EXPERIMENTAL BUSINESS RESEARCH

Experimental Business Research

Marketing, Accounting and Cognitive Perspectives

VOLUME III

Edited by

RAMI ZWICK
Hong Kong University of Science and Technology, China

and

AMNON RAPOPORT
University of Arizona, Tucson, U.S.A.
and
Hong Kong University of Science and Technology, China

Springer

A C.I.P. Catalogue record for this book is available from the Library of Congress.

ISBN-10 0-387-24215-5 (HB) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-10 0-387-24244-9 (e-book) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-13 978-0-387-24215-6 (HB) Springer Dordrecht, Berlin, Heidelberg, New York
ISBN-13 978-0-387-24244-6 (e-book) Springer Dordrecht, Berlin, Heidelberg, New York

---

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Printed on acid-free paper

All Rights Reserved
© 2005 Springer

No part of this work may be reproduced, stored in a retrieval system, or transmitted
in any form or by any means, electronic, mechanical, photocopying, microfilming, recording
or otherwise, without written permission from the Publisher, with the exception
of any material supplied specifically for the purpose of being entered
and executed on a computer system, for exclusive use by the purchaser of the work.

Printed in the Netherlands.

Contents

PREFACE

Rami Zwick
Hong Kong University of Science and Technology

Amnon Rapoport
University of Arizona
And
Hong Kong University of Science and Technology

This volume (and volume II) includes papers that were presented at the Second Asian Conference on Experimental Business Research held at the Hong Kong University of Science and Technology (HKUST) on December 16–19, 2003. The conference was a follow up to the first conference that was held on December 7–10, 1999, the papers of which were published in the first volume (Zwick, Rami and Amnon Rapoport (Eds.), (2002) *Experimental Business Research*. Kluwer Academic Publishers: Norwell, MA and Dordrecht, The Netherlands). The conference was organized by the Center for Experimental Business Research (cEBR) at HKUST and was chaired by Amnon Rapoport and Rami Zwick. The program committee members were Paul Brewer, Kenneth Shunyuen Chan, Soo Hong Chew, Sudipto Dasgupta, Richard Fielding, James R. Frederickson, Gilles Hilary, Ching-Chyi Lee, Siu Fai Leung, Ling Li, Francis T Lui, Sarah M Mcghee, Fang Fang Tang, Winton Au Wing Tung and Raymond Yeung. The papers presented at the conference and a few others that were solicited especially for this volume contain original research on individual and interactive decision behavior in various branches of business research including, but not limited to, economics, marketing, management, finance, and accounting.

The following introduction to the field of Experimental Business Research and to our center at HKUST replicates the introduction from Volume II. Readers familiar with the introduction to Volume II are advised to skip Sections 1 and 2 below.

1. THE CENTER FOR EXPERIMENTAL BUSINESS RESEARCH

The Center for Experimental Business Research (cEBR) at HKUST was established to serve the needs of a rapidly growing number of academicians and business leaders in Hong Kong and the region with common interests in experimental business research. Professor Vernon Smith, the 2002 Nobel laureate in Economics and a current member of cEBR's External Advisory Board, inaugurated the Center on September 25, 1998, and since than the Center has been recognized as the driving force behind experimental business research conducted in the Asia-Pacific region. The mission of cEBR is to promote the use of experimental methods in business research, expand experimental methodologies through research and teaching, and apply these methodologies to solve practical problems faced by firms, corporations, and governmental agencies. The Center accomplishes this mission through three agendas: research, education, and networking and outreach programs.

2. WHAT IS EXPERIMENTAL BUSINESS RESEARCH?

Experimental Business Research adopts laboratory based experimental economics methods to study an array of business and policy issues spanning the entire business domain including accounting, economics, finance, information systems, marketing and management and policy. "Experimental economics" is an established term that refers to the use of controlled laboratory-based procedures to test the implications of economic hypotheses and models and discover replicable patterns of economic behavior. We coined the term "Experimental Business Research" in order to broaden the scope of "experimental economics" to encompass experimental finance, experimental accounting, and more generally the use of laboratory-based procedures to test hypotheses and models arising from research in other business related areas, including information systems, marketing, and management and policy.

Behavioral and experimental economics has had an enormous impact on the economics profession over the past two decades. The 2002 Nobel Prize in Economics (Vernon Smith and Danny Kahneman) and the 2001 John Bates Clark Medal (Matthew Rabin) have both gone to behavioral and experimental economists. In recent years, behavioral and experimental research seminars, behavioral and experimental faculty appointments, and behavioral and experimental PhD dissertations have become common at leading US and European universities.

Experimental methods have played a critical role in the natural sciences. The last fifteen years or so have seen a growing penetration of these methods into other established academic disciplines including economics, marketing, management, accounting and finance, as well as numerous applications of these methods in both the private and public sectors. cEBR is active in introducing these methodologies to Hong Kong and the entire Pacific Basin. We briefly describe several reasons for conducting such experiments.

First and most important is the used of experiments to design institutions (i.e., markets) and for evaluating policy proposals. For example, early experiments that studied the one-price sealed bid auction for Treasury securities in the USA helped motivate the USA Treasury Department in the early 1970 to offer some long-term bond issues. Examples for evaluating policy proposals can be found in the area of voting systems, where different voting systems have been evaluated experimentally in terms of the proportion of misrepresentation of a voter's preferences (so- called "sophisticated voting"). In the past decade, both private industry and governmental agencies in the USA have funded studies on the incentives for off-floor trading in continuous double auction markets, alternative institutions for auctioning emissions permits, and market mechanisms for allocating airport slots and the FCC spectrum auction. More recently, Hewlett-Packard has used experimental methods to evaluate contract policy in areas from minimum advertised price to market development funds before rolling them out to its resellers, and Sears used experimental methods to develop a market for logistics.

Second, experiments are used to test a theory or determine the most useful competing theories. This is accomplished by comparing the behavioral regularities to the theory's predictions. Examples can be found in the auction and portfolio selection domains. Similarly, business experiments have been conducted to explore the causes of a theory's failure. Examples are to be found in the fields of bargaining, accounting, and the provision of public goods.

Third, because well-formulated theories in most sciences tend to be preceded by systematically collected observations, business experiments are used to establish empirical regularities as a basis for the construction of a new theory. These empirical regularities may vary considerably from one population of agents to another, depending on a variety of independent variables including culture, socio-economic status, previous experience and expertise of the agents, and gender.

Finally, experiments are used to compare environments, using the same institution, or comparing institutions, while holding the environment constant.

3. CONTENT

Whereas Volume II contains papers under the general umbrella of economic and managerial perspectives, the present volume includes papers from the fields of Marketing, Accounting, and Cognitive Psychology. Volume III includes 14 chapters. The 33 contributors come from many of the disciplines that are represented in a modern business school.

Chapter 1 by Zhao, Meyer, and Han explores consumers' ability to optimally anticipate the value they will draw from new product features that are introduced to enhance the performance of existing technologies. The research is motivated by the common observation that consumers frequently purchase more technology than they can realistically make use of. Central to their work is the idea that a general over-buying bias may, in fact, have a strong theoretical basis. Drawing on prior work in affective forecasting, they hypothesize that when buying new technologies consumers will usually have a difficult time anticipating how they will utilize a product after it is purchased, and will be prone to believe that the benefits of attribute innovations that are perceived now will project in a simple fashion into the future. Implicit to this over-forecast is a tendency to underestimate the impact of factors that may likely serve to diminish usage in the future, such as frustration during learning and satiation. Consequently, there is a tendency for consumers to systematically evaluate product innovations through rose-colored glasses, imagining that they will have a larger and more positive impact on the future lives than they most often will likely end up having. This general hypothesis is tested in the context of a computer simulation in which subjects are trained to play one of three different forms of an arcade game where icons are moved over a screen by different forms of tactile controls. Respondents are then given the option to play a series of games for money with either their incumbent game platform or pay to play with an alternative version that offers an expanded set of controls. As hypothesized, subjects displayed an upwardly-biased valuation for the new sets of controls; adopters underutilized them and displayed a level of game performance that was not better than those who never upgraded. A follow-up study designed to understand the process underlying the bias indicated that while adopters over-forecasted the degree to which they would make use of the new control, they did not over-forecast performance gains. Hence, the key driver of adoption decisions appeared to be an exaggerated belief in the hedonic pleasure that would be derived from owning and utilizing the new control as opposed to any objective value it might provide.

What is notable about their results is that the evidence for the optimism bias was derived from a context designed to facilitate rational assessments of innovation value. Specifically, subjects were given a clearly-stated metric by which the objective value of the innovation could have been assessed, there was a direct monetary penalty for overstating value (the game innovation was paid for by a point deduction), and the innovation itself was purely functional rather than aesthetic (a new control added to the same graphic game platform). Yet, subjects still succumbed to the same biases.

