Preference Markets: Organizing Securities Markets for Opinion Surveys with Infinite Scalability

Ely Dahan, Arina Soukhoroukova, and Martin Spann

January 9, 2007

Ely Dahan, UCLA Anderson School, Los Angeles, 110 Westwood Plaza, B-514, Los Angeles, CA 90095, (310) 206-4170, fax (310) 206-7422, edahan@ucla.edu

Arina Soukhoroukova, University of Passau, Department of Marketing and Innovation, Innstr. 27, 94032 Passau, Germany, as@ideamarkets.com, + 49 851 509 2423, (fax) 49 851 509 2422

Martin Spann, University of Passau, Department of Marketing and Innovation, Innstr. 27, 94032 Passau, Germany, spann@spann.de, + 49 851 509 2421, (fax) 49 851 509 2422

Keywords: Product Testing, New Product Development, Virtual Stock Markets, Product Design, Preference Markets, Information Markets, Scalable Market Research

Acknowledgements: The authors thank Martin Bernhardt, Barnaby Feder, Craig Fox, Mark Garmaise, John Hauser, Oliver Hinz, Rich Johnson, Donald Morrison, Hyun Shin, Bernd Skiera, Sven Theysohn, Agnieszka Wolk, and EMAC 2006-, Marketing Science 2005-, and Sawtooth Software 2006 Conference attendees for their valuable feedback.

Preference Markets: Organizing Securities Markets for Opinion Surveys with Infinite Scalability

Abstract:

Preference markets address the need for scalable, fast and engaging market research. For faster new product development decisions, we implement a flexible prioritization methodology for product features and concepts, one that scales up in the number of testable alternatives, limited only by the number of participants. Preferences are measured by trading stocks whose prices are based upon share of choice of new products and features. We develop a conceptual model of *scalable preference markets*, and test it experimentally.

Our conceptual model posits that individuals: (a) develop expectations of others based on self preferences, (b) use those expectations when buying and selling stocks, (c) have their opinions differentially weighted by the market pricing mechanism, resulting in a consensus of opinions, (d) learn from trading, and further converge towards consensus. Four studies confirm (a) - (d). Beyond accuracy, advantages of the methodology include speed (less than one hour per trading experiment), scalability (question capacity grows linearly in the number of traders), flexibility (questions in mixed formats can be answered simultaneously), and respondent enthusiasm for the method.

1 Introduction

In an environment of accelerating technology and short product life cycles, one in which a plethora of product features proliferates, new product development teams need fast and accurate marketing research. Smart phones, entertainment devices, information appliances, and other durable goods require development teams to *prioritize* literally hundreds of design decisions (Thompson, Hamilton and Rust 2005). There is a need to bridge the front end- and design phases by narrowing many features and concepts down to those key, make-it-or-break-it success factors. This requires a fast prioritization methodology, one that scales up in the number of testable features and concepts.

The more features or product concepts to be studied, the greater the number of participants and the cost and time required. Limits on the number of questions for participants derive from bounded rationality (Simon 1955), respondent fatigue (Shugan 1980), and time constraints. Faced with too many questions, respondents may resort to simplifying heuristics, even with tasks involving as few as 10-20 product features (Yee et al. 2006; Gilbride and Allenby 2004). .

In this paper we propose a new flexible approach to test preferences for large numbers of product features and concepts through the use of *scalable preference markets*. By engaging in stock trading, in which the price of each stock represents the degree of preference for a product aspect, feature or full concept, participants reveal their own preferences and their expectations of others' preferences, and converge towards a consensus of opinion.

Preference markets address the need for scalable, fast and engaging market research by combining elements of three methods: (1) actual financial markets, in which huge numbers of securities undergo continuous valuation through a fluid network of individuals trading with each other, (2) opinion surveys, which measure individual preferences, and (3) prediction markets, which measure expectations of future events through the market pricing mechanism using virtual stock markets.

While preference markets build upon these three approaches, they differ from them in important ways. Financial markets operate more efficiently when they are "thick," when the number of traders exceeds the number of securities being traded (Fama 1970). Even though no single individual has the capacity to follow every security, every security is still traded by a large number of individuals. Unlike in real financial markets, where traders self-select securities, we control our experiments by *assigning* traders to specific bundles of stocks, thus ensuring that every stock gets traded.

In opinions surveys, respondents answer each question once, do not learn from each other, and typically express self preferences. While stock trading outcomes are related to those measured by opinion surveys, they differ substantially. Preference market participants "answer" each question *multiple* times by buying and selling stocks throughout a trading task. Traders may learn from, and be influenced by, the behavior of fellow traders. And they may base their trading decisions on self preferences, on expectations of others' preferences, or on some combination of both. Indeed, Hoch (1987) and (1988) show that aggregating the opinions of heterogeneous

individuals produces different results than averaging those individuals' expectations of others. Surowiecki's *The Wisdom of Crowds* (2004) qualitatively illustrates the benefits of aggregating opinions of different individuals. While not completely eliminating biases, aggregation of diverse opinions frequently outperforms those of individual "experts", particularly if responses are weighted based on competence or confidence (van Bruggen, Lilien and Kacker 2002).

Scalable preference markets also continue the trend towards Internet-based market research, yielding benefits such as speed, adaptive interactivity, and task engagement (Dahan and Hauser 2002; Sawhney, Verona and Prandelli 2005). Other research has recognized the challenge of respondent fatigue, and addressed it through adaptive questioning (Sawtooth Software 1999; Toubia et al. 2003), more engaging tasks such as user design (Park, Jun and MacInnis 2000; Randall, Terwiesch and Ulrich 2006; Liechty, Ramaswamy and Cohen 2001), and task simplification as in self-explicated questioning (Kivetz, Netzer and Srinivasan 2004). Preference markets build upon these Internet benefits, but add competition and interactivity to enhance the respondent experience and align incentives for truth-telling.

Finally, previous research on prediction markets has used stock trading to forecast actual outcomes such as election results, movie box office receipts, or sporting event outcomes (see Table 1 for a summary of prior research on prediction and preference markets). In addition to this published research, firms such as Microsoft (Proebsting 2005) and Google (Cowgill 2006) employ internal prediction markets.

Preference markets, on the other hand, do not predict actual outcomes, nor are they based upon external information. Rather, they measure expectations of others' preferences, based upon individual self preferences combined with insights about others. While prediction markets typically run for weeks or longer, preference markets require only minutes, as there is no outside "news" to affect the market. For example, Dahan et al. (2006) evaluate product concepts in stock trading task that run less than an hour. Participants are presented with new product concepts then trade securities representing the competing designs. Their results exhibit high consistency and reliability across trading experiments and against independent surveys. However, to the best of our knowledge, no study has tested scalable markets for measuring preferences over a large numbers of concepts and features.

Table 1: A Sampling of Prior Research on Prediction- and Preference Markets

	Prediction Markets (actual outcomes)	***Preference Markets*** (no actual outcomes)
Few Stocks	Spann and Skiera 2003: 5 web-enabled cell services Chen and Plott 2002: 12 HP printers Wolfers and Zitzewitz 2004b: 3 war-related securities	Dahan et al. 2006: Study 1: 11 bicycle pumps; Study 2: 8 crossover vehicles
Many Stocks	Forsythe et al. 1992; Forsythe, Rietz and Ross 1999: multiple political races Pennock et al. 2001: 161 events on the Foresight Exchange Spann and Skiera 2003: 152 HSX.com movies Servan-Schreiber et al. 2004: 208 games on TradeSports and NewsFutures	The present research: 14 - 64 smart phone features and integrated products in four (4) studies and seven (7) stock markets

The aim of this research is to propose a conceptual model and methodology of scalable preference markets to handle large numbers of product concepts and

features, position the methodology in the context of other methods of new product research, validate the approach through empirical tests, and derive insights about the application of scalable preference markets to marketing problems. The paper is structured as follows. We discuss previous research and develop a conceptual model of preference markets in Section 2. Section 3 details the methodology. Section 4 analyzes four empirical studies of preference markets for the test of smart phones and their features. And section 5 concludes with a general discussion, managerial insights, limitations, and areas of future research.

2 Conceptual Model of Preference Markets

A conceptual model of scalable preference markets builds upon prior work on financial markets and experimental economics, including information- and prediction markets, as well as traditional market research. Four hypotheses are linked: (1) individual preferences lead to expectations about others, (2) rational traders use their expectations to decide when to buy and sell securities, (3) market prices aggregate information and beliefs held by individuals, and (4) individuals learn from markets. These four effects combine to explain how market prices measure people's beliefs about others' preferences.

2.1 The Wisdom (and Biases) of Crowds: Connecting the SELF to OTHERS

The task of estimating the market success of new products requires experts to distinguish between their own self-preferences and those of others, which may or may not be similar. Prior research has demonstrated that individuals' self preferences can bias their expectations of others' preferences (Hoch 1987). Yet,

aggregating individual opinions, even biased ones, produces surprisingly accurate and objective estimates of the consensus of opinion (c.f., Surowiecki 2004, Lorge et al. 1958). In financial markets, individual investors let individual biases influence their choices of stocks, for example by over-investing retirement savings in their own firms' stock (c.f. Benartzi and Thaler 2001; Huberman and Sengmueller 2004). And even in prediction markets with actual outcomes, such as the Iowa Electronic Market for the 1988 presidential race (Forsythe et al. 1992), "62% of the Bush supporters bought more Bush stock than they sold, while 68% of Dukakis supporters bought more Dukakis stock than they sold." So we expect self preferences to be strongly correlated to expectations of others, while being subject to the effect of biases.

H1a. Wisdom: *Self- preferences provide insight about others' preferences, therefore mean self-preferences and mean expectations of others' preferences will be highly correlated.*

H1b. Bias: *Individuals who prefer a product option have higher expectations of others' preferences for that option than do individuals who do not choose that option.*

2.2 Rational Expectations: How expectations of OTHERS affect ORDERS to buy and sell

Rational, profit-maximizing investors utilize personal knowledge in determining the value of a stock, and make trading decisions based upon this knowledge (Lucas 1972). This principle, well-established in financial markets, is also evident in experimental markets, where traders have an incentive to reveal their version of the truth (Smith 1982; Plott and Sunder 1982). Experimental markets have been shown to be quite robust to manipulation by some traders because other traders take the possibility of manipulation into account when setting *their own* expectations (Hanson, Oprea and Porter 2005). Even in opinion surveys where participants are

rewarded for making insightful observations, reward-maximizing contributors factor in their expectations of others' reactions (Toubia 2006). Similarly, we expect portfolio-maximizing traders in preference markets to buy and sell based upon their expectations of others, rather than on their own self-preferences for product features and concepts. Therefore, we expect stock prices to more closely correlate with expectations of others than with self preferences.

