 ($M = 8, N = 35$)		Test 2 ($M = 8, N = 55$)		Test 3 ($M = 11, N = 58$)	
	r^2	Rank	r^2	Rank	r^2	Rank
Conjoint Analysis	0.60	0.69	0.06	0.14	0.30	0.61
STOC VWAP	0.80	0.93	0.48	0.88	0.11	0.43

Several results from the three Wii Video Game experiments are worthy of note:

- **STOC measures E[Others]:** In all three tests, the correlations between E[Others] and volume-weighted-average-STOC-prices (VWAP) are significant and higher than those between VWAP and SELF preferences (e.g. 0.91 vs.0.89 in Test 1, 0.79 vs. 0.69 in Test 2, and 0.66 vs. 0.32 in Test 3). Traders rely more heavily on their beliefs about other traders than on their personal preferences when trading. This is especially evident in Test 3, and is reinforced by the individual level results in section 4.3
- **STOC vs. Conjoint Analysis is a draw:** Notwithstanding the prior result, STOC outperforms conjoint analysis in estimating SELF preferences for two of the three Wii Video Game tests, as the values in the lower row of Table 10 exceed those in the upper row for tests 1 and 2. Conjoint analysis outperforms in test 3, although neither method does particularly well in that test. Overall, it seems that having the information from *both* methods would be superior to depending on either one alone.
- **Aggregate Preferences are Heterogeneous:** The correlation between Test 1 and Test 2 measuring preferences for the same (8) concepts are quite low, and, for the most part, not significant. This implies that while all four surveys and STOC prices reflect aggregate preferences of a *particular* group of respondents, the two groups are quite heterogeneous in their preferences over the same concepts. One should not use the results of one STOC test to predict the results of another group with *different*

preferences. This may explain why STOC did poorly predicting crossover vehicle sales in the prior set of tests, and argues for a *representative sample* of traders.

- **Aggregate preferences are largely determined in advance of STOC trading:** The high correlations between $E[\text{Others}]$ and STOC VWAP (0.91, 0.79, and 0.66, respectively for tests 1, 2 and 3) imply that trading itself may not enhance the accuracy of preference measurement. We note that *Pre*-STOC estimates are more highly correlated to individual preferences than *Post*-STOC estimates. Learning may take place during STOC trading, but not necessarily to positive effect.
- **Wisdom of Crowds:** The high correlations between aggregate individual preferences and $E[\text{Others}]$ (0.94, 0.94, and 0.88), and between $E[\text{STOC Prices}]$ and STOC VWAP (0.93, 0.94, and 0.63) confirm strong wisdom-of-crowds effects. While many individual respondents and traders do not accurately estimate others' preferences, aggregating their individual beliefs produces an accurate estimate of aggregate preferences because most of the individuals' errors cancel out.

4.3 Eliciting individual preferences from STOC trading data

Traders' buying and selling decisions may be influenced by their individual preferences. Consider Table 11's four possible outcomes for any given stock trade between two individuals. While traders do not know the VWAP for each STOC while playing the game, the researcher can precisely measure those numbers in hindsight. If each trader's portfolio is valued at these VWAP's, then each trade can be viewed retrospectively as having been profitable or not. Selling shares above VWAP or buying below VWAP, as shown in the shaded areas, would be profitable. But *selling below* VWAP might indicate an "abnormal" dislike for a particular stock. And *buying a stock above* VWAP might reveal a strong preference for the concept associated with that stock. Below, we test these hypotheses empirically by looking at the 50% of transactions in the off-diagonal (non-shaded) cells. The goal is to detect individual preferences in the trading data.