Chapter 2 by Kim and Waller reports on a behavioral accounting experiment on strategic interaction in a tax compliance game. The experiment employed a three-step approach. First, subjects were assigned to the opposing roles of auditor and strategic taxpayer. This step addressed a past criticism of behavioral accounting research: economic mechanisms such as the interaction of players with conflicting preferences potentially eliminate the decision biases found in individual settings. Second, the experiment operationalized a game-theoretic model of the tax compliance problem by Graetz, Reinganum, and Wilde. In the model, the taxpayer chooses a strategy $\{\alpha, 1 - \alpha\}$ when true income is high, whereby he under-reports income with probability α and honestly reports income with probability $1 - \alpha$. The auditor chooses a strategy $\{\beta, 1 - \beta\}$ when reported income is low, whereby she conducts a costly audit with probability β and does not audit with probability $1 - \beta$. The model assumes two types of taxpayer: proportion ρ of strategic taxpayers who maximize expected wealth, and proportion $1 - \rho$ of ethical taxpayers who adhere to an internalized norm for honesty. The auditor maximizes expected net revenue, i.e., tax plus fine minus audit cost. Before conducting an audit, the auditor cannot distinguish between the taxpayer types. When the auditor conducts an audit and detects under-reporting, the taxpayer must pay a fine plus the tax for high true income. The model implies that the optimal audit rate β^* is insensitive to an exogenous change in ρ, as long as ρ exceeds a threshold. The strategic taxpayer fully absorbs the change in ρ by adjusting the optimal rate of under-reporting income α^*. Third, the experiment manipulated two variables that are considered irrelevant by the game-theoretic model, i.e., the level of ρ and uncertainty about ρ, in order to test hypotheses about auditors' choice of the audit rate, β.

Contrary to the model, Kim and Waller hypothesized that an auditor with limited rationality will use ρ as a cue for adjusting β. The hypotheses assume a simple additive process: $\beta = \beta' + \beta''$, where β' depends on ρ, and β'' depends on a belief about the taxpayer's strategy. The results show positive associations between ρ and β', and between auditors' uncertainty about ρ and β'. The auditors formed incorrect beliefs about the taxpayers' responses, which affected β''. The auditors incorrectly believed that the taxpayers increased the rate of under-reporting income as ρ increased, and that the taxpayers expected a higher audit rate when the auditors faced uncertainty about ρ. The taxpayers correctly believed that β increased as ρ increased, and responded by decreasing the rate of under-reporting income.

Chapter 3 by Bodoff, Levevq, and Zhang explores the beliefs that underline policies such as the SEC's Fair Disclosure Rule, and technologies such as SEC EDGAR, that aim to disseminate corporate disclosures to a wider audience.

Rational expectations models have been successful in predicting equilibrium prices in experimental markets of risky assets. In previous work, the authors explored whether such models are also useful in their other predictions regarding welfare in the sense of *ex ante* expected utility. They previously found that they are not, i.e. that subjects did not prefer the predicted market condition. In particular, when subjects could select the environment in which to trade, and the environment was characterized by the proportion of informed traders, subjects' preference for the fraction of informed traders was "Half > None > All", i.e. investors most favored a situation where a random half of investors are informed. Analytical predictions based on theories of non-revealing and full-revealing prices would predict a different preference order: "None > All > Half". In this chapter, the authors explore the tension between the correct predictions of the equilibrium solution and the incorrect predictions of subjects' preferences. In analytical models, predictions of EU follow by definition from the equilibrium prices, so it would be expected that if a theory properly characterizes the equilibrium, then it will properly predict *ex ante* EU. But this is apparently not the case, which suggests an anomaly. If market equilibriums were perfectly accurate, then the anomaly would be total. Because the predictions of market equilibrium are not perfect, the authors explored the possibility that perhaps subjects' preferences were consistent with the expected utility of the actual market equilibriums, if not with the analytically predicted market equilibrium. They found that they still were not. Ultimately, the authors adopt another approach, and propose that subjects have different attitudes toward different sources of risk, a phenomenon which traditional analytical models do not consider.

In Chapter 4, Amaldoss and Rapoport report the results of an experiment designed to investigate the effects of idiosyncartic investments in collaborative networks. The research is motivated by a desire to better understand the emerging phenomenon of networks, rather than individual firms, developing new products. In contrast to the common belief of alliance managers, the authors have shown that in theory the joint investment of network partners does not decrease as a network grows in size. Specifically, if the investments are recoverable, the joint investment should increase as the network size increases. But if they are not, then joint investment should not change with network size. On extending the theoretical model to investigate competition among a large number of networks ($N > 2$), the authors found that the effect of number of competing networks on joint investment depends on whether the investments are recoverable. If they are, it exerts a positive effect, but if they are not, it has a negative impact. In this chapter they describe an experimental test of these predictions in a laboratory setting. The experimental results support the qualitative predictions of the model. That is, they report that the joint investment increases as network size increases when investment is recoverable. But joint investment does not change significantly with increase in network size when investments are nonrecoverable. Amaldoss and Rapoport also detected a trend toward equilibrium behavior over multiple iterations of the stage game, and found that an adaptive learning model (EWA) accounts for the investment patterns of the subjects over time.

Chapter 5 by Hertwig and Ortmann discusses the methodological insights that experimental economists may derive from the debate in psychology about the reality of cognitive illusions. The authors have argued elsewhere that psychologists can learn from the experimental practices of economists. In this chapter, the proposed directional cross fertilization is reversed.

Hertwig and Ortmann discuss the heuristics-and-biases program launched by Kahneman and Tversky in the early 1970s. This program stresses that people have only limited "reasoning power" at their disposal and hence must rely on cognitive heuristics to make judgments and choices. Although these heuristics are highly economical and usually effective, they can lead to systematical and predictable errors that are variously referred to as biases, fallacies, or *cognitive illusions*. The heuristics-and-biases program has attracted the attention of numerous social scientists, including economists and legal scholars. In fact, much of today's work in behavioral economics and behavioral finance draws inspiration and concepts from the heuristics-and-biases program. This attention is warranted because systematic biases may have important implications for economic behavior.

As the heuristics-and-biases program has gained acceptance outside psychology, it has also drawn criticism within psychology. Some critics have suggested that the heuristics-and-biases research strategy has a built-in bias to find cognitive illusions, and others have claimed that some cognitive illusions are themselves illusory. Perhaps the most influential objections were voiced by Gigerenzer, who has argued that the heuristics to which cognitive illusions are attributed are not precise process models; that the heuristics-and-biases program relies on a narrow definition of rationality; and that cognitive illusions can be reduced or made to disappear by representing statistical information differently than it typically had been in the heuristics-and-biases experiments.

Hertwig and Ortmann's focus in this chapter is neither the controversy about cognitive illusions nor its implications for rationality. Instead, it is what they see as the important methodological insights that have emerged from the controversy, which can inform the choices that all behavioral experimenters wittingly or unwittingly make when they sample and represent stimuli for their experiments. In particular, Hertwig and Ortmann discuss the issues of stimulus sampling and the way these stimuli are presented to subjects, and then show that both factors matter in experiments with economical context.

For example, the question whether and how to sample from the environment has not been of much concern to experimental economists. Little attention has been paid to how representative these environments are of their real-world counterparts and the neglect of representative design has been amplified by the practice of using abstract tasks. However, there is now ample evidence that stripping away content and context prevents participants from applying the strategies that they use in their usual habitats.

Similarly, the authors argue that stimulus representation is an important factor in experimental economics and demonstrate how representing the stimuli in different formats (e.g., graphical) can dramatically reduce inconsistent behavior in an Allais type task even if boundary gambles are used.

Chapter 6 by Kramer and Budescu explores the role of vagueness (ambiguity) in choice. Ellsberg's paradox (1961) involves an inconsistent set of choices amongst two urns, each filled with red or blue marbles, but whose composition is known with different levels of precisions. In the "classic paradox" the DMs' choices indicate that the more certain urn is more likely to produce the desired marble *for each color*, implying that Pr(red) + Pr(blue) >1. Several empirical studies have investigated variations of this paradigm, but none have demonstrated conclusively the presence of Ellsberg's paradox in situations where the composition of neither urn is known precisely. In the present study the authors investigate this Vague-Vague (V-V) case, where neither of the urns' color probabilities are specified precisely, but one urn's probabilities are always more precise than the other. They show that people prefer precisely specified gambles and succumb to Ellsberg's paradox in these "dual vagueness" situations. The tendency to avoid the more vague urn and the prevalence of the classic paradox (and all the other two-choice patterns) is similar, but not identical, in the standard P-V (Precise-Vague) and the V-V situations. When conditioning on the midpoint (the middle of the probability range[s]), there is a reversal in vagueness avoidance between P-V and V-V cases. Otherwise, their results indicate that P-V and V-V cases are not qualitatively different, and it is more appropriate to think of them as defining a continuum of "degree of vagueness." The P-V case is just one, admittedly critical and intriguing, endpoint of this continuum. In both P-V and V-V cases, the prevalence of the paradoxical pattern of choices depends primarily on the ranges of the two gambles (i.e., the relative precision and minimal imprecision of the pair) and, to a lesser degree, on the pair's common midpoint.