H2. *Expectations of others' preferences affect individual buy/sell decisions for stocks more than do self preferences, so stock prices will correlate more highly with expectations of others.*

2.3 PRICES: How markets achieve consensus based upon ORDERS

In a competitive economy, the price of a good reveals its value (Hayek 1945). Similarly, in financial markets, the efficient market hypothesis posits that stock prices aggregate all information known to traders (Fama 1970; Fama 1991). Experimental markets confirm the theory and converge towards "truth" within a few iterations, even when information is dispersed (Plott and Sunder 1988; Forsythe, Palfrey and Plott 1982). Importantly, not all market participants are equal in their influence on prices. In opinion surveys, van Bruggen et al. (2002) show that accuracy improves when the opinions of more confident or competent participants are weighted more heavily. In financial-and prediction markets, traders with greater knowledge or certainty, exert greater influence on prices, thus weighting their opinions more heavily. These "informed" traders with private information effectively set market prices for the "less informed" (Oliven and Rietz 2004).

For prices to be truly informative, profit-maximizing participants should be rewarded based upon *real outcomes* (Camerer, Loewenstein and Weber 1989; Smith, Suchanek and Williams 1988). But real outcomes may be hard to come by at the front end of new product development. For example, only a few of many product concepts or features may actually be launched. Product development teams may not be able to afford to wait for actual market outcomes because speed-to-market, and the first-mover-advantage that results from speed, are frequently key success factors.

For such research questions, the prerequisite that securities be linked to actual outcomes may have to be relaxed, and may either be replaced with an alternative such as actual survey results, or at least the belief on the part of traders that their portfolio valuation depends on such external survey results. As long as traders behave consistently with the belief that their portfolio performance will be measured based on actual, observable results, then market prices should reflect traders' expectations of those results. Even lacking observable outcomes, preference markets can utilize the market pricing mechanism to efficiently aggregate consensus expectations.

H3. *The market pricing mechanism summarizes the consensus of individual offers to buy and sell stocks, so market prices will correlate highly to mean expectations of others' preferences*.

2.4 LEARNING from market PRICES

In financial markets, investors continuously observe prices, and rapidly update valuations in response to events (MacKinlay 1997). In this context, stock price

shocks can become events in and of themselves. Upon receiving news or observing stock price changes, traders may respond by adding further volatility (Blanchard and Watson 1982; Timmermann 1993). In experimental- and prediction markets, traders also learn from prices, update their beliefs, and converge towards a common, competitive equilibrium price (Smith 1982; Bondarenko and Bossaerts 2000). Of course, one risk of this price-based communication is the possibility that traders will learn the *wrong thing* from each other, and jump on a "misguided bandwagon." Smith, Suchanek and Williams (1988) show that inexperienced traders may produce market bubbles and crashes. A few marginal traders with strong (but inaccurate) beliefs can influence a majority of traders who are weaker in their beliefs, leading to herding behavior. Further, the underlying individual preferences themselves can be influenced by such communication. Salganik, Dodds and Watts (2006) show that the individual preferences change dramatically when people are exposed to the preferences of others'. Additionally, publicly posted opinion surveys, such as critical reviews, polls, or even web chat, may affect individual opinions (c.f., Chevalier and Mayzlin 2006). In the case of preference markets, the potential for herding behavior suggests that repeated, independent market simulations are called for to verify inter-test reliability. On the other hand, for fashion goods or those with network externalities, where individual preferences are heavily influenced by others, the inter-trader communication inherent in preference markets offers potential insight about how preferences can evolve.

Regarding learning, preference markets are simpler than financial- and prediction markets. Rather than outside news, the only new information revealed to traders is

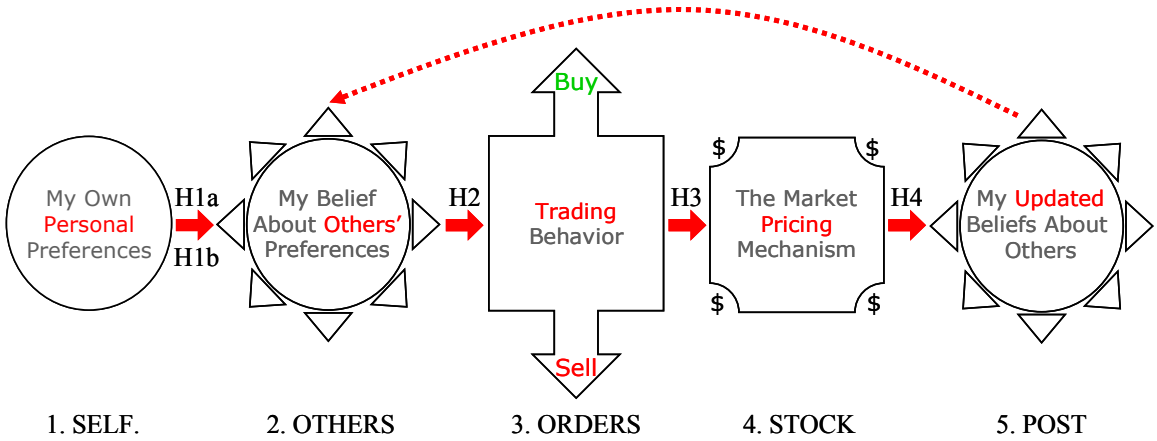
the stock price itself. Therefore, we expect traders in preference markets to learn from stock prices, update their own beliefs, and converge in their expectations of others.

H4. *The process of trading will cause traders to update their expectations of others' preferences, and will reduce the variability in these expectations across traders.*

2.5 A Conceptual Model of Preference Markets

Figure 1 integrates sections 2.1 through 2.4 and our four hypotheses into a conceptual model of preference markets.

Figure 1: A Conceptual Model of Preference Markets



The first circle represents SELF preferences, that is each trader's preferences for him- or herself. These preferences influence expectations of OTHERS' preferences [H1a and H1b] as depicted in the second, "outward-looking" circle. Traders' rational expectations of OTHERS inform buy/sell decisions [H2], as shown in the ORDERS square in the middle of Figure 1. For example, the value of a specific stock traded in a preference market can be linked to the percentage of people who choose to buy the product feature or concept linked to this stock. If a trader believes that more than 20% of people prefer the Motorola brand, for example, he or she should buy the stock

when the market price is $20[1]. Conversely, if a trader believes that less than 20% prefer Motorola, he or she should sell at that point. Market PRICES are determined through executed trades. Traders with stronger convictions hold greater sway in determining stock PRICES [H3]. Traders may engage in LEARNING, due to updated expectations of others based upon newly posted prices [H4].

3 A Methodology for Preference Markets

The primary objectives of running a preference market are to elicit from respondents (traders) the most truthful measurement of their expectations, and to negotiate a consensus amongst traders as to the group's overall preferences.

In order to achieve these objectives, it is imperative that: (1) traders understand what each stock means, and its connection to the product attribute-level or integrated concept it represents, (2) the number of stocks being traded by each trader is manageable, (3) the process of trading evokes decision processes and reactions similar to what respondents might experience when evaluating actual products, i.e. at least some traders must have information about each stock, and (4) each trader has an incentive to reveal "truth" throughout the trading task (c.f., Spann and Skiera 2003; Wolfers and Zitzewitz 2004a). With these objectives in mind, below we identify four design decisions relevant to preference markets[2], discuss our choices for each design decision, and highlight our methodological contributions.

[1] Traders are informed prior to trading that stock prices are defined as the percentage of people who chose to purchase each product or feature, a specific brand in this case, at a given retail price.
[2] Securities markets involve many more than just four design decisions, as delineated in Appendix 1.

3.1 Defining the Securities

In general, we define individual securities so that the price of a stock measures the strength of preference for a particular product, product concept, brand, feature, attribute level, or bundle of attributes. Traditionally, market research studies impose parallelism by focusing on one question type for a given task. For example, conjoint analysis asks respondents to evaluate product attribute levels, while concept testing has them compare product concepts, and sometime actual products (Huber et al. 1993). In fact, as the next section shows, it is relatively easy to "mix-and-match" actual products, product concepts, and all manner of features and attributes in a single preference market.

Stocks can represent *binary* product features (e.g., "FM Tuner Included" or "FM Tuner Not Included"), or *mutually exclusive* product features (e.g., forms such as "Brick", "Slide Open", or "Flip-Phone").

In order to accurately capture preferences, stocks must be defined in such a way that traders understand the connection between a stock's price and strength of preference. Stock prices can be defined in many ways such as "average rating on a 1-100 scale," "number of units that will be sold in a given period," or "percentage of people who would choose this option." To make the trading task a bit easier, the scale for stock prices should be common across all stocks.

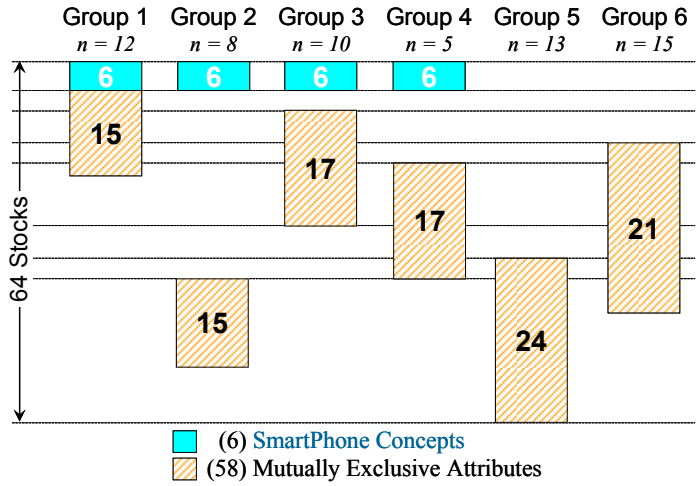
Further, the stimuli describing features and concepts must be sufficiently vivid to maximize salience. Once decisions have been taken about the specific aspect being

measured, the question format, and the exact definition of stock price need to be communicated clearly to traders.