Table 11: Trades between individuals are profitable for one person and unprofitable for the other; unprofitable trades may reveal individual preferences

	Buy a Stock	Sell a Stock
Below VWAP	Profitable Trade	Reveals Dislike
Above VWAP	Reveals Preference	Profitable Trade

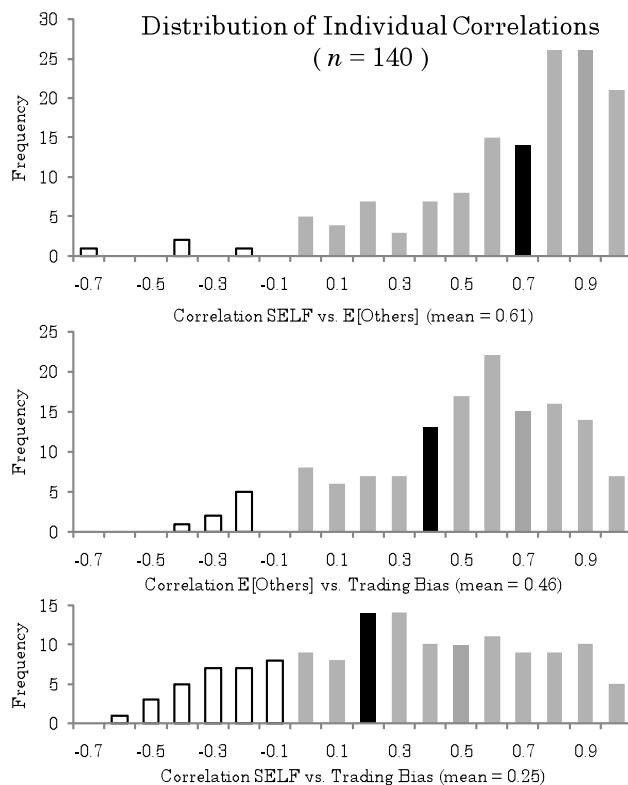
Below, we identify a trading-based metric (more fully developed in Appendix 3) for each trader, and compare this metric against known preferences from the individual-level survey data. The key result, depicted in Figure 13, is that trading reveals preferences at the individual level.

We define trader n 's bias for stock m as:

$$(2) \frac{VWAP_m + (\tilde{z}_{m,n} \times \tilde{\sigma}_m)}{\sum_{m=1}^M VWAP_m + (\tilde{z}_{m,n} \times \tilde{\sigma}_m)}$$

where the $\tilde{z}_{m,n}$ measures trader n 's relative bias for stock m using the volume-weighted trades made by that person for that stock in the off-diagonal cells of Table 11, and $\tilde{\sigma}_m$ measures the volume-weighted variance of stock m 's price relative to its volume-weighted mean (VWAP). We note that each trader's biases for all m stocks sum to 100%, and can therefore be compared to the constant sum survey results.

Figure 13: Individuals' trade based on their own expectations of others, which depend on self preferences due to false consensus



Two primary findings regarding individual level preferences include:

- There are strong positive correlations between the survey of self-preference and each respondent's expectation of others. This implies that most individuals' expectations of others is strongly biased by self-preference, consistent with the false consensus bias. Note that we saw a similar effect at the aggregate level in Table 8, where the 0.94 correlation between mean self preferences and mean expectations of others were extremely significant.
- More interestingly, individual trader preferences measured with the trading bias metric are positively correlated with the survey of self-preference data (mean $r = 0.25$), and even more so with the survey of $E[\text{Others}]$ (mean $r = 0.46$), which in turn are highly correlated with self preferences at the individual level (mean $r = 0.61$). It appears that individuals trade stocks biased by their expectations of others, which are consistent with false consensus beliefs. While STOC was not primarily designed to measure preferences at the individual level, the method enables traders do reveal their personal biases for and against the product concepts being tested. As with conjoint analysis, the STOC method measures implicit preferences rather than requiring explicitly stated preferences.