In Chapter 7, Levy and Levy experimentally test the overweighing of recent return observations in an investment experiment with business school students and financial practitioners. They find that it is mainly the most recent observation that is overweighed, and that this overweighing is very strong. They estimate the decision weight attached to the most recent observation as approximately twice the objective probability. In this framework, probabilities are subjectively distorted on the basis of the temporal sequence of the observations, unlike the distortion that takes place in single-shot lottery type decisions (as in Prospect Theory, Cumulative Prospect Theory, or Rank Dependent Expected Utility models). This framework is applicable to circumstances where individuals are given observations as time series, as they are in financial markets, rather than a "given" set of outcomes and probabilities, as in many decision-making experimental setups. The case of the temporal probability distortion seems more relevant to actual economic decisions because in practice investors observe time series data regarding corporate earnings, mutual fund returns, etc., and their decisions are based on these time series. The findings of this paper suggest a simple explanation to several important economic phenomena like momentum (the positive short run autocorrelation of stock returns) and the relationship between recent fund performance and the flow of money to the fund. The results also have strong implications to asset allocation, pricing, and the risk-return relationship.

Chapter 8 by Blume, DeJong, and Maier concerns cognitive processes in common-interest spatial dispersion games in which the agents' common goal is to choose distinct locations. The games are characterized by multiple, non-strict equilibria. It is an open question whether players can select and attain equilibrium in such games and if equilibrium can be achieved, how long will it take and what are its characteristics. A further question is whether the insights from matching games extend to dispersion games. The authors report on an experiment designed to answer these questions. In their setup, cognition matters because agents may be differentially aware of the dispersion opportunities that are created by the history of the game. Their main finding is that strategic interaction magnifies the role of cognitive constraints. Specifically, with cognitive constraints, pairs of agents fail to solve a dispersion problem that poses little or no problem for individual agents playing against themselves. When they remove the cognitive constraints, pairs of agents solve the same problem just as well as individuals do. In addition, they report that when playing against themselves agents do not change the mode by which they solve the dispersion problem when their design removes the cognitive constraints.

In chapter 9, Chong, Camerer, and Ho further develop their cognitive hierarchy (CH) model. Strategic thinking, best-response, and mutual consistency (equilibrium) are three key modeling principles in non-cooperative game theory. In a previous paper, the authors relaxed mutual consistency to predict how players are likely to behave in one-shot games before they can learn to equilibrate. They introduced a one-parameter cognitive hierarchy (CH) model to predict behavior in one-shot games. The CH approach assumes that players use k steps of reasoning with frequency $f(k)$. In their previous paper they assumed $f(k)$ to be a one-parameter Poisson distribution. This chapter investigates and lends support to the generality and precision of this Poisson CH model in three ways: 1. An unconstrained general distribution CH model is found to offer only marginal improvement in fit over its Poisson cousin and hence this suggests that the Poisson approximation is reasonable. 2. The steps of thinking players use in games are found to positively correlate with response time and schools they attend which suggests that cognitive hierarchy captures realistically a reasoning mechanism that goes on in the brain of these players. 3. Several classes of interesting economic problems, including asset pricing and business entry, can be explained by the iterated reasoning of the Poisson CH model. When compared to the Quantal Response Equilibrium model, which relaxes the best-response assumption of equilibrium theory, the better fit of Poisson CH model seems to suggest that mutual consistency is a more plausible assumption to relax in explaining deviation from equilibrium theory.

Chapter 10 by Fox, Bardolet, and Lieb explores a wide range of judgment and decision tasks in which people are called upon to allocate a scarce resource (e.g., money, choices, belief) over a fixed set of possibilities (e.g., investment opportunities, consumption options, events). The authors observe that in these situations people tend to invoke maximum entropy heuristics in which they are biased toward even allocation. Moreover, they argue that before applying these heuristics, decision makers subjectively partition the set of options into groups over which they apply even allocation. As a result, allocations vary systematically with the particular partition that people happen to invoke, a phenomenon called *partition dependence.* The authors review evidence for maximum entropy heuristics and partition dependence in the following domains: (1) decision analysis in which the degree of belief and importance weights must be distributed among possible events and attributes, respectively; (2) managerial decision making in which money and other organizational resources are allocated among risky projects, divisions, and organizational stakeholders; and (3) consumer choice in which individuals select among various consumption goods and consumption time periods.

In Chapter 11, Gneezy investigates the influence of prior gains and losses on the risk attitude of people. Empirical findings suggest that in decisions under uncertainty people evaluate outcomes relative to a reference level: they are risk-seeking in the domain of losses and risk-averse in the domain of gains. The finance literature uses this finding to predict/explain the "disposition effect," which is the tendency of investors to sell assets that have gained value ("winners") too early and ride assets that have lost value ("losers") too long. The purpose of the experiment reported in this chapter is to investigate the influence of prior gains and losses on the risk attitude of people. Unlike the case of real market data, the stylized experimental setup allows the author to gain insight into the decision-making process of individuals. Furthermore, using a stylized decision problem makes the benchmark prediction very clear and testable. One of the main goals was to find evidence on how prior gains and losses influence the risk behavior of people, by shifting the reference level. The results show that prior gains and losses do influence the risk attitude, and in a different way from that predicted by the rational theory (expected utility). The disposition effect prediction that people will be reluctant to sell losing assets found strong empirical support with the traditional assumption that the reference level is the initial purchase price of the stock. This finding supports the empirical research done on real market data. The use of a stylized process also allows for more refined tests about the way reference levels are formed. In particular, it is possible to learn about how it depends on the history of gains and losses. This is important because, for example, prospect theory is useless as a descriptive theory without a "good" assumption about the reference levels. It was found that when the peak of the process was used as a reference level, the descriptive power of the theory increased dramatically.

Chapter 12 by Dufwenberg and Gneezy investigates the relationship between gender and coordination. Groups of six females or six males played the minimal effort coordination game for ten periods. Little difference was found between the groups of men and women with regard to their ability to avoid the least efficient equilibrium. The results show some differences in the initial stages of the game, but these differences quickly disappear and no difference is found in later stages. In addition to reporting this result, the authors raise a methodological issue: Is there a bias in the research community against reporting or publishing results that document the absence of a gender effect? The results reported in this chapter are not "positive," in the sense that no difference in behavior between females and males was found. The authors believe that in order to truly understand the differences in behavior between genders, one should not only report or publish experiments and results that show positive differences because such practice would bias perceptions about the magnitude and the limits of the differences.

The last two chapters discuss the use of laboratory- and class-based experiments intended to enhance teaching and learning. Successful attempts to teach business related courses through experiments and projects conducted in computerized laboratories (e.g., the Economic Science Laboratory at the University of Arizona, the Laboratory for Economic and Political Research at the California Institute of Technology) all testify to the benefit of integrating this new methodology in the teaching of business related courses at the undergraduate, graduate, and MBA levels. There is by now ample evidence that "hands-on" learning through experimentation, in which different economic scenarios are created under controlled laboratory conditions, is a very effective way of acquiring new concepts and procedures, gaining insight into business practices, and learning how to make better decisions.

The basic idea that underlines this new teaching methodology is that actual experience in carefully designed experiments, whether they are selected to test basic theoretical concepts or mirror business problems that appear in practice, is critical for effective teaching of business. The experience takes two forms: a personal experience of participation and a supply of data produced by the participants. The personal experience is invaluable for maintaining the student's attention and motivating his/her understanding of the material, but it is the data produced by the group that truly make clear the power of economic principles in understanding markets, bargaining, and other business decision environments.

Chapter 13 by Croson, Donohue, Katok, and Sterman describes an experiment that illustrates the challenges of supply chain management. Supply chain management involves the management of orders and shipments of goods through a supply chain; for example, shipping beer from the manufacturer to the distributor to the wholesaler and then to the retailer for sale to customers, and transmitting the orders for beer back up the supply line. A large body of research investigates these issues theoretically. However, in addition to the theoretical operational challenges, there are also cognitive limitations that managers face which prevent them from optimally managing their supply chains. Chapter 13 describes an in-class experimental game that can be used to illustrate a number of these challenges, operational and cognitive, that managers face in supply chain management. The experiment is well-suited for undergraduate, MBA, or executive teaching, and has been used in all those forums. Exactly which treatments to choose, and how deep the debriefing should be, will depend on the sophistication of the audience as well as the manner in which the teacher chooses to implement the experiment (physical or computer).

The last chapterby Erev and Livne-Tarandach describes an innovative approach to the use of experimentally derived findings in experiment-based exams in the social sciences. The authors have analyzed GRE exams and highlighted an important difference between the natural and the behavioral sciences. Most questions in Physics ask the examinee to predict the results of particular experiments. On the other hand, nearly all questions in Psychology deal with abstract terms. The analysis in Chapter 14 clarifies this difference, and proposes two related steps that can lessen the gap.