3.2 Experimental Design

Next, we must connect traders with stocks. In financial- and prediction markets, traders self-select stocks, typically trading only a tiny percentage of the universe of securities. We seek both the scalability of financial markets and the superior control of experimental research, and achieve this by *assigning* small groups of traders to small groups of stocks. A contribution of the present research is the unique and highly flexible experimental design in which participants trade *certain* stocks *within* their subgroup and other stocks *across* multiple subgroups. Traders can be assigned to the subgroups randomly, or based on interest, product expertise, or market segment. A sample experimental design appears in Figure 2.

Figure 2: Organizing Traders and Stocks into Groups, each Trading a Subset



Upon reflection, one realizes that this experimental design offers many potential advantages in the areas of scalability and control. Trading stocks across subgroups ensures communication through the price mechanism, in effect putting all stocks on

a common basis. In fact, preferences in two or more product categories can be measured simultaneously by assigning some traders to one product category, other traders to the second category, and a third group of traders to both categories. A similar experimental design may assess price elasticity by measuring preferences over identical products and attributes at varying price points. Such "price-elasticity" stocks may be traded independently by separate groups of traders, to avoid context- and anchoring effects. An additional use would be to measure inter-test reliability by having multiple groups trading identical sets of stocks, just not with each other (while still having *some* stocks common across groups in order to maintain the common basis alluded to earlier).

Respondent sample

When recruiting respondents to be traders, we must consider the insight and expertise they bring, their membership in market segments, and whether to open trading to the public, or to restrict participation to pre-recruited respondents.

Regardless of how traders are matched to stocks, it is important to recruit at least some traders who have strong preferences in the product category, or who can provide insight about others' preferences. Since preference market prices represent the consensus view of the traders, one might assume that traders should comprise a representative sample drawn from the market for the product. But, as previously shown in political stock markets (Forsythe et al. 1999), and in prediction markets for business forecasting (Spann and Skiera 2003), a representative sample is *not* required, as long as traders possess insight about customers in that product market.

Further, a sufficient *number* of traders is needed, determined by the need for liquidity in each stock and the number of stocks each trader can handle. For example, if an individual trader can manage twenty stocks, and each stock needs to be traded by twenty people, then twenty stocks could be traded by twenty people in a trader subgroup.

The population from which participants are drawn depends on the product expertise and degree of secrecy required. Access might be restricted to invited consumers or employees in order to guard intellectual property embedded in the stock definitions. Or a narrow sample may be recruited out of a population with particular interest in the product category. On the other hand, an open access market, such as HSX or TradeSports allows users to self-select based on their degree of interest in each stock, potentially improving motivation.

3.3 Market pricing mechanism

Once a list of stocks has been specified, and groups of traders matched to groups of stocks, the mechanism for buying and selling stocks must be designed since it may profoundly effect preference measurements. Specifically, effective market mechanisms optimize "FUSS": (a) FEEDBACK: providing traders sufficient market information without overloading them, (b) USABILITY: keeping things simple enough for novices, yet allowing experts to "express" themselves through trading, (c) SALIENCE: having stock trading evoke responses consistent with the process of buying products, and (d) SPEED: completing the trading task quickly enough to keep traders engaged, while still collecting rich data.

Appendix 1 illustrates many design options for market mechanisms and user interfaces, but here we focus on three that stand out with respect to preference markets. First, taking advantage of the fact that no exogenous information enters a preference market, the duration of trading can be extremely short, particularly if an Internet-based double-auction mechanism is in place (c.f., Guarnaschelli, Kwasnica and Plott 2003 and Sunder 1992). A consensus regarding stock values should form literally within minutes, and most stock prices should stabilize in under an hour. To avoid potential boundary effects at the end of trading, traders are given a rough timeframe for trading, and the market is stopped randomly, after say 30-45 minutes.

While more exotic trading options such as short-selling, options, and derivatives are technically feasible, they veer away from the need for usability and salience, especially for novices.

Third, market boundary conditions affect outcomes. Specifically, endowments of stocks and virtual cash, and the overall level of liquidity in the market, may either hinder trading if liquidity is too low, or encourage excessive speculation if it is too high.

3.4 Incentives

Incentives should induce each trader to reveal his or her true expectations and actively involve throughout the experiment. In financial and prediction markets, traders attempt to maximize portfolio value because their payoffs are directly proportional to the liquidation value of all stocks and cash. Preference markets do not necessarily provide payoffs to every trader, nor do they reveal actual outcomes.

So rather than basing incentives on actual outcomes, they can be based on closing stock prices, which are endogenous to the market, and act as a surrogate for revealed outcomes. The lack of actual outcomes may leave preference markets vulnerable to anomalies such as bubbles and gaming. Alternatively, we can generate exogenous "truth" by conducting an independent preference survey, and using its results as the actual outcome for each stock. In addition, beyond rewarding trader performance and accuracy, we might reward effort (e.g., number of trades).

To reduce the cost of compensating every trader based on final portfolio values, we might randomly select prize winners based on the ranking of each portfolio within a trading subgroup. Maximizing one's expected-reward would still be consistent with maximizing one's portfolio value, even if one is not the top trader within the subgroup. The short duration of preference markets add to the *intrinsic* reward of competing, since within minutes of completing the market, one discovers one's ranking among all traders. The incentives need to be high enough to attract traders. Details are provided in section 4.3.

4 Empirical Studies

In order to test the feasibility and accuracy of scalable preference markets and to test our conceptual model from Section 2, we ran four studies within the smart phone product category, mapped out in Table 2, beginning with a relatively basic test of fourteen smart phone features with MBA students, and culminating in a test of sixty-four smart phone features, attributes and concepts with expert managers

and designers. Each study evolved from and improved upon earlier ones, and addressed additional hypotheses. The results from the four studies support the hypotheses and conceptual model.

Table 2: Study Roadmap

	Number of Stocks	Stock Types	Participant Types	Key Insights
Study 1	14	Binary Feature Levels only	4 groups of MBA Students	Verify the conceptual model
Study 2	14	Binary Feature Levels only	Executives	Outsiders also perform well
Study 3	56	Binary Features, Full Products, and Mutually Exclusive Feature Levels	MBA Students	Markets scale well; allow flexibility, [H1]-[H4] tested
Study 4	64	Only Mutually Exclusive Products and Feature Levels	Expert Managers & Engineers	Real world feasibility; Remote trading works, [H1]-[H4] tested

4.1 Study 1: Do Preference Markets Measure Actual Preferences?

Our first objective is to verify the two endpoints of Figure 1's conceptual model, namely that individual preferences are captured through stock prices in a preference market.

Study Design and Procedure

Study 1 included four replications of the same basic experiment. First, the fourteen feature levels of personal digital assistants (PDA's) shown in Figure 3 were identified based on individual- and group discussions about the key factors influencing the purchase decision.

Next, individual respondents designed an "ideal" device that optimized the tradeoffs against price, weight, and battery-life. The percentage of respondents choosing each feature level was then calculated. Fourteen stocks were defined in terms of these

percentages, so that more popular feature levels should command higher stock prices. For example, during user design, upgrading a smart phone to Bluetooth would add $49 to the retail price and would reduce battery life by 5%. If 35% of people made such an upgrade, then the value of the "BLUETOOTH" stock would be $35.

In replication 1A, 241 MBA students from a US East coast business school participated in a user design study, and 25 of them traded the related stocks two months later. Replications 1B, 1C and 1D took place two years later (so that some preferences for this fast-improving technology may have changed) at a US West coast business school.

Figure 3: (14) Feature Levels in the User-Design and Stock Experiments

			% Upgrade	Price	Weight	Battery
SIZE	compact	full-size		$10	3.0 oz	-
COLOR	monochrome	color		$99/149	-	-20%
MEMORY	16MB	32MB		$25	-	-
MEMORY	16MB	64MB		$50	-	-
OS	Palm	Pocket		$40	-	-5%
CELL	cell			$99	1.0 oz	-
HANDSFREE	hands free			$50	-	-5%
BATTERY	extended battery			$99	0.5 oz	+300%
WIRELESS	wireless			$99/149	0.5 oz	-10%
BLUETOOTH	blue tooth			$49	-	-5%
KEYBOARD	Thumb keyboard			$25	-	-10%
CF SLOT	CF slot			$15	0.5 oz	-5%
SD SLOT	SD slot			$15	-	-
GPS	GPS			$129	1.0 oz	-10%

Table 3 summarizes the correlations from all four replications of Study 1, along with the results from Study 2, described below. For example, we see that the 14 closing stock prices from Study 1A ("1A PRICES") have a correlation of 0.862 ($p < 0.001$) with

the market shares for the 14 feature levels ("1A SELF") among the 241 individual respondents.

Table 3: Correlations based on the Closing Prices of five Preference Markets

n =	241	25	41	37	39	49	47	44	33	34	19
	1A SELF	1A PRICES	1B SELF	1B PRICES	1B POST	1C SELF	1C PRICES	1C POST	1D PRICES	1D POST	2 PRICES
1A PRICES	.862***										
1B SELF	.804***	.640*									
1B PRICES	.557*	.491	.819***								
1B POST	.756**	.719**	.836***	**.880***							
1C SELF	.786***	.677**	.835***	.719**	.876***						
1C PRICES	.532	.669**	.608*	.750**	.852***	.758**					
1C POST	.782***	.795***	.776**	.794***	.947***	.919***	**.916***				
1D PRICES	.554*	.693**	.700**	.665**	.789***	.770**	.866***	.823***			
1D POST	.647*	.747**	.716**	.716**	.870***	.892***	.881***	.923***	**.920***		
2 PRICES	***.931***	.803***	.776**	.542*	.724**	.667**	.498	.701**	.578*	.567*	

SELF: User-design survey; means of individual MBA student choices

PRICES: Preference market results; closing stock prices

POST: Post-trading survey; mean expectations of others' preferences

*: $p < .05$, **: $p < .01$, ***: $p < .001$

Discussion

The results support the conceptual model, since in all four studies, SELF preferences are highly correlated with preference market stock PRICES, and with POST-trading expectations of others. Replications 1A, 1B and 1C, yielded Pearson correlations of 0.862, 0.819 and 0.758 (all significant at $p < 0.01$ level), respectively, between the user-design study of SELF preferences and closing stock PRICES. More remarkable were the increased correlations between stock PRICES and POST-trading expectations of others (**bolded** values) for studies 1B and 1C, respectively, of **0.880** (which is higher than 0.819) and **0.916** (which is higher than 0.758). These higher

correlations could result from either or both of two key effects, (1) learning or anchoring effects, in which traders observe stock prices and set their expectations of others accordingly, or (2) from traders' greater reliance on expectations of others when deciding which stocks to buy, and which to sell. Studies 3 and 4 tease apart these effects, and show that both are at play.

