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[♦♦]

First Draft: July 1, 2001

This Draft: July 20, 2008

[*] (*corresponding author*) Assistant Professor of Marketing, B-514, UCLA Anderson School, 110 Westwood Plaza, Los Angeles, CA 90095-1481, edahan@ucla.edu

[**] Postdoctoral Associate, MIT Sloan School of Management, MIT, Cambridge, MA 02139

[***] 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 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 partially funded by 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.

Securities Trading of Concepts (STOC)

Abstract

Market prices are well known to efficiently collect and aggregate diverse information regarding the economic value of goods, services, and firms, particularly when trading financial securities. We propose a novel application of the price discovery mechanism in the context of marketing research: to use pseudo-securities markets to measure consumer preferences for new product concepts. This is the first research to test potential new product concepts using virtual markets and the first to validate such an approach using eight years of stated-choice and longitudinal revealed preference data. We directly 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 seen to resemble traditional market research more than they resemble prediction markets.

As a measure of internal validity, each of three product categories is tested in 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 many concepts. To address the need for external validity we empirically test three approaches comparing STOC trading results against preferences measured through: (1) virtual concept testing (of bicycle pumps and crossover vehicles), (2) stated-choices (of crossover vehicles) 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). These experiments reveal that the market prices of securities designed to represent product concepts are remarkably efficient, accurate, and internally consistent measures of preferences, even when conducted with relatively few traders. We also note that while STOC prices do measure preferences, 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

Markets are well-known to be an efficient tool for collecting and aggregating diverse information regarding the value of commodities and assets (Hayek 1945). They have been particularly successful in the domain of financial securities. In this paper, we explore a novel application of the price-discovery mechanism of financial markets to marketing research using securities trading of concepts (STOC) to collect consumer preferences on product concepts. 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 also exploits the incentive-compatible nature of markets, i.e. the fact that over- or undervaluing securities reduces the participants' rewards, and the preferences of most participants for competing in games over responding to direct surveys.

In particular, we present results for multiple market experiments with live subjects in which participants express their preferences over new product concepts by trading virtual securities. The resulting securities prices are compared against preferences measured in separate studies of the same products using various stated-choice and revealed preference approaches: rank-ordered choice, the virtual concept testing (VCT) methodology developed by Dahan and Srinivasan (2000), and actual purchase decisions under controlled conditions and in actual product markets. We find that results across different market experiments in three product categories--bicycle pumps, laptop computer messenger bags and crossover vehicles--are, with one notable exception, reliable across repeated tests and predictive of stated-choice and revealed preference results. 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 essence of the STOC methodology centers around the establishment of virtual stock markets that trade virtual securities, each associated with an underlying product or service concept that can either exist or be at the conceptual or prototype stage. 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 (stocks) 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 last price prior to the closing of the market. Markets are typically open for 20 to 30 minutes. 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.

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 2000, 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.

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 description (and visualization) 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. Individual heterogeneity is not captured well in the end, even though it enters the trading process in the form of differences in security valuation. 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). 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 useful 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 and longitudinal revealed preference sales data for crossover vehicles. We then summarize relevant research from the prediction markets and experimental economics literature. Section 3 introduces the three 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 4 we conjecture on how STOC works by considering alternative strategies that traders might employ, and show how stock prices capture consensus preferences. Section 5 presents results from multiple STOC experiments, and develops a taxonomy of internal and external validity testing. We conclude in Section 6 with a discussion of the results, 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. The goal of their study is to identify the most preferred of nine bicycle bump concepts versus the two commercially available products depicted in In our eight 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).

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 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 4and 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 2002, as summarized in Table 1. Traders in the first six tests were MBA students, but additional tests included attendees from the MIT Center for Electronic Business conference (crossover vehicle test 3) and more senior managers attending executive education classes (crossover vehicle test 4). All eight tests were run under controlled conditions in a business school trading laboratory.