The first step addresses the difficulty of developing experiment-based questions in the behavioral sciences. The authors assert that the main stumbling block, from the developer's point of view, lies in identifying questions with unambiguous correct answers. The solution proposed here is technical. It requires focusing each question on a particular experiment that has been run. With this focus in mind, the correct answer is crystal clear: It is the observed experimental result. Their analysis suggests that the discriminative power of experiment-based questions based on this technical solution is at par with the discriminative power of more typical abstract questions. The second step requires some changes in the information collected by researchers and presented to students. The authors assert that the discriminative power of experiment-based questions can be improved through the standardization of descriptive models and experimental procedures. The standardization of descriptive models as suggested, for example, by Erev, Roth, Slonim, and Barron is expected to have three benefits: It would allow unbiased selection of experimental tasks; it would clarify the boundaries of descriptive models; and it would provide guidance where models conflict with intuition, introspection, and or personal experience. The standardization of experimental procedures is expected to be beneficial in that it would facilitate clear and parsimonious presentations of experiment-based questions.

Erev and Livne-Tarandach believe that the use of experiment-based questions to evaluate students in behavioral science courses is likely to have many attractive outcomes. In addition to making behavioral science exams more similar to those in the natural sciences, this effort will advance the behavioral sciences in substantial ways. A focus on predictions in exams is likely to have a similar effect on courses, textbooks, and mainstream research.

ACKNOWLEDGEMENTS

We owe thanks to many for the successful completion of this volume. Most importantly, we express our gratitude to the contributors who attended the conference and participated in insightful discussions. The conference was supported financially by a grant from the Hong Kong University Grant Commission to cEBR (Project No. HKUST-3, Experiential based teaching for networked economy), and by an RGC Direct Allocation Grant (Project No. DAG02/03.BM78) to Rami Zwick and Soo Hong Chew. Additional financial support was given by HKUST. Special thanks are due to Professor K C Chan the Dean of the HKUST Business. We wish to thank Maya Rosenblatt and Maggie Chan, the conference secretaries, without their help the conference would have been a total chaos, and Chi Hang Chark for the splendid and dedicated work in preparing and formatting all the chapters for publication. We also thank Kluwer for supporting this project.

Chapter 1

THE RATIONALITY OF CONSUMER DECISIONS TO ADOPT AND UTILIZE PRODUCT-ATTRIBUTE ENHANCEMENTS: WHY ARE WE LURED BY PRODUCT FEATURES WE NEVER USE?

Shenghui Zhao
University of Pennsylvania

Robert J. Meyer
University of Pennsylvania

Jin Han
Singapore Management University

Abstract

The ability of consumers to optimally anticipate the value they will draw from new product features that are introduced to enhance the performance of existing technologies is explored. The work tests a hypothesis that when consumers are given the opportunity to buy a new generation of a products that offers enhanced features consumer will overvalue them, a bias the accrues to a tendency to overestimate both the extent that they will utilize these new features and the impact they will have on utility. This general hypothesis is tested in the context of a computer simulation in which subjects are trained to play one three different forms of an arcade game where icons are moved over a screen by different forms of tactile controls. Respondents are then given the option to play a series of games for money with either with their incumbent game platform or pay to play with an alternative version that offered an expanded set of controls. As hypothesized, subjects displayed an upwardly-biased valuation for the new sets of controls; adopters underutilized them and displayed a level of game performance that was not better than those who never upgraded. A follow-up study designed to resolve the process underlying the bias indicated that while adopters indeed over-forecast the degree to which they would make use of the new control, they did not over-forecast performance gains. Hence, the key driver of adoption decisions appeared to be an exaggerated belief of the hedonic pleasure that would be derived from owning and utilizing the new control as opposed to any objective value it might provide.

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 1–33.
© 2005 *Springer. Printed in the Netherlands.*

As consumers we have always had something of a love-hate relationship with new generations of products. On one hand, innovations that hold the promise of being the latest and best in a class of technologies often hold an allure that seems to go beyond the objective incremental benefits they provide. Manufacturers of new gaming systems seem never to produce enough units to meet initial demand, we brag about the multitude of features that endow our new cell phones (even if we never use them), and the wealthy compete to see who can fill their homes with the most advanced technological gadgets. Even those who lack the wealth to acquire technological enhancements are no less subject to their appeal; society surrounds us with images of innovations in magazines, television ads, and billboards.

Yet, it is equally transparent that whatever appeal consumers may see in acquiring new technologies, it is an appeal that has real limits. As attracted as we may be to the idea of acquiring that which is new and innovative, we are also often averse to incurring the switching costs that are often associated with adopting innovations – an effect cognitive scientists term *lock-in* (e.g., Norman 1998; Johnson, Bell, and Lhose 2003; Zauberman 2003). Hence the origin of Klemperer's (1987) paradox of the early innovator: individuals who are the first to adopt new technologies often turn into laggards, inhibited from keeping up with the pace of innovation by the need to constantly incur switching costs.

How do consumers balance these instincts when forming assessments of their willingness to adopt product innovations? The answer to this question is uncertain. On one hand, there is ample anecdotal evidence that would seem to support the often-heard claim that consumers over-estimate the degree that they will make use of enhanced features carried by new technologies. For example a 2003 *Harris Poll* revealed that 45% of cell phones owners never use voice mail features, and 50% have never exercised the option of setting their phones to silent or vibrate[1]. But the mere fact that consumers make limited used of the advanced features of new products, of course, does not necessarily imply that a forecasting error had been made at the time of purchase, or that they would be happier if they put them to greater use. An un-used feature may have been acquired simply because it was part of a sales bundle, or the feature may have been purchased for its option value. That is, only by acquiring the feature could the consumer learn whether they would be useful or not, or gain access to it at an uncertain later point in time. Finally, it should be noted that, by definition, the reciprocal error of *under*-forecasting is difficult to document; while it is easy to observe attributes that are purchased but never used, we never observe attributes that would have been used had they been purchased.

The purpose of this paper is to take a step toward resolving this research uncertainty by systematically investigating the quality of consumer decisions to adopt and then subsequently utilize innovative features in new products. We undertake our investigation in a controlled laboratory setting where subjects are trained to play a new arcade game for a monetary incentive where game tokens are moved using a certain set of computer controls. Subjects are then given the opportunity to purchase alternative versions of the platform that offer expanded sets of controls. The objective of the paradigm is to identify biases in consumers' willingness-to-pay for product attribute enhancements as well as how these attributes are subsequently utilized. In addition, we also examine biases that arise in the reverse case where the product innovation offers a design *simplification*.

The core finding of the work is strong support for what might be termed an *enhancement bias* in new-product adoption decisions. When given the opportunity to purchase an enhanced game platform subjects reveal levels of willingness-to-pay that are greatly in excess of that which can be explained based on either their own best forecasts of score improvement or a simplified options-value analysis of the adoption decision. In essence, subjects act as if mere access to the new set of controls – regardless of their functional value – provides a source of prospective utility worth paying for. Yet, once this ability is in place few seem to utilize it; players who acquire the enhanced platform withdraw use of the new controls after overly-short periods of experimentation, and do not realize higher levels of performance compared to those who never had the chance to upgrade.

We organize our presentation of our research in three phases. We first develop a more complete background for the research by reviewing the normative basis for consumer new product-adoption decisions and exploring prior behavioral research that suggests how actual decisions may depart from this benchmark. We then test these hypotheses using data drawn from two laboratory experiments. We conclude with a general discussion of the implications of the work for both basic research in consumer response to product technologies as well applied work in new-product design.

1. THE PSYCHOLOGY OF NEW-PRODUCT ADOPTION DECISIONS

In this work we consider how consumers solve a class of new-product adoption problems that have the following structure. A consumer currently owns a durable good that conveys utility through the utilization of a set of features (such as options in software or capabilities of a home entertainment device). A manufacturer offers the consumer the opportunity to purchase an enhanced version of the good that retains the features of the old but also offers a new set of discrete attributes of uncertain value. The existence of these new attributes does not affect the functionality or utility derived of the older attributes, however they do compete for usage time. That is, the new attributes cannot be used simultaneously with the old. Hence, analogies might be software packages that provide users with the option to utilize either older or newer interfaces (similar to *Windows XP*), or digital cameras that give users the option to operate it with basic or advanced settings.

We can formally model the consumer's problem as follows. Assume that the utility that the consumer realizes from consuming an incumbent good with attribute α at any point in time t is scaled to be 0. Let $d_t \in \{0,1\}$ denote the consumer's decision whether or not to utilize some new feature δ given its ownership at time t, let $x_t = u(\delta) - c(\delta)_t$ be the net utility that is realized given a decision to utilize δ at t, and z_t denote the consumer's beliefs about the probability distribution associated with x_t^2. In addition, let $\pi_T = d_0, \ldots d_T$ be a sequence of attribute-usage decisions defined over a T-period ownership horizon, and let $V_0(\pi_t)$ be the total discounted expected utility implied by this sequence, defined as follows:

$$V_0(\pi_T) = E_0 \sum_{t=0}^{T} \beta^t v(x_t, z_t, d_t) \quad (1)$$

The decision maker's goal would then be to find that sequential decision policy π_T^* that maximizes expression (1), yielding an optimal ownership valuation ($V^* = V_0(\pi_T | \pi_T^*)$). The consumer would then be prescribed to buy the new product if $V^* > C$; that is, if the lifetime expected value of the new product that follows from assuming optimal utilization of the innovative feature δ exceeds good's purchase price.