Interestingly, the highest correlation came from participants who, unlike those in Studies 1A, 1B and 1C, had not completed a user design survey of their own SELF preferences, but only of their POST-trading expectations of others (study replication 1D). The closing stock PRICES for this group had a correlation of **0.920** with the POST survey.

We note that while most of the results in Table 3, both between methods and between studies, show high correlations, some of the correlations between Study 1A and Studies 1B/1C/1D, which took place two years later, are lower. For example, Study 1A's stock prices only had a correlation of 0.491 (not significant) to Study 1B's stock prices.

4.2 Study 2: Must Traders be Sampled from the Target Market?

Study 1 revealed that preference markets work well when the survey respondents and stock traders are sampled from the same underlying population. We seek to learn how well preference markets perform when the traders come from a different population, one that may possess insight about the target market, but are not necessarily part of it. And we also want to learn whether such "experts" can proceed directly to stock trading without priming with the original survey instrument.

Study Design and Procedure

Utilizing the same 14 product feature levels as in Study 1, we recruited a group of nineteen executives who happened to be participating in a one-week program on e-Business held at a US East coast business school shortly after the time of Study 1A. They proceeded directly to trading stocks after hearing a brief description of the original user design study (but without completing the survey instrument itself) and receiving instructions regarding the user interface for the stock market.

Results and Discussion

As seen in Table 3 the high correlation of ***0.931*** between Study 2's closing stock PRICES and Study 1A's SELF preferences indicates that the executive traders who had not interacted with the 241 individual respondents nor completed a user-design study themselves, were nevertheless able to accurately predict the percentage take-up rates for the 14 feature levels. We can infer that for preference markets to be effective, traders may either be drawn from the target market itself, or from a population with insight about that market.

4.3 Study 3: Can We Increase the Number of Stocks and Question Types?

Building upon Studies 1 and 2, Study 3 explores the potential advantages of preference markets over other market research methodologies: *scalability, flexibility and learning*. Theoretically, the number of feature levels and product concepts in a preference market scales up in the number of traders. Additionally, alternative approaches to defining stocks provide flexibility to ask multiple question types

within a single study. We aim at identifying which *types* of questions provide the most useful insights about underlying preferences.

Study Design and Procedure

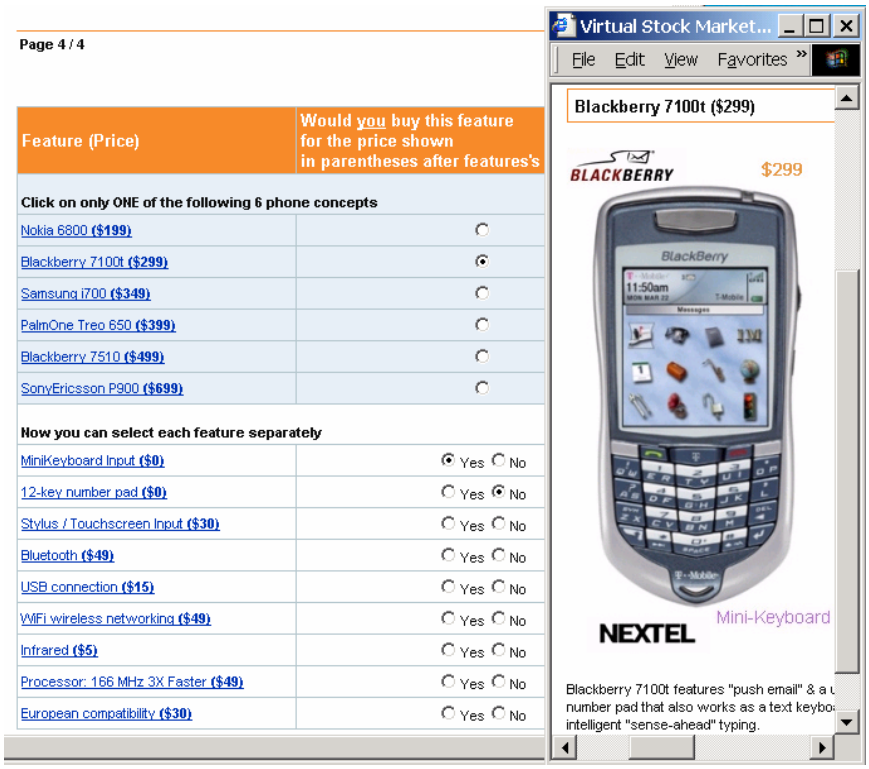
First, the questions addressed were expanded into 31 binary feature levels, 19 mutually exclusive feature levels, and 6 full phones, for a total of 56 questions, (Appendix 2). 116 MBA student respondents (a 38% response rate) were recruited to complete two surveys in advance of stock trading. In advance of trading, each participant completed (1) a SELF survey, as shown in Figure 4(a) for all 56 options, to be compared with (2) a second survey of expectations of OTHERS for 20 of the 56 options.

Binary feature levels such as "Bluetooth ($49)" were surveyed using radio buttons with only two options, "yes" or "no." "Yes," the response given by 40% of the 116 people surveyed, meant that the respondent would add Bluetooth were it priced at $49 retail. Mutually exclusive feature levels also used radio buttons, with between two to six alternatives, exactly one of which had to be selected. We note that SELF preferences were highly heterogeneous, evidenced by a median coefficient of variation of 76% across all 56 stocks (a coefficient of variation of 100% being the highest possible in this context, which would occur if 50% of respondents chose an option and the other 50% rejected it).

After the SELF survey, those respondents who would be trading a particular stock also answered a question about OTHERS' preferences, "What percentage of participants would buy this feature?" For example, among the 41 respondents who

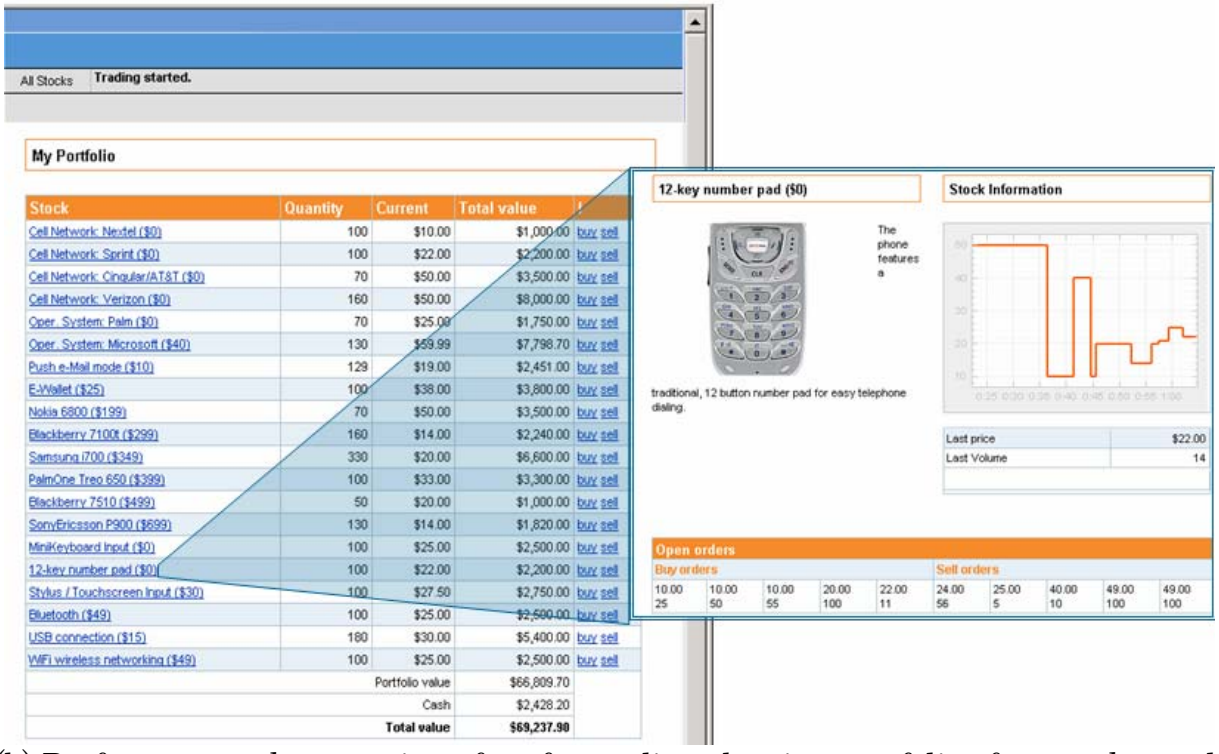
were about to trade the "Bluetooth" stock, the average answer to the question about OTHERS was 26% (s.d. 19%). After trading, a POST survey (56% response rate[3]) asked traders to provide updated estimates of others' preferences. This is the first experiment in which surveys of SELF-, OTHERS-, and POST can be compared against stock trading, enabling us to test the four hypotheses that comprise the conceptual model in Figure 1.