Table 1: Data Collected for each Product Category

Method *Product type*	Experiment	STOC Method	Virtual Concept Test	Self-Stated Choices	Simulated Store	Longitudinal Sales Data
Bike Pumps	Tests 1 & 2 *n* = 28	9 Pumps; Same traders tested twice	Dahan and Srinivasan '00 Physical, Web *n* = 102, 87			
Laptop Bags	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				
Crossover Vehicles	Test 1 *n* = 49	8 vehicles *No* Prices	VCT with and without Prices	Top 3 of 8 with prices		Cumulative units sold for each of 8 vehicles from 2001-2006 per Ward's *Auto News*
	Test 2 *n* = 43	8 vehicles *No* Prices	VCT with and without Prices	Top 3 of 8 with prices		
	Test 3 *n* = 42	8 vehicles *With* Prices	VCT with and without Prices	Top 3 of 8 with prices		
	Test 4 *n* = 16	8 vehicles *No* Prices	VCT with and without Prices			

For each of the above tests, Figure 7'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

[1] The Iowa Electronic Markets, http://www.biz.uiowa.edu/iem/

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], predict 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 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 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 STOC price movements.

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

[2] *BusinessWeek*, 11/11/96
[3] The Foresight Exchange, http://www.ideosphere.com/fx

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 share some characteristics with such experimental economics markets, and information aggregation and dissemination provide compelling explanation for the success of our STOC market. 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.

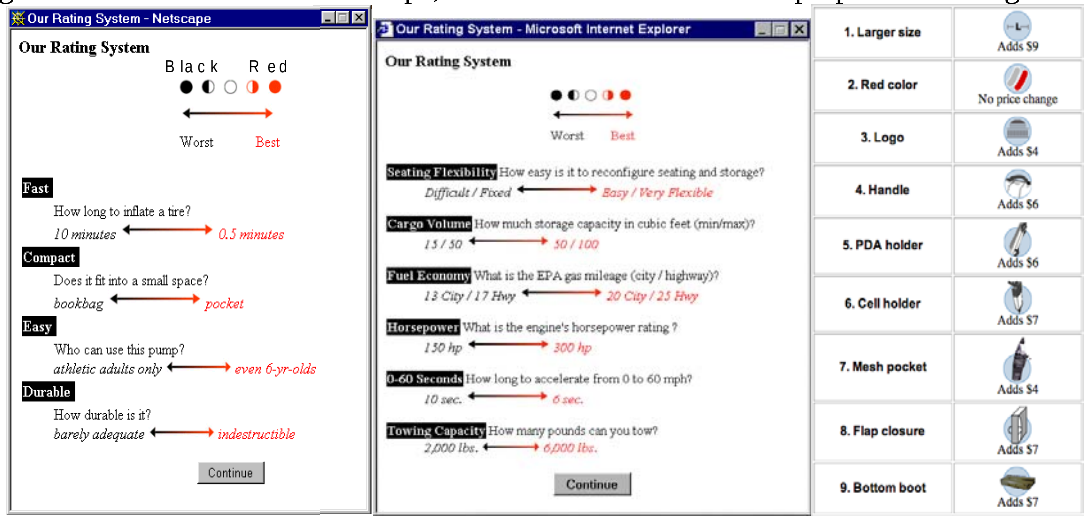
Alternatively, individual traders may form their own beliefs about the products, acknowledging that market prices will depend on aggregate beliefs. This is similar to the information aggregation scenario in which there are no "insiders", but where all traders are partially informed. Even in this case, 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 our STOC market 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.