Figure 14: Individuals who are better estimators of others' preferences tend to rely more on their own expectations of others when trading

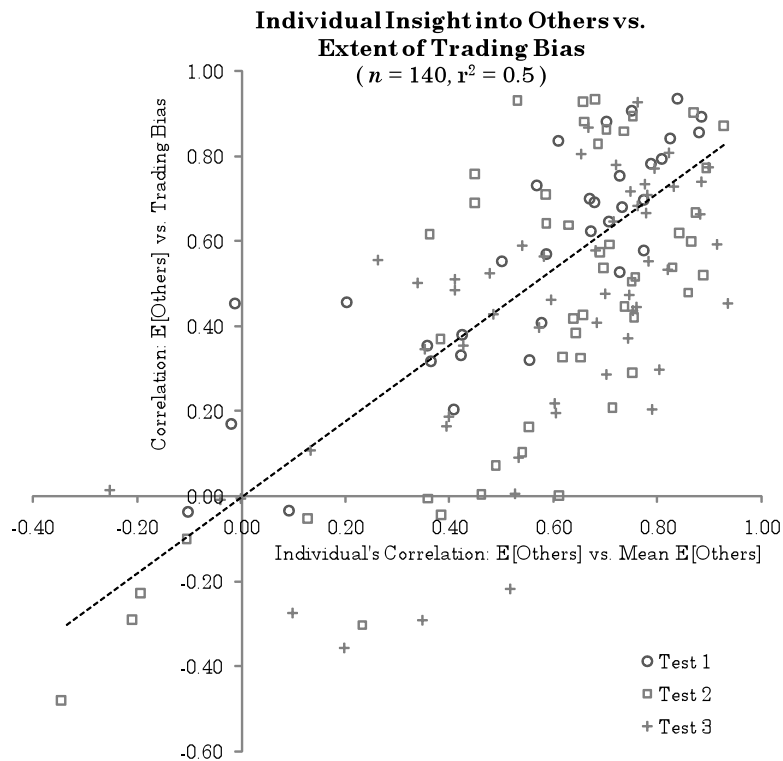


Figure 14 compares the ability of individuals to estimate others' preferences against the extent to which they rely on those estimates when trading, and reveals that individual who

estimate better tend to base their trading on their own estimates, while traders who had poor predictions of others' preferences tend *not* to rely as much on their beliefs when trading. One explanation for this remarkable result is that market prices tend to confirm the beliefs of those who start trading with more accurate estimates, while those whose estimates are far off the mark adjust their trading after learning market prices. That is, they learn from the STOC market.

5 Discussion and Conclusions

In this paper we study a novel application of the market mechanism: the use of securities markets to aggregate and infer diverse consumer preferences. We implement this idea in the specific context of four product-concept testing studies that aim to predict potential market share for between eight to eleven product prototypes. We observe that, unlike financial and prediction markets, stock prices have stationary means in the STOC data due to a lack of outside information. The results from ten of the eleven STOC experiments show remarkably high consistency among themselves, and significant correlation with independent preference measurement techniques. We note the importance of clear and salient stimuli, and the need for training and priming traders prior to the start of trading. STOC also appears to reveal heterogeneity of individual preferences as expressed in the form of trading biases. We caution that while the STOC methodology is particularly effective at identifying the most preferred concepts from among a larger set, it appears to be less effective at accurately measuring price sensitivity or predicting actual sales. Of course, *prediction* markets can be designed to perform the latter, and other choice-based market research techniques such as conjoint analysis are ideal for measuring price sensitivity and tradeoffs at the individual level.

STOC offers two key Advantages relative to traditional market research methods:

- **Cost efficiency:** The fact that respondents prefer competing in the STOC game to taking surveys, combined with the efficiency of STOC's market pricing mechanism in aggregating preferences, reduces recruiting and respondent compensation costs by as much as 75% for a given number of concepts.
- **Scalability:** The ability to conduct a large number of concept tests quickly and simultaneously, limited only by the number of respondents, is particularly beneficial in the context of early phase new product development and Web 2.0 input from potential customers.

The efficacy of STOC markets at identifying winning concepts may not be particularly surprising to economists. After all, Keynes (1936) commented on the similarities between stock selection and a beauty contest over seventy-four years ago:

...professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole ...”