Figure 4: Updated Multi-Screen User Interface for Trading



(a) Survey of self preferences showing a mutually exclusive phone choice at top, and nine binary choices at bottom

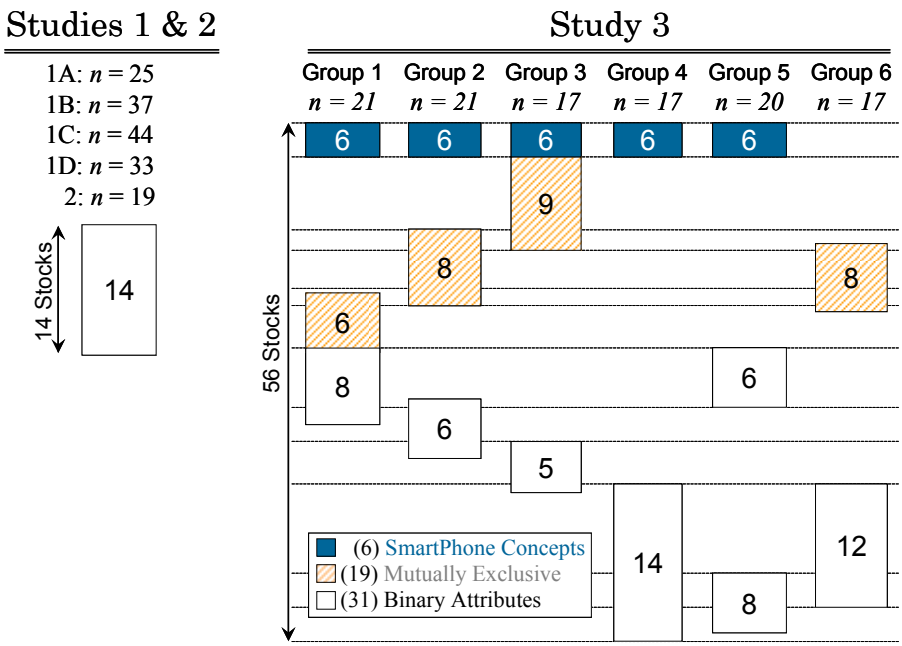
[3] No significant differences were observed between the respondents and non-respondents to the POST survey in terms of trading activity, offers to buy and sell, and trader performance.



(b) Preference market user interface for trading showing portfolio of 20 stocks, and stock detail such as name, image, pricing history, and open order book

113 of the 116 survey respondents opted to participate in the stock trading experiment on a university holiday, 93 of them in person in two classrooms, and 20 off-site logged into the market over the Internet. To test scalability, we develop an experimental design consisting of six groups of traders and six overlapping groups of stocks, as shown in Figure 5.

Figure 5: Making Preference Markets Infinitely Scalable



A continuous double auction market mechanism was implemented, as in Forsythe (1992). In such a system, a trade is executed only when a seller's asking price is at or below a buyer's bid price. No initial prices or orders were set up in advance, and all endowments within a trader group were identical: 100 shares of each of twenty securities, and $15,000 in virtual cash. The $15,000 amount was chosen to provide sufficient liquidity, while not encouraging excessive speculative behavior, and represented approximately 25% of the expected portfolio value.

The user interface, depicted in Figure 4(b), provided traders with short descriptions and images, real-time trading information, and supply- and demand data in the open order book. During the fifty minute duration of the experiment, traders attempted to maximize their respective portfolios, including the market value of all stocks and cash, by executing a total of 1,680 trades. Each stock was traded

between 5 and 150 times, confirming the effectiveness of the experimental design in scaling up in the number of traders.

As an incentive, a $50 reward was offered to the "winners," who were randomly drawn from amongst all traders within each of the six groups in Figure 5. The probability of winning depended on the trader's rank within his or her trading group, based upon total portfolio value at ending market prices, at a randomly chosen ending time, so that those who had performed well had a higher probability of winning. Further, traders with the highest total portfolio values were announced publicly, a form of recognition within this competitive peer group. These incentives were designed to induce traders to reveal their true beliefs, even if they were not performing particularly well. Additionally, an award was offered for the highest portfolio based on actual survey results using the SELF survey data.

Results

Prediction markets use closing prices since the efficient market hypothesis suggests that the most recent stock price summarizes all known information (Fama 1970), but in the context of preference markets, given their lack of external information, every offer to buy and sell expresses the belief of an individual trader, and each executed trade represents an agreement between at least two traders. So, analyzing the entire data set with volume-weighted average prices (VWAP), and not just closing prices as in studies 1 and 2 (due to data availability), captures a broader range of opinions.[4] The first four rows and columns of Table 4 summarize the

[4] All of our key results hold when using closing prices, as they are highly correlated with VWAP, but we note that VWAP slightly improves the accuracy of these results.

relationships between stock trading and the SELF-, OTHERS-, and POST surveys in Study 3 and for the next Study 4. All correlations are significant at the $p < 0.001$ level.

Table 4: Correlations based on Volume-Weighted Average Prices (VWAP) for Studies 3 & 4

n =	116	116	113	63	110	102	63	29
	3 SELF	3 OTHERS	3 PRICES	3 POST	4 SELF	4 OTHERS	4 PRICES	4 POST
3 SELF								
3 OTHERS	.880***							
3 PRICES	.622***	*.750****						
3 POST	.717***	.832***	**.924*****					
4 SELF	.653***	.677***	.723***	.651***				
4 OTHERS	.685***	.752***	.769***	.733***	.863***			
4 PRICES	.525***	.620***	.661***	.626***	.707***	*.829****		
4 POST	.560***	.647***	.741***	.696***	.714***	.837***	**.910*****	

SELF: Mean survey results for individuals' choices

OTHERS: Mean pre-trading expectations of others' choices

PRICES: Preference market results; volume-weighted average stock prices (VWAP)

POST: Mean post-trading expectations of others' choices

***: $p < .001$

A subset of 39 stocks were common to both Studies 3 and 4.

Given the highly heterogeneous preferences of the 116 respondents, we conclude from the Pearson correlation of 0.880 between the SELF-and OTHERS surveys, that respondents are accurate in estimating each other's preferences. Thus, H1a is supported, as is Figure 1's link between SELF-and OTHERS. As for H1b, the hypothesis that expectations of others are biased by self preferences, the surveys

confirms this hypothesis as well, since individuals who chose an option for themselves had higher expectations of the percentage of others who would choose that option (37% higher on average across all 56 stocks) than those who rejected an option, and those who rejected an option for themselves had 18% lower expectations, on average. These differences were significant at the $p < 0.001$ level for 46 of the 56 stocks. These biases work against overall accuracy, and yet are overcome by the "wisdom of crowds" effect (H1a) and by the market pricing mechanism [H3].

As seen in row three of Table 4, preference market PRICES correlate better with expectations of OTHERS (ρ = *0.750*) than with SELF preferences(ρ = 0.622), supporting [H3].

Figure 6: Study 3 Comparison of Three Stock Types:
Binary Aspects, Mutually Exclusive Feature Levels, and Full Products
(Overall ρ=.750 (p<.001), r^2 = 0.56)

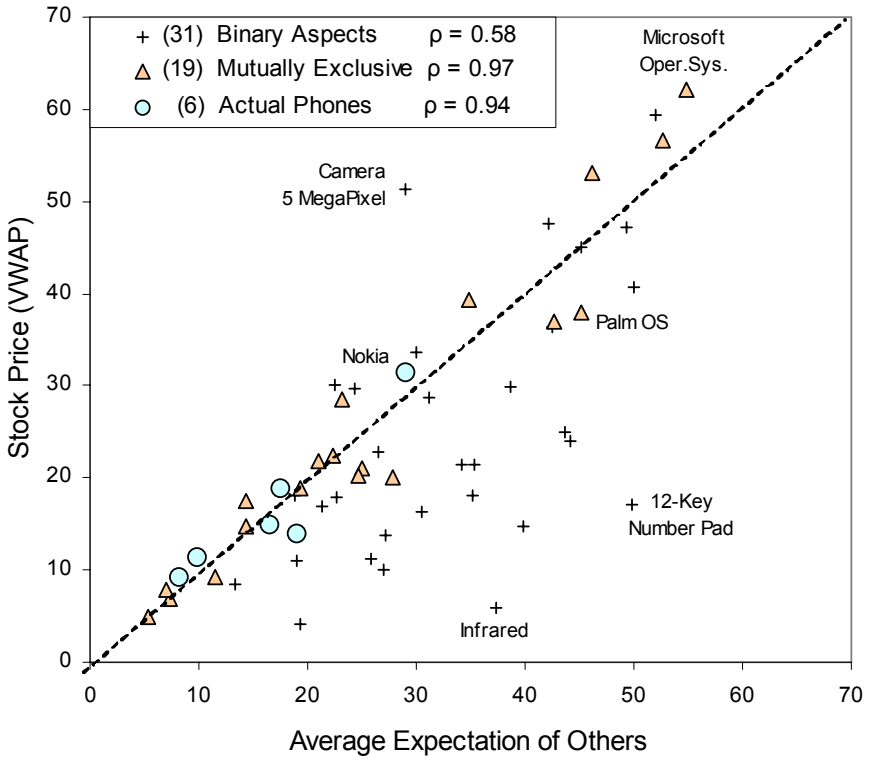


Figure 6 offers interesting insights to the effectiveness of various question types, namely that mutually exclusive questions appear to produce significantly more accurate results than binary questions. The six smart phones, shown as circles in the figure, had ρ = 0.94 ($p < .01$) between stock PRICES and the OTHERS survey. Similarly, the correlation for the nineteen mutually exclusive features was 0.97 ($p < .001$), and the best-fit line describing the twenty-five data points for these two question types is not statistically different from the 45-degree line.

For the thirty-one binary feature levels, depicted as "+"'s in Figure 6, ρ drops to 0.58 and the best fit line has a slope of 0.77 (s. e. 0.20). We note however, not only are these data points more dispersed, but that the vast majority of them lie below the 45-degree line. So not only do the binary stocks in this experiment exhibit lower predictive accuracy. Further, as shown by Payne, Bettman and Johnson (1993), respondents trade off effort versus accuracy, and binary questions may require more effort than mutually exclusive questions. Statistically, we observe a higher average coefficient of variation (0.74) for binary responses than for mutually exclusive ones (0.54). This higher variation holds in the POST survey as well.

Study 3 also supports H2, the hypothesis that traders base their stock-buying and -selling decisions upon expectations of OTHERS more so than on SELF preferences. By aggregating all of the orders to buy and sell each stock, and comparing the volume-weighted mean value of the order prices against the self and others survey, stock by stock, we observe that the correlation between ORDERS and OTHERS (ρ = 0.76 at the $p < .0001$ level) is higher than the correlation between ORDERS and SELF (ρ = 0.65 at the $p < .01$ level).