3 Design of Markets and Securities

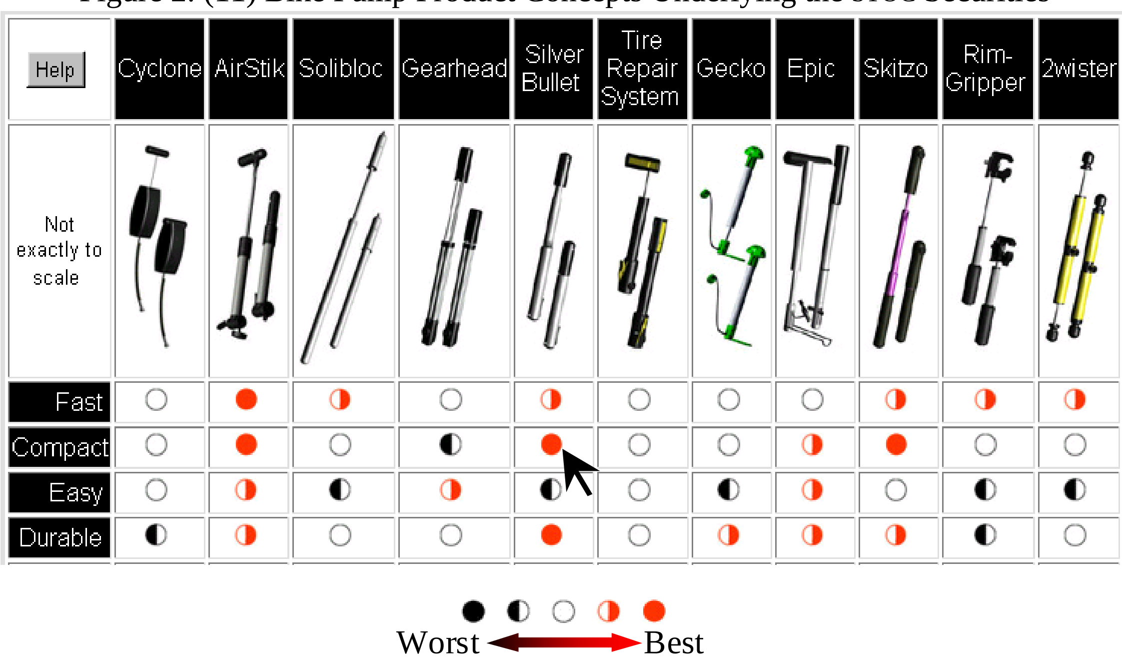
In addition to presenting full product concepts, we educated traders about the product attributes and attribute levels using Consumer Reports-style ratings as in Figure 1.

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



In our eight 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).

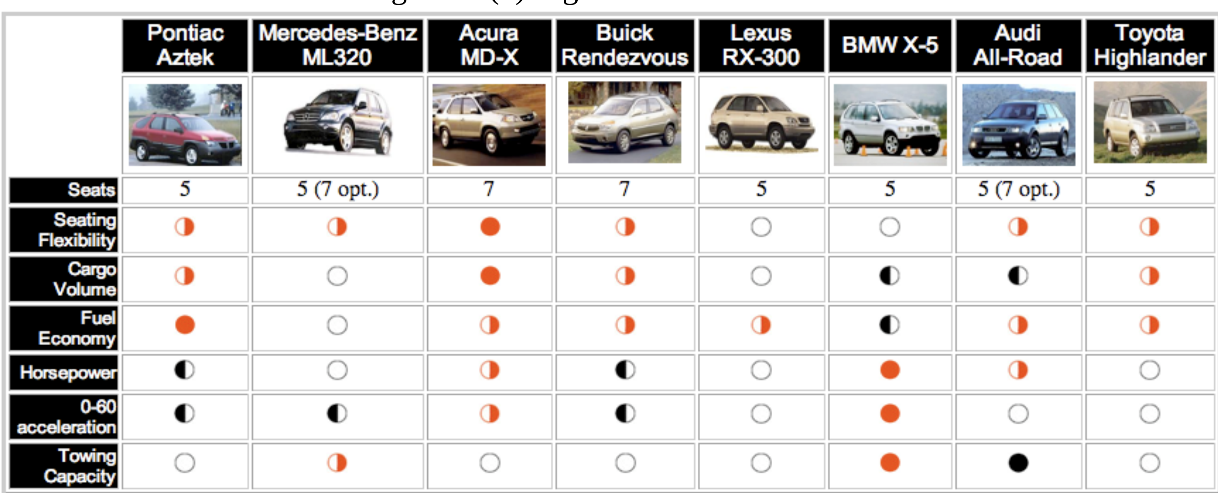
Figure 2: (11) Bike Pump Product Concepts Underlying the STOC Securities



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.