It goes without saying that making a new-product adoption decision in this manner would be a formidable feat of cognition. One would need to possess good skills not only in intuitive dynamic programming (to derive the optimal ownership policy π_T^*), but also hedonic forecasting – accurately anticipating the various possible states of long-term pleasure one might come to associate with a new technology (the distribution over net asymptotic values of $u(\delta) - c(\delta)$) as well as how this pleasure may change over time in the course of ownership.

How potentially damaging would failure of these assumptions prove? On one hand, the literature is replete with examples of intuitive decisions that closely correspond with those prescribed by highly complex normative models (e.g., Hogarth 1981; Meyer and Hutchinson, 2001; Rust 1992). Yet, there is growing evidence that this same robustness may *not* extend to tasks – like the current – where decision makers are required to forecast their future preferences. Specifically, as skilled intuitive decision makers we may be in many domains, predicting how we will feel and act in the future does not appear to be one of them (see, e.g., Loewenstein and Schkade 1999; Wilson and Gilbert 2003). A core hypothesis of this research is that when making product-adoption decisions biases in hedonic forecasts will yield systematic inefficiencies in both the quality of initial decisions to buy new goods and their subsequent utilization after purchase.

We will briefly review lines of evidence that suggest systematic biases that may arise when consumers attempt to develop two kinds of forecasts that would be central to the normative solution to expression (1): forecasts of the mean potential value of an innovative attribute (beliefs about x and z); and forecasts of the dynamic utilization of the new attribute (beliefs about the decision policy π).

1.1. Intuitive forecasts of new attribute values

Assessing what one should be willing to pay for new product features is not an easy task. Such assessments should rationally reflect not just the pleasure one anticipates drawing from the feature over the expected future of ownership, but also the costs that will be incurred learning to use the feature, and, most critically, the long-term utility of *not* acquiring it; keeping the current device and spending the money on something else. How skilled will consumers be in making these kinds of assessments? While no work has examined this question directly, research that has examined the quality of human hedonic forecasts would not seem encouraging (e.g., Kahneman 1999; Loewenstein and Schkade 1999; Wilson and Gilbert 2003). Prior evidence suggests that not only will consumer assessments of the likely future value of attribute innovations often depart from normative benchmarks, but that these departures will have a distinct bias: toward overvaluation.

The core argument is as follows. One could view the above normative framework as requiring consumers to hold three kinds of expectations when valuing product enhancements: an initial short-term expectation of the relative value offered by the innovation, an expectation of how these beliefs will evolve over time through ownership, and a belief about the option value of the attributes – the utility of being able to decide in the future *not* to use the feature if its value turns out to be limited. We argue that consumers will commonly systematically overvalue new-product features because of the cascading effect of congruent distortions in each of these judgments: a tendency to hold overly optimistic priors about value, under-assess the likelihood that pleasure may diminish in the future, and over-assess future option values.

Consider, first, the direction of affect consumers will first associate with a product innovation. There are strong normative and psychological arguments that predict that these assessments will routinely be positively biased, with consumers feeling a lure to acquire the new good that is not based in any objective knowledge of value. Common experience, of course, offers numerous anecdotes that would seem to support this idea: we are attracted to new rides at amusement parks and new flavors of ice cream, and are anxious to read about the latest innovations in computer technology. In many cases these kinds of reactions have a sound rational basis in information economics: one *should* be tempted to try new options that appear in markets because it is only through the experience of trial will we know which options will give us the highest utility in the future.

There is also evidence, however, that the lure consumers feel toward product innovations is triggered by more than curiosity: novel products also often are evoke heuristic expectations of heightened quality. To illustrate, Miller and Kahn (2003) offer data showing that merely affixing novel names to the color or flavor of an otherwise familiar product enhances its perceived quality among consumers. They suggest that the effect arises not as a result of a rational desire for information but rather by a simpler effect of conversational norms (Grice 1975). Given a communication that is seen as potentially ambiguous (in their case, a color or flavor name), consumers implicitly assume that it holds relevance to the purpose of the communication (conveying something about the nature of the product), and its valence is inferred from the presumed intended consequence of the communication (that the consumer would be more inclined to buy the good). In the case of innovative product attributes conversational norms would predict a similar result; even if consumers were not lured by curiosity, most would believe that if a firm took the time and effort to add new features to a good it was with the intention of enhancing its value.

This same effect is likely to be compounded by yet another documented bias in decision making: the tendency of individuals to overvalue options that allow for flexibility (e.g., Lowenstein and Alder 1996; Simonson 1990). Translated to product design, such a preference would reinforce a "more is better" heuristic in evaluating new product attributes: even if one suspects that that an expanded set of a feature offered by a product innovation carry little immediate value (e.g., an imbedded camera in a cell phone), one might nevertheless desire having it as a hedge against future changes in preference or usage norms.

Of course, such assessments *per se* are far from wrong; recall that in a normative analysis the prospective value of an innovative product feature depends not just on the utility that one expects to receive from it given its use ($u(\delta)$), but also the option value of *not* using it. The problem comes from the fact that individuals routinely overvalue the merits of such flexibility.

For example, Simonson (1990) and Loewenstein and Alder (1996) report data showing that when consumers are asked to make a one-time choice of a basket of product flavors that will be consumed in the future they tend to choose a wider assortment than is actually consumed when these choices are made individually over time. Likewise, Shin and Ariely (2003) report a sequential search task where people are willing to pay to keep search routes open even when the odds that they will be utilized is small. Finally, Gilbert and Ebert (2002) and Wilson and Gilbert (2003) offer evidence that consumers tend to strongly prefer transactions that allow for revocability (e.g., liberal exchange policies), even when they are unlikely to be exercised. Hence, while there is indeed a rational basis for desiring products that offer a flexible assortment of features, the value that consumers place on this capability may be excessive.

Of course, one might argue that these kinds of visceral assessments of product value might fade once consumers begin thoughtful analyses of the real *net* utility they would draw from an innovation given its purchase price. Consumers might (and should) come to recognize, for example, that with these new features comes the cost of having to learn how to use them, and recall times in the past when they were lured to buy new goods in the belief that they would dramatically enhance pleasure, only to find that the enhancement was modest at most. Yet, the weight of evidence is that consumers will under-attend to these considerations, perpetuating a positive assessment bias.

Supporting this idea is empirical evidence that affective forecasts are often subject to what that Loewenstein, O'Donoghue, Matthew Rabin term a *projection biases*, a tendency to presume that one will feel in the future much as how one feels today. What seems to drive this bias is an effect that Wilson and Gilbert (2003) call *focalism*: when a decision maker is in one affective state it is difficult to imagine being in another, or project the preferences one will have at future points in time (see also Kahneman and Snell 1992). Gilbert, Gill, and Wilson (2002) and Read and van Leeuwen (1998) illustrate thus effect by showing there is real truth to the old adage that one should never shop on an empty stomach; shoppers who are hungry systematically buy more than those who are full, presumably due to inability to anticipate how they will feel in the future when they begin to consume the goods they are purchasing. Likewise, DellaVigna and Malmendier (2002) and Gourville and Soman (1998) offer evidence from health-club attendance patterns that people systematically underweight future costs in the form of effort. Specifically, subscribers pay large up-front fees to join a gym (implying high expectations of usage), but then underutilize it after joining, implying an under-forecast of the effort required to attend. The implication here is that while learning costs may ultimately play a major role in influencing how new-product attributes are actually used, they will tend to be undervalued at the time product-adoption decisions are made.

Taken together, these streams of work suggest a straightforward hypothesis about how consumers will prospectively value new attributes carried by product innovations:

H1: The Innovation Bias. *When given the opportunity to purchase a new product that possesses an expanded set of attributes relative to an incumbent, consumers will display an overvaluation bias, revealing rates of adoption and levels of willingness-to-pay in excess of those would be justified by both actual subsequent utilization patterns and a rational a priori options valuation.*

The logic that underlies **H1** rests, on an assumption that the most salient initial reaction that consumers will have when exposed to a product innovation will always be that of optimism about its value. The degree to which this would hold in natural settings, of course, would be expected to vary from consumer to consumer. For example, a consumer who has recently incurred extremely high learning costs when adopting an innovation might have far more tempered – or even negative – visceral reactions to a product that offers yet another new set of features. In the same way that focalism predicts that optimistic consumers will be prone to underweighting future learning costs when valuing products, pessimistic consumers may be prone to underweighting future pleasure. Given this, we might expect that individual differences in difficulties encountered when learning to use new product features in the recent past could serve to moderate the general prediction in H1. Formally,

H1a: The moderating effect of past learning costs: *the mean tendency of consumers to overvalue product innovations will be moderated by past learning costs, with the bias being tempered among decision makers who have experienced associate steep learning curves with innovations.*

Now that they've bought it, will they use it?