Finally, Study 3 also reveals that trading stocks results in a significant amount of learning among traders, and that H4 is supported. Specifically, traders update their beliefs about others based on the stock prices they observe, as seen in the increased ρ of **0.924** ($p < .001$) between stock PRICES and the POST survey, as compared to the lower correlation of 0.832 between PRICES and the pre-trading version (OTHERS) of the same survey. Further, the coefficient of variation in estimates of others was reduced from an average of 65% in the OTHERS survey to 55% in the POST survey across all 56 stocks (a statistically significant reduction for 40 of the 56 stocks at the $p < 0.001$ level). So, it appears that the process of trading causes participants to converge towards a consensus of opinion. The learning aspects of scalable preference markets could be particularly useful for product categories in which individual preferences are shaped by others, such as fashion goods or those with network externalities.

Discussion

Study 3 achieved both objectives: (1) preference markets were demonstrated to be *scalable* by virtue of an experimental design that matches traders with a convenient number of stocks, and creating trading links between the groups, (2) and multiple questions types were combined into a single study, with the result that mutually exclusive questions, both for features and full products, significantly outperformed binary (yes/no) features. Importantly, study 3 supports hypotheses H1a through H4, consistent with Figure 1's conceptual model of preference markets.

4.4 Study 4: How do Scalable Preference Markets Perform Under Real World Conditions?

Our final study tests how well scalable preference markets perform outside of a purely academic environment, under real world conditions. Building on the lessons from the prior study, we focus entirely on mutually exclusive feature levels, and test the limits of scalability by reducing the number of traders, most of whom trade remotely, while increasing the number of stocks. To add to the realism of the experiment, we conduct it with managers, designers, and engineers within a large firm that is directly engaged in the smart phone industry. As in study 3, we test the four hypotheses of our conceptual model.

Study Design and Procedure

With the help of an internal innovation team at the firm, 63 people participated in the on-site experiment, of whom 15% were in marketing & sales, 65% in technical positions, and 5% in finance and the remaining 15% in other functional areas. Participants reported an average of 4.5 years of industry experience and 53% claimed a management position. The experiment was conducted at the firm's corporate headquarters, with over 60% of participants accessing the market remotely from their offices, after having completed SELF- and OTHERS surveys in advance. The remote participants learned how the experiment worked through a live, 15-minute video web cast with audio questions and answers. The experiment employed the same user-interface and experimental design as Study 3. Six groups were formed, ranging in size from 5 to 15 traders, with 21 to 24 stocks each (Figure 2). While Study 3 had approximately two traders per stock (113 traders, 56 stocks),

Study 4 was twice as intensive, with an average of only *one* trader per stock (63 traders, 64 stocks). In discussion with the firm's executives we defined 64 mutually exclusive stocks, 39 of which could be compared against those in Study 3, and 25 of which were new, and included recent advances and features of interest to the firm.

Results

Referring to the fifth through eighth rows and columns of Table 4, the correlation of 0.863 between the SELF-and OTHERS surveys confirms that respondents are accurate in estimating each other's preferences, supporting H1 as in Study 3. Again, preference market PRICES capture expectations of OTHERS ($\rho = .829$) better than SELF preferences ($\rho = .707$), supporting H2 and H3. For the 29 of respondents (46% response rate[5]) who completed the POST survey, the hypothesis that they learned from trading stocks, H4, is supported by the higher ρ of **0.910** between the POST survey and STOCK prices (as compared with $\rho = .829$), and by the reduction in the average coefficient of variation (c.v.) from 68% for the OTHERS survey to 41% in the POST survey. In fact, learning in the form of statistically significantly reduced coefficients of variation for 61 of the 64, or 95% of the smart phone features.

For the 39 stocks common to Studies 3 and 4, the students and firm participants diverged somewhat in their SELF preferences ($\rho =$ only .661 between Study 3 and Study 4), so it is not surprising that Study 4's STOCK prices were weaker predictors of Study 3's SELF preferences ($\rho = .525$) than they were of study 4's SELF preferences ($\rho = .707$). Differences between the groups' results may be due to time-varying

[5] There are no significant differences between the 29 individuals who responded to the POST survey and the 34 who did not in the amount of trading activity, offers to buy and sell, and performance.

preferences (Studies 3 and 4 took place 20 months apart), differences in how stocks were defined, and the distinction between students and professionals. Considering all of these differences, we are encouraged by the degree of convergent validity between Studies 3 and 4.

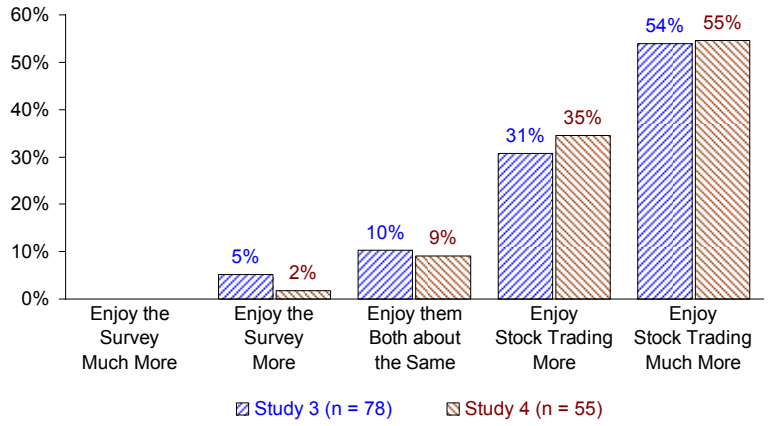
Discussion

In effect, Study 4 was a real-world replication of Study 3. First, it demonstrated that scalable preference markets perform well in the field, with managers and employees trading in an efficient manner. Specifically, we learned that the majority of traders mastered the user interface and were able to trade remotely from their offices. Further, the high ratio of one stock-per-trader was still sufficient to achieve accurate results. Mitigating the high stock-per-trader ratio and remote participation rate were the high level of participants' market expertise and the use of easier, mutually exclusive questions. Study 4 produced remarkable results in a very short time, with fewer people, over a very larger number of questions. The wisdom-of-crowds-, expertise-aggregation-, and learn-from-trading effects were all evident.

4.5 Which to employ, surveys or scalable preference markets?

In addition to the POST survey questions about updated expectations of OTHERS, we asked respondents in Studies 3 and 4 (69% and 87% response rates, respectively) about their relative preference between surveys and stock trading. The results are shown in Figure 7.

Figure 7: Which Method Did Respondents Prefer: Survey or Stock Trading?



There was near-unanimity in preference for stock trading over surveys. Scalable preference markets differ from surveys in that they include elements of competition, interaction, gaming, learning, and the opportunity to gain recognition and win prizes, which might explain the strong result. We would expect that the cost of recruiting and compensating respondents would be lower for stock trading than for individual markets research surveys, given the same quantity of research questions (see Appendix 3). In addition, 75% of the industry experts in Study 4 expressed a willingness to participate in a preference market again.

5 Discussion

We developed a conceptual model of preference markets and tested a scalable version of the method that worked well in practice. Through seven stock trading experiments within four studies, we validated the model that SELF preferences influence expectations of OTHERS, which in turn are reflected in stock PRICES. Of course, reverse causality, in which one's expectations of others' tastes and preferences may help form self preferences, may also explain some of our results. Were that the case, measuring expectations of others would be all the more

important. But given that we observed high variation between traders in their expectations of others it seems likely that individuals have more confidence in their self preferences than in their expectations of others, so that causality is more likely to be from SELF to OTHERS.

Our results suggest that scalable preference markets offer an effective tool for product development teams, especially when large numbers of design decisions need to be prioritized. For example, the top 5-10 stocks may merit further study via conjoint analysis. The number of features and concepts that can be tested scales in the number of traders, with one trader per stock representing a minimum. Respondents express a strong preference for trading stocks over answering surveys. And they learn from *each other* while trading, updating their expectations in a way that converges towards a clearer consensus.

Our research demonstrates that scalable preference markets have the potential to reduce the costs of recruiting and compensating respondents, as sample sizes were reduced, a ratio of one question per participant was achieved, and respondents expressed a high willingness to participate, even with minimal compensation. Appendix 3 illustrates how these improvements impact respondent-related costs.

Despite these promising results, some issues remain: external validity, comparison with conventional methods and directions for future research.

5.1 External Validity

Validating methods with actual, external data poses a challenge in new product development research, as many of the ideas tested may not exist. And even in the case of existing features and concepts, access to accurate data may be limited. Instead, we look to new product releases and comparisons to prior studies for at least some degree of validation of the accuracy of our results.

Looking across all seven experiments, several clear trends emerge in the data. Table 5 shows that five smart phone traits were preferred by the majority, even at a price premium, in virtually every survey and preference market, and can be interpreted as "must have" features, while ten aspects were consistently rejected by over two thirds of respondents, and represent low-priority, or niche, design considerations. From a market research perspective, the features in the middle represent differentiation opportunities that may merit further study. Scalable preference markets facilitate "triage"; design teams may prioritize opportunities and focus their product development efforts.

Table 5: "Triage" of Smart Phone Preferences

Preferred by a Majority	Heterogeneous Preference	Rejected by a Majority
Small Size & Weight (3-4")	Oper. System (Microsoft rising)	Hands Free Operation
Color Display (320x240+)	Memory Capacity & Battery Life	Bluetooth, Infrared, USB
Camera (quality rising)	Mini-Keybd. vs. 12-key vs. Stylus	GPS (but rising)
Verizon Cell Network	WiFi Capability and Push Email	FM radio, Video Camera
Black or Silver Phone	Slot types (SD rising)	Changeable Faceplates
	MP3 vs. TV	European Compatibility
	Phone Brands and Models	e-Wallet

Table 5 also presents an interesting example of external validity, in that one would expect leading smart phone manufacturers to launch new products conforming to

these results. And, as shown in Figure 8, Nokia, Motorola, and BlackBerry launched smart phones in 2006 that largely fit the table and appear to be converging towards a dominant design.