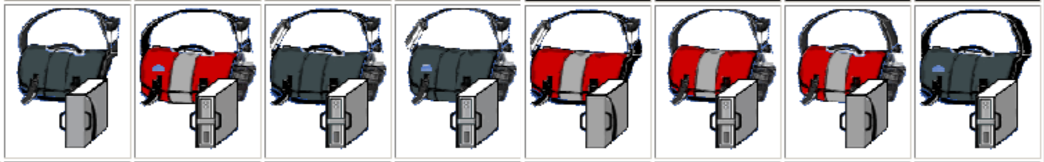
The eight crossover vehicles in Figure 3 were tested in 2000 and 2001 and consisted of three already released at the time (Lexus, Mercedes and BMW) and five yet-to-be-released vehicles (Pontiac, Acura, Buick, Audi and Toyota).

Figure 3: (8) Eight Crossover Vehicles



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



	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

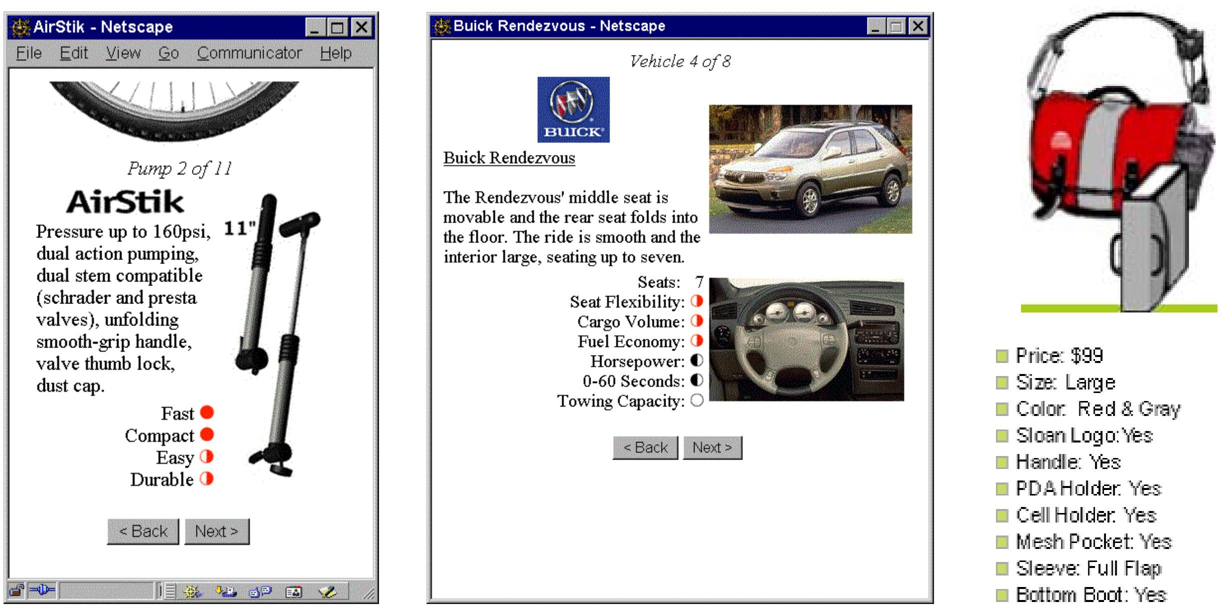
In the first laptop bag STOC test, traders saw the eight laptop bags in the table shown in Figure 4. In test 2, the eight laptop bags were depicted as simpler images, four of which are reproduced in Figure 5, leaving out the table of product attributes and simply showing nine product attributes visually 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 Without the Attribute Data Table

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.