Central to the work on affective forecasting that forms much of the basis of **H1** and **H1a** is the idea that biased forecasts arise because individuals are poor at anticipating how they will make decisions in a future world where the on-going judgment tasks and inputs are substantially different from those that are faced today. In the case of new-product judgments this disconnect would seem particularly acute. At the time of purchase the consumer's cognitive efforts are focused on solving a rather formidable normatively decision problem: that of whether the option value of acquiring a new generation of a technology is worth the purchase price, given assessments of the likely horizon of ownership, likely utilization over that horizon, and the affect associated with loss of the incumbent good and liquidity. But once an affirmative decision to acquire the innovation is made, cognitive efforts shift to solving a quite different – and seemingly much simpler–task: making moment-to-moment decisions about whether to make use of the innovative attributes of the good now that it is owned. These judgments, in turn, will be influenced by a range of hedonic factors that were not salient at the time of the initial choice, such as the frustration of learning how to use a new product attribute, and the appeal of momentarily deferring this learning to a future time period during ownership.

The implication is that consideration of these new factors will not only lead to levels of attribute utilization that are below those envisioned at the time of purchase, but also below those that would maximize the absolute long-term utility of ownership. Specifically, when a consumer who has purchased an new product is deciding whether or not to try utilizing one of its new features the decision is not simply one of whether this action might yield long-term benefits, but whether these benefits – which are uncertain – will be higher than those afforded by continuing to use older, more familiar, attributes. A systematic finding of work on technological utilization is that when consumers have well-developed skills in utilizing one technology they often find it difficult to learn new ones, and are frequently averse to learning – an effect called termed cognitive lock-in (e.g., Johnson, Bell, and Lhose 2003; Norman 1998; Zauberman 2003). The usual explanation is that expertise with using one generation of a technology tends to increase as logarithmic function of practice (the power law), implying that the more familiar one becomes with one technology, the higher the short-term relative cost of learning to utilize new technologies (Klemperer 1987; Zauberman 2003).

The fact that new technologies involve switching costs, however, does not by itself imply that consumers will be prone to error in how they initially value technologies or how they utilize them once acquired. As we noted earlier, normative assessments of the value of innovations *should* anticipate such costs (through the consumer's beliefs about how $u(\delta)$ and $c(\delta)$ will evolve over time), and after purchase the observed magnitude of switching remain a normatively-relevant consideration in usage decisions. For limited utilization to be judged an *error*, therefore, the effect of switching costs on usage must be greater than what would be anticipated in a rational analysis.

Prior work on dynamic decision making in other contexts provides strong hints that processing of switching costs may well be biased in just such a manner. First, one of the most pervasive findings in the study of decision making over time is that people frequently undervalue the long-term benefits of learning and experimentation (see, e.g., Meyer and Hutchinson 1994; 2002). For example, in experimental armed-bandit tasks decision makers frequently cease gathering data on unfamiliar options after overly-short periods of experimentation (e.g., Meyer and Shi 1985), and melioration experiments find a similar aversion to making short-term costly investments when the benefits are long-run and distant (e.g., Herenstein and Prelec 1992). Hence, while consumers might well concede the long-term benefits of learning about new technologies at some abstract level, day-to-day decisions about the attribute utilization may be dominated by short term assessments of which product features yield the greatest benefit at the lowest cost – leading to underutilization.

A related influence that may further contribute to underutilization is the fact that in product-adoption settings learning is deferrable. In other words, for most consumers the decision about whether to take up learning about a new product feature is not one of whether it will *ever* be beneficial to learn (for most, the answer would be, "probably yes"), but rather whether *now* is the best time to start. From a normative perspective, of course the answer to this question will always be "yes"; one should always want to resolve uncertainty as early as possible so as to allow the benefits of information can be realized over the longest-possible time horizon. Yet, this is an instinct that is often lost on real decision makers (e.g., Meyer and Hutchinson 2002).

Specifically, there is extensive evidence showing that when individuals are presented with a choice between a set of uncertain alternatives versus deferral, growing indecision leads to a growing preference for postponement (e.g., Dhar 1997; Tversky and Shafir (1992). Hence, one might speculate that the more consumers are unsure whether the benefits offered by a new attribute are worth the learning costs, the greater will be their urge to delay the onset of experimentation. What is particularly attractive about delay in this context is that it allows consumers to mentally justify the short-term action of utilizing familiar attributes while still retaining the abstract goal of wanting to learn new technologies. By deferring one is not abandoning this long-term commitment, just delaying its onset to an unspecified future time when costs be lower (e.g., "I'll read the manual over the weekend").

A final factor that would contribute to under-utilization errors is if learning and usage costs at the time of initial purchase turn out to be much larger than was anticipated. In some cases this under-forecast will arise due to the inherent difficulty that comes from envisioning future affective states that we discussed above (e.g., DellaVigna and Malmendier 2002 and Gourville and Soman 1998). But an even more acute basis for under-forecasts would be if consumers use an inappropriate analogic-reasoning process to generate expectations about learning costs. That is, assume that knowledge about product usage gained in one domain can be directly transferred to the new one to greater degree than is the case (e.g., Moreau, Lehmann, and Markman 2001). While learning-by-analogy can often greatly reduce learning costs, it can also substantially raise them if the assumed analogies prove inappropriate; for example, assuming that short-cuts useful in one text editor holds for others (e.g., Norman 1988). In such cases consumers have the added burden of not just learning how to use the new technology, but also *un*learning interfering mappings to old ones – mappings for which they may be unaware (e.g., Wood and Lynch 2002). A relevant suggestive illustration of this effect has recently offered by Zauberman (2003), who reports data showing that people tend to over-forecast how productive they will likely be using new web interfaces, implying that the difficulties involved in switching to new formats went largely unanticipated.

Taken together, these discussions lead to the following general hypothesis about post-purchase utilization of new-technology attributes:

H2: The Under-Utilization Bias. *Given a decision to acquire an innovation that possesses a mixture of innovative and familiar attributes, utilization of the new attributes will be- downwardly biased relative to the levels implied by stated wiliness-to-pay for the good, direct forecasts of benefits, and objective benefits that would come from optimal usage.*

An intriguing consequence of the discussion we offered about how projection biases might influence both prior new-product valuations and subsequent utilizations is that it implies a possible paradox in how individual differences in post-purchase attribute utilization might relate to pre-purchase willingness to pay. In **H1a** we proposed that consumers who had more positive experiences when consuming past technologies would produce the most optimistic assessments of the prospective value of a new product that offered an attribute innovation. Yet, because much of this optimism comes from the under-forecasting of learning and switching costs (as above), it is these same consumers who would most likely experience the greatest disappointment when they come to utilize attribute innovation that they paid for. This disappointment, in turn, would lead to more rapid decisions to abandon use of the new attributes relative to those who entered ownership with more modest expectations. We summarize this idea in following hypothesis:

H2b The Paradox of the Technological Optimist: *Consumers who reveal the greatest optimism in their willingness to pay for a technological innovation will also be the most prone to abandon trial usage of attribute innovations given ownership.*

2. EMPIRICAL ANALYSIS

2.1. Overview and Design Consideration

In this section we describe the results of three experiments designed to test the empirical validity of the research hypotheses summarized in **H1**, **H1a**, **H2**, and **H2a**, as well as provide descriptive insights into the process by which consumers make decisions to buy and then subsequently utilize product innovations. These issues were examined by observing how a sample of experimental subjects learned to play an original arcade-like computer game where performance was rewarded by a monetary incentive. After a period of training with one of several basic platform designs subjects were given the opportunity to purchase an enhanced platform that offered a combined set of features that were drawn from the basic platforms. In a third experiment we examine the reciprocal case: subjects trained on the enhanced platform are given the opportunity to exchange it with a reward for a simplified platform containing only them most-used controls.

The game was called "Catch'em" and bore similarities to the popular late 70's, early 80's arcade game *Pac Man*. In the game players viewed a square grid on which, at the start, was superimposed a number of stationary green dots called "cookies". Also on the grid were two larger red and black dots that depicted the staring position of the player and his or her robotic opponent, termed the "Monster". Upon triggering the start of the game both the Monster's and player's icons began moving over the grid. While the Monster moved at a random speed and direction, the player controlled the speed and direction of his or her icon. Each time either the player's icon (or the Monster) moved over a cookie a point was scored for the player (or the Monster). If all of the cookies were consumed from the board by the player and/or the Monster, the play ended and the player received a point total equal to the number of cookies he or she had captured. If, however, at any point the Monster's icon touched the player's icon, the player's icon was declared "caught" and play also ended, with all points having been earned to that point being forfeited. The basic board layout and instruction are reproduced in Appendix 1.