Figure 8: Three New Smart Phones Launched in 2006

We compare stock prices from Studies 3 and 4 against self-stated preferences for 15 and 11 features, respectively, from two individual surveys conducted independently of the present research. The 2004 study, with 518 MBA student respondents, is forthcoming in a leading journal, and the more recent study is part of a working paper. Both focus on new methods of conjoint analysis. The ρ's, which range from 0.714 to 0.885, relate the stock market results to the external studies, and provide further evidence of external validity. We note that in Table 6, Study 3's stock trading, also conducted with MBA students, correlates better with the external data than does Study 4, which was conducted with industry experts.

Table 6: Evidence of External Validity
(Pearson correlations between stock results and two independent studies, 14 and 11 attributes compared)

	3 PRICES	4 PRICES
2004 Study (n=518)	.802**	.714*
2005 Study (n=206)	.885**	.769*

**: p <.05, *: p <.10

5.2 Comparison of Preference Markets with Conventional Methods, Limitations and Directions for Future Research

Table 7 compares preference markets with other methods, and highlights their scalability and consensus orientation. Preference markets complement other methods by narrowing a large number of potential product features and concepts to a manageable set that can be further analyzed at the individual level using the other approaches. A limitation of scalable preference markets is that they do not measure preference heterogeneiy. Our results demonstrate that markets achieve a consensus about expectations of average preferences, and do not provide insight about distinct individuals. To measure heterogeneity, methods such as conjoint analysis are better suited to the task. Of course, one could devise markets that attempt to estimate preferences at the market segment level by cleverly defining stocks and trader groups using the approach taken in Figure 5, but this would still be a blunt instrument for measuring detailed differences at a more granular level.

Table 7: Comparison with conventional methods

	User Design	*Conjoint Analysis*	*Self Explicated*	*Preference Markets*
Description	Individuals customize optimal products	Individuals rate-, rank- or choose feature bundles	Individuals rate importances of *unbundled* features	Trader groups achieve consensus through trading
Advantages	Identifies optimal feature bundles from a large number of combinations; engaging task	Quantifies trade offs over a limited number of features; measures individual utility	Quantifies individual trade offs over more features; easier task	Measures consensus preferences over many features and concepts; scalable; engaging, fun task
Disadvantages	Does not measure trade-offs; setup costs can be high	Task difficulty, inconsistent answers, complex post analysis	Potential problem of "everything is important"	No individual preferences; need for simultaneous participation
Best Fit Applications	Customized goods; identification of ideal feature bundles; Go/No Go on features	Optimizing design, pricing, positioning over a narrow number of decisions	When conjoint is too difficult or costly, or number of features is too high	Narrowing many options to a few; achieving consensus; when speed matters;

Logistically, one challenge of preference markets is coordinating *simultaneous* participation. If traders cannot connect in real time, a "market maker" might be necessary, or a longer trading horizon in which orders do not clear within seconds or minutes. But this could reduce the engagement level of participants. One might ask whether our dynamic market mechanism is necessary at all. Might not the OTHERS survey suffice? The answer is that such a survey could be used to prioritize a long list of potential product features and concepts, but one would lose the benefits of interaction, competition, and learning. And we would expect that significantly more respondents would need to be recruited and compensated, raising the cost and time required. More importantly, preference markets scale up in the number of respondents much more easily than would the survey of OTHERS.

Implementation in firms poses its own challenges. The firm must develop or invest in trading software and infrastructure. Respondents need to be taught the mechanics of trading and the underlying meaning of each stock. The key outcomes, the stock prices themselves, become known to all traders immediately, so data security may pose a problem. And the market mechanism itself pulls no punches: the consensus view, whether positive or negative, becomes instantly transparent. Champions of specific product ideas may not readily accept negative outcomes, a challenge with any market research, but one which might be exacerbated by the immediacy of preference markets.

Even after overcoming objections about validity, aggregate measurement, the need for simultaneity, and internal politics, optimizing preference markets will require

further research. We hope that fellow researchers will experiment with the design options listed in Appendix 1, so that the accuracy and applicability of scalable preference markets can be improved. Specifically, future research might study: (a) new stock types such as "no buy" or "price-elasticity" stocks (i.e., product and feature take-up rates at varying retail prices), (b) the ability to buy and sell additional shares from a "bank," short-selling, options trading, and other exotic alternatives, (c) new incentives and rewards, (d) better tests of internal and external validity, (e) the impact of product-related "news" revealed during trading, and (f) the connection between individual trading behavior and individual preference.

Considering the accuracy, scalability, flexibility, speed, and attractiveness to respondents of preference markets, we anticipate that the methodology will gain adherents over time, enabling firms and their product development teams to prioritize the features and concepts that address the consensus opinions of the market.

References

Benartzi, Shlomo and Richard H. Thaler (2001), "Naive Diversification Strategies in Defined Contribution Saving Plans," *American Economic Review*, 91 (1), 79-98.

Blanchard, Olivier J. and Mark W. Watson (1982), "Bubbles, Rational Expectations and Financial Markets," *NBER Working Paper No. W0945*.

Bondarenko, Oleg and Peter Bossaerts (2000), "Expectations and Learning in Iowa, " Journal f Banking and Finance, 24, 1535-1555.

Camerer, Colin, George F. Loewenstein, and Martin Weber (1989), "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *The Journal of Political Economy*, 97 (5), 1232-54.

Chen, Kay-Yut and Charles R. Plott (2002), "Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem," *California Institute of Technology*.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales," *Journal of Marketing Research*, forthcoming.

Cooper, Robert G. (1999), "The Invisible Success Factors in Product Innovation," *Journal of Product Innovation Management*, 16, 115-33.

Cowgill, Bo (2006), "Putting Crowd Wisdom to Work," (accessed 10/10/2006), [available at http://googleblog.blogspot.com/2005/09/putting-crowd-wisdom-to-work.html].

Dahan, Ely and John R. Hauser (2002), "The Virtual Customer," *Journal of Product Innovation Management*, 19 (5), 332-53.

Dahan, Ely, Andrew W. Lo, Tomaso Poggio, Nicholas T. Chan, and Adlar Kim (2006), "Securities Trading of Concepts (STOC)," University of California at Los Angeles, Working Paper.

Fama, Eugene 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.

Forsythe, Robert, Forrest Nelson, George R. Neumann, and Jack Wright (1992), "Anatomy of an Experimental Political Stock Market," *American Economic Review*, 82 (5), 1142-61.

Forsythe, Robert, Thomas R. Palfrey, and Charles R. Plott (1982), "Asset Valuation in an Experimental Market," *Econometrica*, 50 (3), 537-68.

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

Gilbride, Timothy J. and Greg M. Allenby (2004), "A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules," *Marketing Science*, 23 (3), 391-406.

Guarnaschelli, Serena, Anthony M. Kwasnica, and Charles R. Plott (2003), "Information Aggregation in Double Auctions: Rational Expectations and the Winner's Curse," *Information Systems Frontiers*, 5 (1), 61–75.

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

Hayek, Friedrich August von (1945), "The Use of Knowledge in Society," *American Economic Review*, 35 (4), 519-

Hoch, Stephen J. (1987), "Perceived Consensus and Predictive Accuracy: The Pros and Cons of Projection," *Journal of Personality and Social Psychology*, 53 (2), 221-34.

---- (1988), "Who Do We Know: Predicting the Interests and Opinions of the American Consumer," *Journal of Consumer Research*, 15 (3), 315-24.

Huber, Joel, Dick Wittink, John A. Fiedler, and Richard Miller (1993), "The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice," *Journal of Marketing Research*, 30 (1), 105-14.

Huberman, Gur and Paul Sengmueller (2004), "Performance and Employer Stock in 401(k) Plans," *Review of Finance*, 8, 403-43.

Jamieson, Linda F. and Frank M. Bass (1989), "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods," *Journal of Marketing Research*, 26 (3), 336-45.

Kivetz, Ran, Oded Netzer, and V. Srinivasan (2004), "Alternative Models for Capturing the Compromise Effect," *Journal of Marketing Research*, 41 (3), 237-57.

Liechty, John, Venkatram Ramaswamy, and Steven Cohen (2001), "Choice-Menus for Mass Customization: An Experimental Approach for Analyzing Customer Demand With an Application to a Web-Based Information Service," *Journal of Marketing Research*, 38 (2), 183-96.

Lorge, I., D. Fox, J. Davitz, and M Brenner (1958), "A Survey of Studies Contrasting the Quality of Group Performance and Individual Performance, 1920-1957," *Psychological Bulletin*, 55 (6), 337-72.

Lucas, Robert E. (1972), "Expectations and the Neutrality of Money'," *Journal of Economic Theory*, 4 (2), 103-24.

MacKinlay, Craig M. (1997), "Event Studies in Economics and Finance," *Journal of Economic Literature*, 35 (3), 13-39.

Oliven, Kenneth and Thomas A. Rietz (2004), "Suckers Are Born but Markets Are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market," *Management Science*, 50 (3), 336-51.

Park, C. Whan, Sung Youl Jun, and Deborah J. MacInnis (2000), "Choosing What I Want versus Rejecting What I Don't Want: An Application of Decision Framing to Product Option Choice Decisions," *Journal of Marketing Research*, 37 (5), 187-202.

Payne, John W., James R. Bettman, and Eric J. Johnson (1993), The Adaptive Decision Maker: Cambridge University Press.

Pennock, David M., Steve Lawrence, Lee C. Giles, and Finn Arup Nielsen (2001), "The Real Power of Artificial Markets," *Science*, 291 (5506), 987-88.

Plott, Charles R. and Shyam Sunder (1982), "Efficiency of Experimental Security Markets with Insider Information: An Application of Rational Expectations Models," *Journal of Political Economy*, 90 (4), 663-98.

---- (1988), "Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets," *Econometrica*, 56 (5), 1085-118.

Proebsting, Todd (2005), "Tee Time with Admiral Poindexter. Microsoft Research," *Presentation at DIMACS Workshop on Markets as Predictive Devices (Information Markets)*.

Randall, Taylor, Christian Terwiesch, and Karl T. Ulrich (2006), "User Design of Customized Products," *Marketing Science*, forthcoming.

Salganik, Matthew, J., Peter Sheridan Dodds, and Duncan J. Watts (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," Science, 311, 854-856.