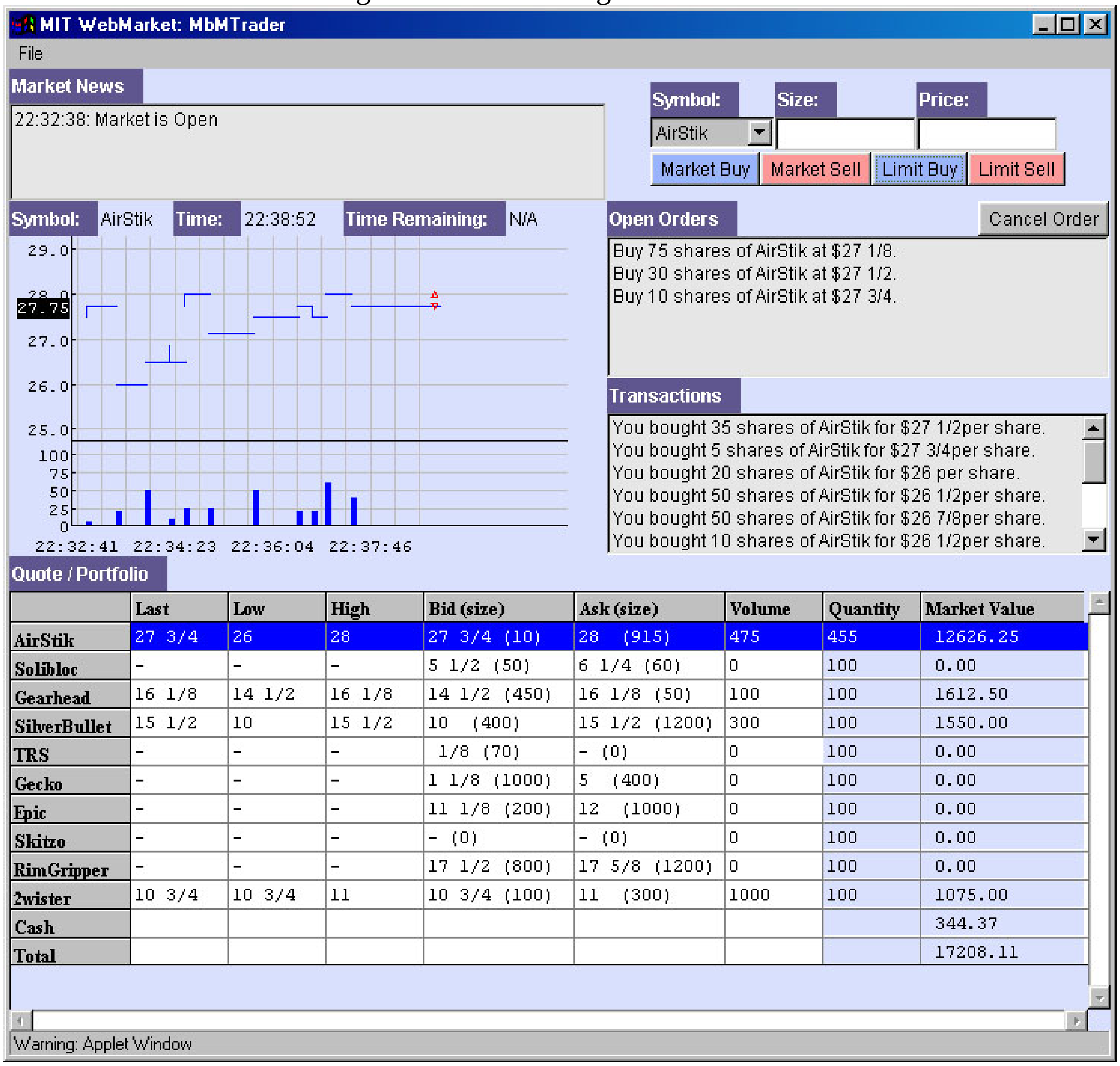
Each stock security represents a particular bike pump, crossover vehicle, or laptop PC messenger bag. The objective of traders in the STOC game is to maximize the value of one's portfolio at market close. The value of a portfolio is calculated as the sum of the cash on hand plus the total market value of the stocks, as determined by the closing market price. 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 eight tests, peer recognition was the only incentive offered.

Figure 6: 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 7.