The analogy is perhaps more accurate for describing what happens in financial markets and STOC games in the *short run*. After all, over the long run financial stock prices depend not only on investors' subjective beliefs and expectations of others, but also on other objective information such as companies' earning potentials and valuations of assets. On the other hand, the trading experiments presented in this paper are precisely “beauty contests,” since values of the virtual securities are derived endogenously from the preferences of the market participants, and their expectations of others' preferences, both of which are largely subjective. To improve the reliability of STOC markets in other marketing applications, one may need to anchor the

values of the securities to some objective fundamental variables of the corresponding product concepts. To test market share predictions, for example, one could compare security values with the realized market shares of the subset of products that already exist, or, barring the existence of real market share data, with the outcomes of external customer choice surveys. We hope to refine STOC market methods along these and other lines in future research.

6 References

- “Box Office Muddle: How Will Films Fare From Week to Week,” (1996) *Wall Street Journal*, July 21, 2000. *Business Week*, November 11.
- Burchill, G. W. and C. H. Brodie (1997). *Voices into Choices*. (Cambridge, MA: Center for Quality Management.
- Calder, B. J. (1977). “Focus Groups and the Nature of Qualitative Marketing Research.” *Journal of Marketing Research*, 14, 353-64.
- Chan, N. (2001) Ph.D. Thesis: Artificial Markets and Intelligent Agents, February.
- Chan, N., E. Dahan, A. Lo and T. Poggio (2001). “Experimental Markets for Product Concepts,” CBCL Working Paper #200/AI Memo #2001-013, Massachusetts Institute of Technology, Cambridge, MA, July.
- Chen, K. and C. R. Plott (2002), “Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem,” California Institute of Technology.
- Dahan, E. and J. R. Hauser (2002), “The Virtual Customer,” *Journal of Product Innovation Management*, 19 (5), 332-53.
- _____ and V. Srinivasan (2000), “The predictive power of internet-based product concept testing using depiction and animation,” *Journal of Product Innovation Management* 17(2), 99-109.
- Davis, D. and C. Holt (1993). *Experimental Economics*. Princeton, NJ: Princeton University Press.
- Fama, E. F. (1970), “Efficient Capital Markets: A Review of Theory and Empirical Work,” *Journal of Finance*, 25, 383-417.
- _____ (1991), “Efficient Capital Markets: II,” *Journal of Finance*, 46 (5), 1575-617.
- Farmer, D. (2002), “Market force, ecology and evolution”, *Industrial and Corporate Change* 11, 895–953.
- Farmer, D. and A. Lo (1999), “Frontiers of finance: Evolution and efficient markets”, *Proceedings of the National Academy of Sciences* 96, 9991–9992.
- Fern, E. F. (1982). “The Use of Focus Groups for Idea Generation: The Effects of Group Size, Acquaintanceship, and Moderator on Response Quantity and Quality.” *Journal of Marketing Research*, 9, 1-13.
- Forsythe, R. and R. Lundholm (1990), “Information aggregation in an experimental market”, *Econometrica*, 58, 309-47.

_____, F. Nelson, G. Neumann, and J. Wright (1993), "The Iowa Presidential Stock Market: A Field Experiment," *Research in Experimental Economics*, 1-43.

_____, T. Palfrey, and C. Plott (1982), "Asset Valuation in an Experimental Market," *Econometrica*, 50 (3), 537-68.

_____, T. Rietz, and T. Ross (1999), "Wishes, Expectations and Actions: A Survey on Price Formation in Election Stock Markets," *Journal of Economic Behavior & Organization*, 39, 83-110.

Green, P. E. and Y. Wind (1981), "New way to measure consumers' judgment," in Y. Wind, V. Mahajan & R. N. Cardozo eds., *New Product Forecasting*, D.C. Heath and Company, Lexington, MA, 89-108.

_____ and V. Srinivasan (1990), "Conjoint Analysis in Marketing: New Developments With Implications for Research and Practice," *Journal of Marketing*, 3-19.

Griffin, A. J. and J. R. Hauser (1993), "The Voice of the Customer," *Marketing Science*, Winter, 1-27.

Grossman, S. J. (1981), "An introduction to the theory of rational expectations under asymmetric information," *Review of Economic Studies*, 48, 541-559.