Sawhney, Mohanbir, Gianmario Verona, and Emanuela Prandelli (2005), "Collaborating to Create: The Internet as a Platform for Customer Engagement in Product Innovation," *Journal of Interactive Marketing*, 19 (4), 4-17.

Sawtooth Software, Inc. (1999), *CBC User Manual*. Sequim (WA).

Servan-Schreiber, Emile, Justin Wolfers, David M. Pennock, and Brian Galebach (2004), "Prediction Markets: Does Money Matter?," *Electronic Markets*, 14 (3), 243-51.

Shugan, Steven M. (1980), "The Cost of Thinking," *Journal of Consumer Research*, 7 (2), 99-111.

Simon, Herbert A. (1955), "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, 69, 99-118.

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

Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams (1988), "Bubbles, Crashes and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, 56, 1119-51.

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

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

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

Thompson, Debora Viana, Rebecca W. Hamilton, and Roland T. Rust (2005), "Feature Fatigue: When Product Capabilities Become Too Much of a Good Thing," *Journal of Marketing Research*, 42 (4), 431-42.

Timmermann, Allan G. (1993), "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices," *Quarterly Journal of Economics*, 108 (4), 1135-45.

Toubia, Olivier (2006), "Idea Generation, Creativity, and Incentives," *Marketing Science*, 25 (5), 411-25.

Toubia, Olivier, Duncan I. Simester, John R. Hauser, and Ely Dahan (2003), "Fast Polyhedral Adaptive Conjoint Estimation," *Marketing Science*, 22 (3), 273-303.

van Bruggen, Gerrit H., Gary L. Lilien, and Manish Kacker (2002), "Informants in Organizational Marketing Research: Why Use Multiple Informants and How to Aggregate Responses," *Journal of Marketing Research*, 39 (4), 469-78.

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

---- (2004b), "Using Markets to Evaluate Policy: The Case of the Iraq War," *Stanford GSB and NBER, Working Pa*

Yee, Michael, Ely Dahan, John R. Hauser, and James Orlin (2006), "Greedoid-Based Noncompensatory Inference," *Marketing Science*, forthcoming.

Appendix 1: Design of Preference Markets

Marketing Research Objective

Design of Virtual Securities

Stock definition (Stimuli)
- Product features
- Product concepts or real products

Stock structure
- Binary
- Mutually exclusive
- Bundles

Include price for the option or not

Experimental Design

Respondents
- How many respondents?
- Which population: Convenient sample, consumers, insiders, outsiders, users, experts, managers
- Open or closed access
- How many repetitions?
- Remote access or central location?

Assignments of stocks
- Random
- Self selection / Rule-based
- Sorting / Filtering

Information display
- Allow outside information
- Show last price, quantities, volume, order book, past data, rankings

Market Mechanism Design

Trading actions
- Limit orders, market orders, restrictions
- Short selling, options
- Position limits and price limits
- Trading fees or no trading fees

Trading mechanism
- Call auction
- Market maker (i.e., dealer)
- Pari-mutuel engine
- Double auction

Duration, Trading hours
- Initial conditions
- Initial prices
- Endowment
- "Bank"

Incentives

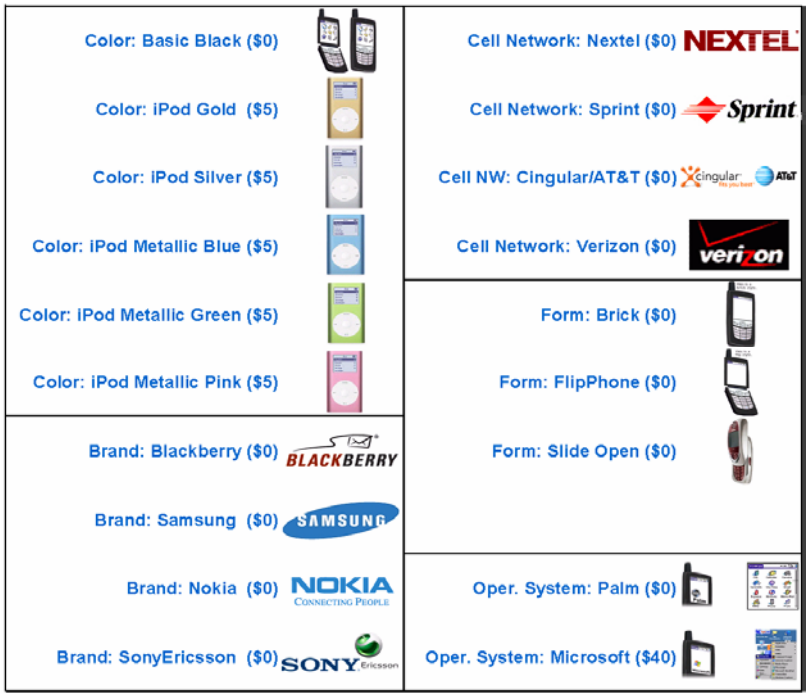
Reward type
- Non-monetary (Recognition, enjoyment)
- Monetary rewards

Rewarding rules
- Not based on performance: everyone, random sample
- Performance based: tournament, sample, everyone, proportional to the portfolio value

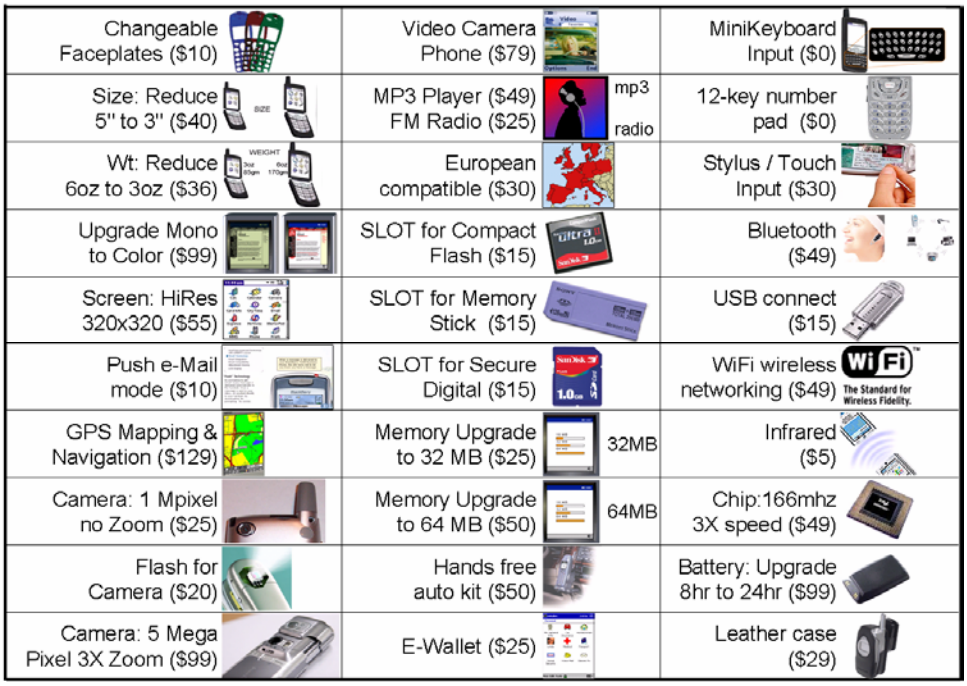
Rewarding
- Accuracy: best predictors
- Effort: trading behavior

Winner determination
- Actual prices
- Other replications
- "truth"
- Exogenous: parallel market research

Appendix 2: Three Question-types for Study 3



(19) Mutually Exclusive Smartphone Feature Levels and (6) Mutually Exclusive Smart Phones
(each of the 6 categories sums to 100%)



(31) Binary Smartphone Feature Levels
(each garners between 0% and 100% "share" at the feature price shown)

Appendix 3: Respondent-related Costs of Marketing Research

Frequently, the cost of recruiting and compensating survey respondents represents the single largest component of overall market research expense. Consider the following simplified equation characterizing the total cost (*TC*) of recruiting and compensating N_{sample} respondents to answer a minimum of Q_{min} new product-related questions:

$$\text{(1)}\quad \begin{aligned} Total\ Cost &= \overbrace{\left(\begin{matrix}Number\ of\\ respondents\ needed\end{matrix} \times \frac{recruits}{respondent} \times \begin{matrix}Cost\ per\\ recruit\end{matrix}\right)}^{\text{Recruiting Costs}} + \overbrace{\left(\begin{matrix}Number\ of\\ respondents\ needed\end{matrix} \times \begin{matrix}Compensation\\ per\ respondent\end{matrix}\right)}^{\text{Respondent Compensation}} \\ &= \begin{matrix}Number\ of\\ respondents\ needed\end{matrix} \times \left(\frac{Cost\ per\ recruit}{response\ rate} + \begin{matrix}Compensation\\ per\ respondent\end{matrix}\right) \\ &= \overbrace{\left(\begin{matrix}\#\ Questions\\ being\ asked\end{matrix} \times \begin{matrix}Sample\ size\\ required\end{matrix} \div \frac{question\ capacity}{respondent}\right)}^{\text{Number of respondents needed}} \times \overbrace{\left(\frac{Cost\ per\ recruit}{response\ rate} + \begin{matrix}Compensation\\ per\ respondent\end{matrix}\right)}^{\text{Average cost per respondent}} \end{aligned}$$

$$\text{(2)}\qquad TC = \left(Q_{\min} \cdot \frac{N_{sample}}{q_{respondent}}\right) \times \left(\frac{c_{recruit}}{r\%} + c_{respondent}\right),$$

Where N_{sample} is the minimum number of people required for statistically valid results, $q_{respondent}$ is the capacity of each respondent to answer questions, $r\%$ is the response rate, i.e. the percentage of potential recruits who choose to participate in the study, and $c_{recruit}$ and $c_{respondent}$ are the respective costs of recruiting and compensating each person.

The equation reveals that, given Q_{min} questions that need to be answered, the respondent-related costs of a study can be reduced in any of five ways: (a) reducing the minimum sample size needed for valid results (reducing N_{sample}) (b) increasing the number of questions each respondent can answer (increasing $q_{respondent}$), (c) lowering the cost of recruiting each potential respondent (lowering $c_{recruit}$), (d) improving the response rate (increasing $r\%$), or (e) reducing the required compensation per respondent (reducing $c_{respondent}$).