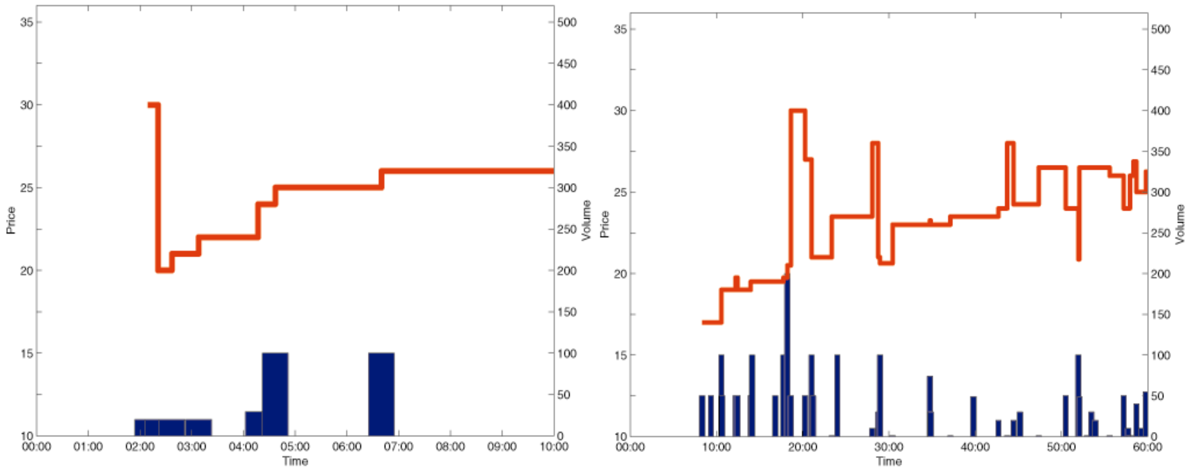
Figure 7: STOC Trading User Interface



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.

The trading prices (lines) and volumes (bars) for a sample security in the two bike pump tests are shown in Figure 8. We note that price converges to approximately $25 in both tests.

Figure 8: Price and Volume History of AirStik



Test 1 Test 2

For analysis, we focus on the trading data, which consists of the time series of trading prices and quantities $(p_{1,i}, q_{1,i}), (p_{2,i}, q_{2,i}), \ldots, (p_{T_i,i}, q_{T_i,i})$, 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 $p_{1,i}, p_{2,i}, \ldots, p_{T_i,i}$; the volume-weighted average price (VWAP) is computed as follows:

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

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 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.

Now 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,

$$\mathrm{MS}_i = \frac{\tilde{p}_i n}{\sum_{j=1}^{N} \tilde{p}_j n} = \frac{\tilde{p}_i}{\sum_{j=1}^{N} \tilde{p}_j},$$

where N is the number of securities comprising the market and n is the total number of shares for a security, which is equal for each security (each trader is given an identical endowment of each security prior to trading). Among the four price statistics, we expect the median price and VWAP to be particularly robust against potential price volatility.

4 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, may derive from three related elements**:**

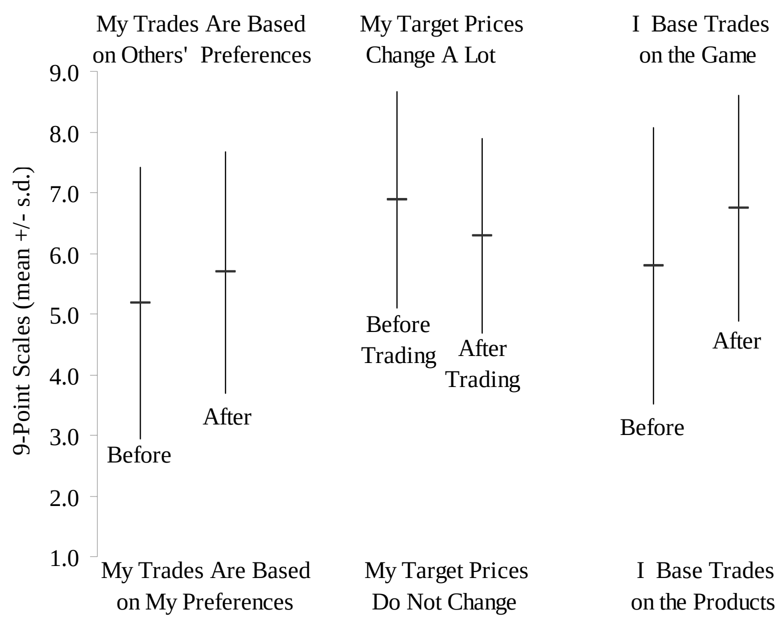
1. **Product Information**. This is what a participant knows about the underlying products. All participants are provided with the same facts and specifications of the products, but they may have obtained extra product information from their personal experience outside the experiments.