Hanson, R. (2003), "Combinatorial Information Market Design," *Information Systems Frontiers* 5(1) January, 105-119.

Hanson, R. (2007), "Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation," *Journal of Prediction Markets*, 1(1) February, 3-15.

Hanson, R. (2009), "On Market Maker Functions," *Journal of Prediction Markets*, 3(1), 61-63

Hanson, R., R. Oprea, and D. Porter (2005), "Information Aggregation and Manipulation in an Experimental Market," *Journal of Economic Behavior and Organization*, 60 (4), 449-59

Hayek, F. (1945), "The Use of Knowledge in Society", *American Economic Review*, XXXV, No. 4; September, 519-30.

Keynes, J.M. (1936), Chapter 12 of *General Theory of Employment, Interest and Money*, London: Macmillan, 401 pp.

Maloney, M. and H. Mulherin (2003), "The Complexity of Price Discovery in an Efficient Market: The Stock Market Reaction to the Challenger Crash", *Journal of Corporate Finance* 9, 453-479.

O'Brien, J. and S. Srivastava (1991), "Dynamic stock markets with multiple assets," *Journal of Finance*, 46, 1811-38.

Pennock, D. M. and M. P. Wellman (1997), "Representing Aggregate Belief through the Competitive Equilibrium of a Security Market," in *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, 392-400.

Pennock, D. M., S. Lawrence, C. Giles, and F. Nielsen (2001), "The real power of artificial markets," *Science*, 291, 987-988.

Plott, C. R. and S. Sunder (1982), "Efficiency of experimental security markets with insider information: An application of rational-expectations models," *Journal of Political Economy*, 90, 663-98.

_____ and _____ (1988), "Rational Expectations and the aggregation of diverse information in laboratory security markets," *Econometrica*, 56, 1085-1118.

Servan-Schreiber, E., J. Wolfers, D. M. Pennock, and B. Galebach (2004), "Prediction Markets: Does Money Matter?," *Electronic Markets*, 14 (3), 243-51.

Shin, H. and E. Dahan (2008), "A time-varying model of securities trading of concepts", UCLA working paper.

Smith, V. L. (1982), "Microeconomic Systems as an Experimental Science," *American Economic Review*, 72 (5), 923-55.

Spann, M. and B. Skiera (2003), "Internet-Based Virtual Stock Markets for Business Forecasting," *Management Science*, 49 (10), 1310-26.

Srinivasan, V. and A. D. Shocker (1973), "Estimating the Weights for Multiple Attributes in a Composite Criterion Using Pairwise Judgments," *Psychometrika*, 38, 4, (December), 473-493.

Sunder, S. (1992), "Market for Information: Experimental Evidence," *Econometrica*, 60 (3), 667-95.

Surowiecki, J. (2004), *The Wisdom of Crowds*. New York: Doubleday.

Toubia, O., D. I. Simester, J. R. Hauser, and E. Dahan (2003), "Fast Polyhedral Adaptive Conjoint Estimation," *Marketing Science*, 22 (3), 273-303.

Urban, Glen L., J. R. Hauser and J. H. Roberts (1990), "Prelaunch Forecasting of New Automobiles: Models and Implementation," *Management Science*, Vol. 36, No. 4, (April), 401-421.

Ward's Automotive News, monthly unit sales data by vehicle, 2001-2006.

Wolfers, J. and E. Zitzewitz (2004a), "Prediction Markets," *Journal of Economic Perspectives*, 18 (2), 107-26.