We chose this – admittedly unusual – stimulus context because it was one that satisfied four ideal design criteria:

1. It provided us with experimental control over the design and familiarity subjects had with a basic generation of a technology;
2. It allowed experimental introduction over the value of enhanced features in a new technology;
3. It provided a natural objective for measuring performance that could be used for providing a monetary incentive to subjects; and
4. The task context – an arcade game – was one that was likely to be seen as highly involving and familiar to the subject pool, primarily undergraduate college students.

The technology in this case was the nature, complexity, and quality of the controls available to subjects for moving their icon. A basic technology was one where subjects had access to only one kind of control at one calibrated level of performance, while the enhanced technology was one where subjects had access to multiple controls – both those with which they were familiar and a "new" set that was derived from one of the other basic models (the existence of which was unknown to subjects).

Our analysis focuses on the results of two experiments conducted within this paradigm. The purpose of Experiment 1 was to conduct a basic test of the four hypotheses in a setting where there was minimal measurement intervention; we observed learning paths, the dynamics of control utilization, and adoption decisions in the absence of direct elicitations of either forecasts of behavior or elicitations of reasons for decisions – interventions that might influence behavior. In Experiment 2 we attempt to more deeply probe the process that underlies the data uncovered in Experiment 1 by gathering such process measures.

3. EXPERIMENT 1

3.1. Design, Subjects, and Procedure

Subjects were 149 business-school undergraduates who volunteered to complete the task for a monetary incentive. Subjects performed the experiment seated in computer cubicles in the school's behavioral research lab. At the outset of the experiment subjects were told that the purpose of the experiment was to learn how consumers such as themselves learned to play gaming devices, and that they would be paid depending on their performance in the game. Subjects were told that there would be a show-up fee of $5 (US) per subject, and they could earn up to $10 more depending on how well they learned to play the game.

All subjects were told that they would be playing the "Catch'em" game a total of 30 times, with the first 15 being practice rounds that would not count toward their final earnings, and the second 15 being money rounds on which their pay would be based. After reading this basic instruction subjects were randomly assigned to either a control or treatment condition, with which they were also assigned to play one of three different basic game platforms (described below). Subjects in the control condition played with the same platform over all 30 rounds of the experiment. Subjects in treatment condition played the first 15 training rounds with one platform, but were then given the opportunity to pay to play the money rounds with a new platform that offered a broader range of controls. The opportunity to pay to switch to a new platform was offered only once; if a subject declined the purchase he or she played the 15 money rounds with the same game platform that they trained on, the same as those in the control condition.

The game platforms. The three basic game platforms on which subjects trained on were defined by the physical form and reliability of the controls used to move the player's icon. There were three mechanisms:

1. *A Scroll Bar Control (Figure 1a)*: Subjects continuously adjusted the speed and direction of movement of their icon by moving each of two horizontal scroll bars displayed on the computer screen. Use of the directional control was aided by a steering-wheel-like graphic that displayed the current directional heading of the icon.

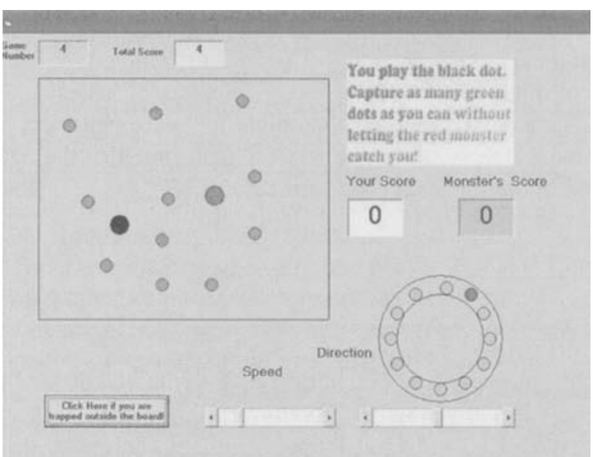


Figure 1. The Three Game Platforms.
1a: Scroll-Bar Control

2. *A Button Control with high reliability (Figure 1b).* Subjects adjusted speed and direction by repeatedly clicking two sets of button controls. One pair of buttons allowed subjects to reverse the current heading of their icon either horizontally or vertically, while the other pair induced discrete increases or decreases in speed. High reliability meant that the icon's movement responded 80% of the times to player actions in the intended manner given activation of any control.

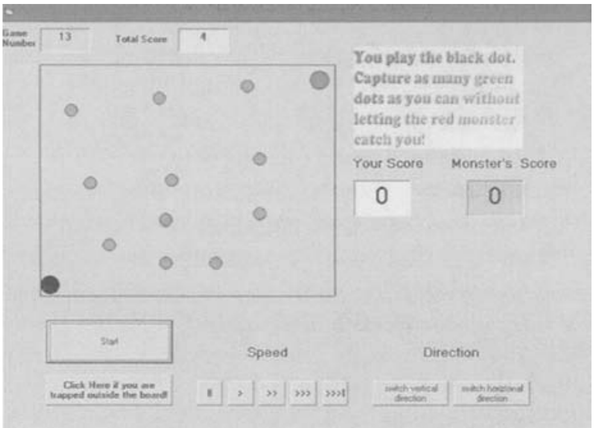


Figure 1b. Button Control.

3. *A Button Control with low reliability (Figure 1b).* The appearance and function of this platform was identical to (2), except that random noise was added to the

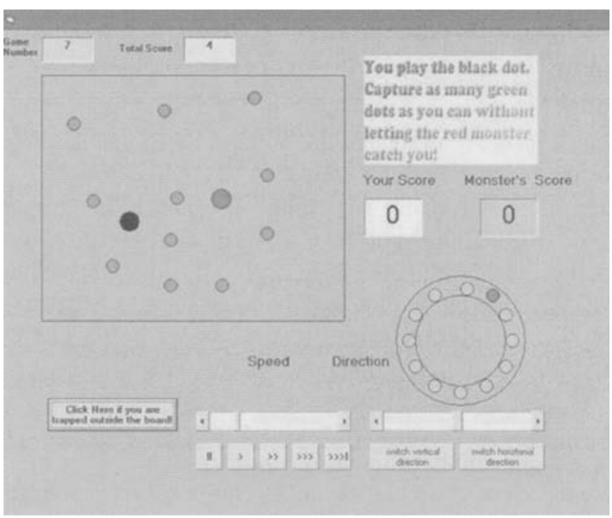


Figure 1c. The Enhanced Platform: Combined Controls.

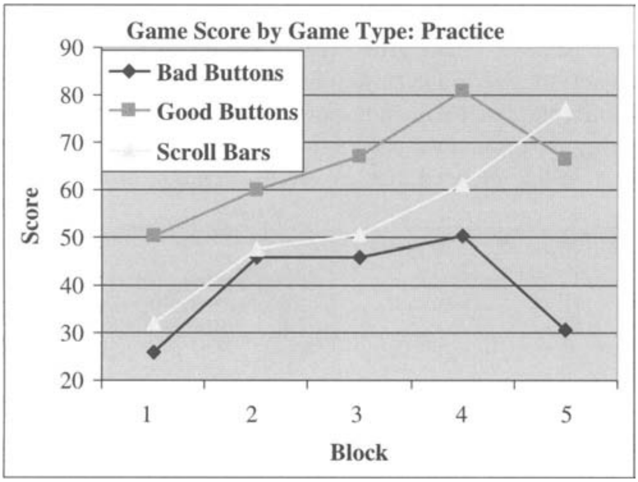


Figure 2. Performance over time during the training rounds by game platform type, Experiment 1. Time is reported in blocks of three trials.

responsiveness of controls. Specifically, given activation of a given control there was a 60% chance that it would momentarily fail, resulting in no change in movement of the icon.

In Figure 2 we plot the average performance attained by subjects using each of these control formats during the training rounds. The figure yields an important feature of this training manipulation: in addition to varying the tactile experience with controls that subjects had entering the money rounds, the three control conditions also manipulated the qualitative nature of their learning experience. Specifically, subjects found the button controls to be a more natural way of moving the icon than the scroll bar, and when the buttons were reliable they realized high levels of performance after a short period of familiarization. For subjects given the scroll-bar control, however, their learning experience was quite different: while they ultimately developed the same level of skill as those displayed by subjects who trained on the reliable buttons (as measured by average realized scores) this achievement was achieved only after they incurred more substantial learning costs as evidenced by the low average scores realized at the outset of training. Finally, subjects who trained on the low-reliability buttons would have found the training rounds to be a far more frustrating experience; while there was tactile ease in using the buttons, they would have experienced little improvement in achievement over time movement was inherently difficult to control.

The enhanced game. The central interest in the experiment was how subjects in the treatment groups responded to the opportunity to play their money rounds of the game with a new platform that offered an expanded set of controls. The version–called the *combo platform* – provided subjects with access to *both* sets of controls that appeared in the basic platforms: buttons as well as scroll bars (Figure 1c). Note that since subjects trained on only one kind of control and were unaware of the existence of the other, the added controls that appeared on the combo version represented an innovation: the scroll bars would have been novel to those who trained on buttons, and the buttons novel to those who trained on scroll bars.