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 the two crossover vehicle STOC games, as summarized in Figure 9.

Figure 9: 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 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.

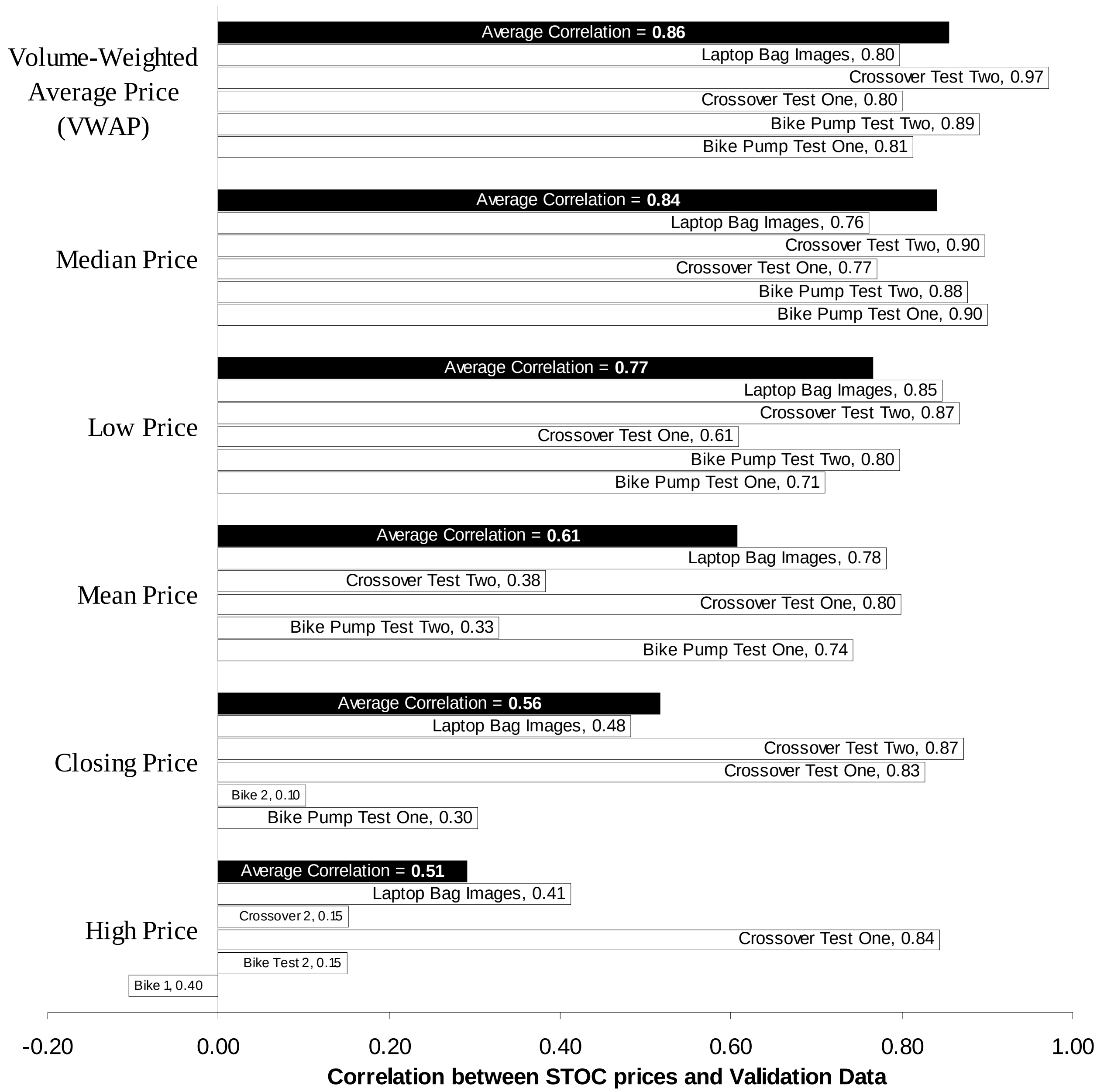
In summary, STOC participants may include some combination of naïve traders, long-term investors, and predatory arbitrageurs. The dynamics of the interactions between different groups within a given population is quite complex (Farmer & Lo 1999, Farmer 2002), and are beyond the scope of our study, but 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.