Appendix 1: Derivation of Equation (1)

Equation (1) can be derived by considering the two primary, people-related costs of running a market research study: (a) those costs associated with *recruiting* potential respondents and (b) those for *compensating* people who actually participate.

$$\begin{aligned}
 \text{Total Cost} &= \overbrace{\left(\frac{\text{Number of}}{\text{respondents needed}} \times \frac{\text{recruits}}{\text{respondent}} \times \frac{\text{Cost per}}{\text{recruit}} \right)}^{\text{Recruiting Costs}} + \overbrace{\left(\frac{\text{Number of}}{\text{respondents needed}} \times \frac{\text{Compensation}}{\text{per respondent}} \right)}^{\text{Respondent Compensation}} \\
 &= \frac{\text{Number of}}{\text{respondents needed}} \times \left(\frac{\text{Cost per recruit}}{\text{response rate}} + \frac{\text{Compensation}}{\text{per respondent}} \right) \\
 &= \overbrace{\left(\frac{M \text{ Concepts}}{\text{being tested}} \times \frac{\text{Sample size}}{\text{required}} \div \frac{\text{question capacity}}{\text{respondent}} \right)}^{\text{Number of respondents needed}} \times \overbrace{\left(\frac{\text{Cost per recruit}}{\text{response rate}} + \frac{\text{Compensation}}{\text{per respondent}} \right)}^{\text{Average cost per respondent}},
 \end{aligned}$$

which we summarize with the following notation:

$$(1) \quad TC = \left(M \cdot \frac{N_{\text{sample}}}{q_{\text{respondent}}} \right) \times \left(\frac{c_{\text{recruit}}}{r\%} + c_{\text{respondent}} \right).$$

where,

Symbol	Description
N_{sample}	Number of required Respondents
$q_{\text{respondent}}$	question capacity per respondent
c_{recruit}	cost to recruit each person
$c_{\text{respondent}}$	compensation for each respondent
$r\%$	response rate

Appendix 3: The Trading Bias Metric

Below, we develop a trading-based metric to measure individual biases and the preferences those biases imply.

- Denoting stocks using $m = 1, \dots, M$ and individual traders using $n = 1, \dots, N$

- Let $P_{m,n}^i$ be the price of the i^{th} trade of stock m for trader n

- $VWAP_m$: volume-weighted average price for stock m .

- $V_{m,n}^i$: volume (# of shares) of i^{th} trade of stock m for trader n

- $\tilde{\sigma}_m$: volume weighted standard deviation of stock m , which equals

$$= \sqrt{\sum_{i=1}^K \sum_{n=1}^N \left[(P_{m,n}^i - VWAP_m) \times I(VWAP_m, P_{m,n}^i, s_{m,n}^i) \right]^2 \times V_{m,n}^i} / \sqrt{\sum_{i=1}^K \sum_{n=1}^N V_{m,n}^i \times I(VWAP_m, P_{m,n}^i, s_{m,n}^i)}$$

- $\tilde{z}_{m,n}$: volume weighted z-score of stock m for trader n , which equals

$$= \sum_{i=1}^K \frac{P_{m,n}^i - VWAP_m}{\tilde{\sigma}_m} \times V_{m,n}^i \times I(VWAP_m, P_{m,n}^i, s_{m,n}^i) / \sum_{i=1}^K V_{m,n}^i \times I(VWAP_m, P_{m,n}^i, s_{m,n}^i)$$

- $s_{m,n}^i$: side of i^{th} trade (buy or sell) stock m for trader n .

- $I(VWAP_m, P_{m,n}^i, s_{m,n}^i)$: indicator function which returns 1 if trader *bought* shares *above* the $VWAP_m$ or *sold* shares *below* the $VWAP_m$ and 0 otherwise.

$$I(VWAP_m, P_{m,n}^i, s_{m,n}^i) = \begin{cases} 1 : s_{m,n}^i = \text{buy}, P_{m,n}^i > VWAP_m & \text{or} & s_{m,n}^i = \text{sell}, P_{m,n}^i < VWAP_m \\ 0 : \text{otherwise} \end{cases}$$

We now define trader n 's bias for stock m as:

$$(2) \quad \frac{VWAP_m + (\tilde{z}_{m,n} \times \tilde{\sigma}_m)}{\sum_{m=1}^M VWAP_m + (\tilde{z}_{m,n} \times \tilde{\sigma}_m)}$$

Note that each trader's biases for all m stocks sum to 100%, and can therefore be compared to the constant sum survey results.