5 Results of STOC Tests

The outcomes of the eight STOC tests for three product categories which are summarized in Table 1 led to our first key result regarding which metric best summarizes trading. In Figure 10, we see that Volume-Weighted Average Prices (VWAP) fit the validation data better and more consistently than five alternative metrics.

Figure 10: 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 implies stationary on non-stationary security prices. 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. The α coefficients in their model are highly correlated to our VWAP data, and their results support the conclusion that an ideal STOC metric should include all trades.

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

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

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

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 original Web Static 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 a pair-wise sample Pearson correlation of 0.86 between market share predictions 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, respectively, for tests 1 and 2 are slightly higher than those between STOC and the physical prototype results (0.75 and 0.82, respectively). This is consistent with the fact that the two groups of STOC traders were shown only the web static images and not the physical prototypes. Also noteworthy is that test-to-test reproducibility is quite high, with a correlation of 0.86.

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, timing and geography. 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 students versus Stanford students).

The two STOC tests of laptop PC bags tell a slightly different, and quite remarkable story. The first laptop bag STOC test, the one in which the stimuli were presented in the form of a table with small images and feature details (Figure 4) failed miserably in correlating to the simulated store sales. The second STOC test, however, the one in which traders learned about the eight laptop bags by viewing full-sized images without the feature table (Figure 5) performed quite well, yielding a correlation of 0.80 with the simulated store data.

Table 3: 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: **$p < 0.05$

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.

The most complete data set we analyzed comes from the crossover vehicle case and the four STOC tests conducted using those eight stimuli. Key correlation results are summarized in Table 4, with significant results in bold, and all are calculated within each group of traders. "Self-stated" survey data represent the normed 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 normed marker shares based on the volume-weighted average prices of each of the eight securities.

Table 4: 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

Several significant results are captured in Table 4, including the following five:

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 a form of 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 for. STOC seems to zero in on preference rather than willingness-to-pay. Also, vehicle prices were not emphasized in 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 whopping 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 extremely heterogeneous, and group sample sizes were small (n = 16 to 49), aggregate preferences across groups were quite similar.

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 those vehicle prices.

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.

6 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 three product-concept testing studies that aim to predict potential market share for between eight to eleven product prototypes. The results from three seven of eight tests 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 STOC games. We also caution that while the STOC methodology is particularly effective at screening most preferred concepts from among a larger set, it appears to be less effective at 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 are ideal for measuring price sensitivity.

The efficacy of STOC markets at identifying winning concepts may not be particularly surprising to economists. After all, Keynes (1958) commented on the similarities between stock selection and a beauty contest over a half-century 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, one may need to anchor the values of the securities to some objective fundamental variables of the corresponding products. To test predictions of market shares, for example, one could compare security values with the realized market shares of the subset of existing products, or, barring the existence of real market share data, with the outcomes of actual customer choice surveys. We hope to refine STOC market methods along these lines in future research.

7 References

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.

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.

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.

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.

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

Keynes, J. M. (1958), "The General Theory of Employment, Interest and Money," *Harcourt Brace.*

Mahajan, V. and Y. Wind (1992), "New Product Models: Practice, Shortcomings, and Desired Improvements," *Journal of Product Innovation Management, 9 (June), 128-139.*

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

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.

Shefrin, H. (2005), "A Behavioral Approach to Asset Pricing Theory," *Elsevier North-Holland.*

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

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.

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.