To insure that the locus of perceived benefits of the combo platform would be isolated to the new control, the function and reliability of the more familiar controls was identical to that which subjects had experienced during the training rounds. Hence, the reliability of the button controls in the combo platform was low for those who trained on low-reliability buttons and high for those who trained on high-reliability buttons. For subjects who trained on the scroll bar, the new button controls were of medium reliability. In addition, the physical appearance of the combo platform was identical to that of each of the basic platforms with the exception of the presence of a second set of controls (Figure 1c).

It should be observed that the design implied that the *objective* incremental value of the new combo platform thus varied depending on the platform on which subjects trained. For subjects who trained on the low-reliability buttons the combo platform subjects access to a more reliable control (the scroll bar) that could potentially allow them to realize significantly higher scores in the money rounds. For subjects who trained on the scroll bars or the high-reliability buttons the objective advantage of the combo version was simply tactile flexibility; since both controls yielded comparable asymptotic levels of achievement (see Figure 2), higher mean achievement mean could be expected only if subjects differed in their natural aptitude for each of the two controls, and made optimal self-selection decisions upon ownership. Of course, subjects could only discover these comparative benefits if they chose to purchase the combo platform and then experimented with the performance of the new control.

The pricing and purchase mechanism. After completing the training phase of the game subjects in the control groups moved on to the money rounds of the game, while those in the treatment read a mock news announcement that a new version had been developed which they had the opportunity to purchase for play during the money rounds rather than the platform they trained on. Subjects were given an illustration of what the new game platform looked like. It was emphasized that the more familiar controls would function just the old ones did, and no statement was made about whether the new control would yield better or worse game results than the old one; subjects were told that the new controls simply gave them greater flexibility in how they controlled their icon.

After reading this announcement subjects were then told that they could acquire the new platform by paying a point handicap that would be applied to their realized score in the money round. Before being shown what this price would be, however, they would have to indicate the maximum price that they would be willing to pay for the game, and they will obtain it if the actual price turns out to be less than this value – an elicitation procedure akin to that suggested by Becker, de Groot, and Marschak (1964). To insure that subjects fully understood how the process would work subjects first participated in a practice round where they set a *WTP* price and an illustrative actual price was drawn by lottery. Subjects were given the opportunity to repeat this exercise until they felt comfortable with the procedure.

The actual price of the combo game was held constant for all subjects at 120 points, a price at which subjects would break even if the new game allowed them to realize a modest (8 point-per-game) increase in performance over the incumbent platform. This price thus implied that subjects who saw the prospect of either only nominal or no improvements in performance with combo platform would play the money rounds with their existing game, whereas subjects with more optimistic estimates would play with the combo game. After subjects submitted *WTP*s, those who submitted valuations greater than 120 were informed that they would be playing with the combo platform, and this the purchase price was immediately reflected as a negative number in the cumulative score box on their game screen (see Figure 1c).

3.2. Results

Among the 68 subjects in the treatment condition who were given the opportunity to purchase the new game platform, 57 (84%) provided willingness-to-pay levels that were sufficient to attain ownership of the combo platform (valuations greater than 120). Hence, on the whole subjects were quite optimistic about the score improvement they could potentially realize by playing the version. A subsequent analysis of the performance of the 11 non-adopters during the money rounds revealed a pattern of achievement similar to that observed among those in the control condition, hence these two groups were pooled in subsequent analyses.

The efficiency of adoption decisions. Subjects' stated willingness to pay for the new product platform is, of course, an implicit forecast of how having the ability to use a second control will improve their score beyond that which could be realized by the basic platform. Since the raw measure of WTPs is highly skewed, we utilize and report log-transformed WTPs in all subsequent analyses unless otherwise noticed. In Figure 3 we plot the mean *WTP* of subjects who adopted the innovation by training condition relative to two standards of achievement: the improvement in scores they actually realized relative to that realized over the last 6 games of the training round (Figure 3a), and the improvement relative to the scores realized by control subjects who did not upgrade (Figure 3b). The figure yields two insights that suggest initial support for **H1** and **H1a**:

1. *Excessive mean optimism in the projected benefits a new control.* The mean stated *WTP* for the new platform across training conditions was 345 game points, equivalent to an expectation that having access to a second control would allow subjects to realize a nearly 20% improvement in score over retaining the basic

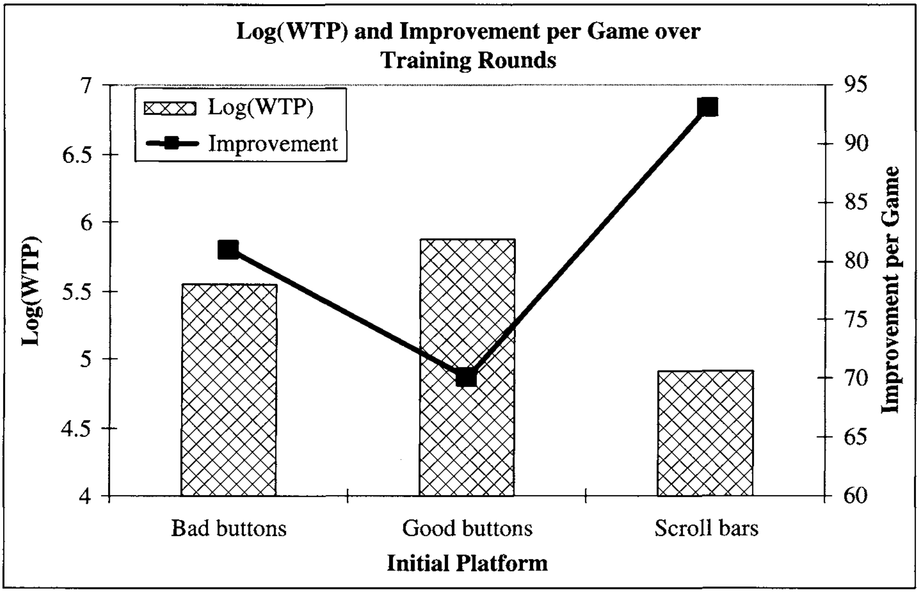


Figure 3. Stated willingness-to-pay for the new combo platform by training platform type: Experiment 1.
Figure 3a. Stated Willingness-to-pay for the new platform and improvement over training period by platform type during training period.

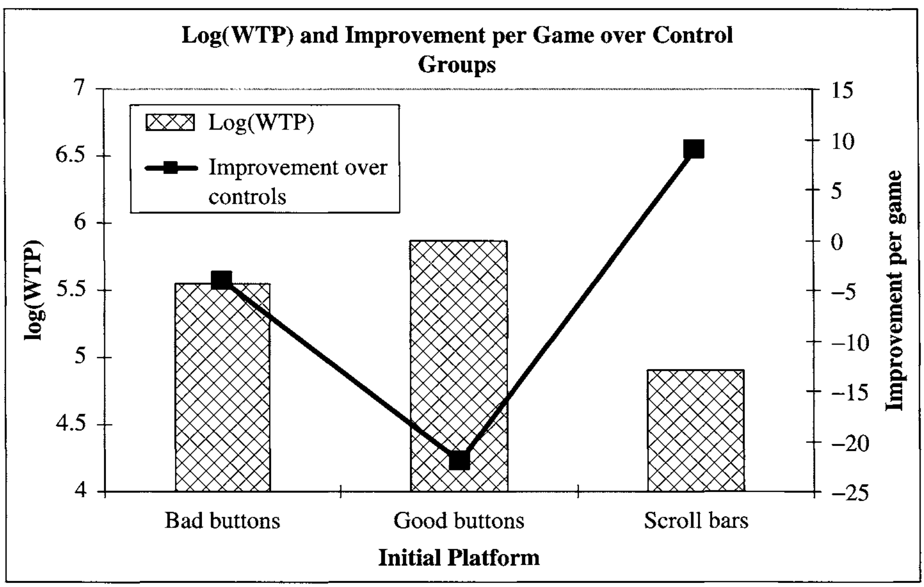


Figure 3b. Stated willingness-to-pay for the new platform and improvement over control groups by platform type during training period.

platform. These implicit forecasts, however, turned out to be quite poor: on average treatment subjects who bought the new platform (henceforth "adopters") realized a mean performance that was on average 130 points *lower* than that realized by those who never upgraded. In addition, the mean *WTP* also exceeded the average increase in raw score by 18% (292 versus the mean WTP of 345). Moreover, WTP was negatively correlated with raw increase in total score ($r = -.40$, $p = .0007$, $N = 68$).

2. *The optimism bias was conditioned by the training platform.* By visual inspection, Figure 3a (the cross-checked bars) offers some initial support to **H1a**. That is, those starting with high-reliability buttons which offered the least frustrating experience also tended to give higher WTPs for the new platform than those starting with scroll-bar platform that is most difficult to learn (389 for high-reliability buttons and 304 for scroll-bar).

To more rigorously explore the effect of training experience on WTPs, we modeled individual estimates as a function of the initial platform and subjects' experience during the training period (we used the maximum score over the last six games during training rounds as the [MAX6] as the proxy for experienced ease of learning). The regression results are presented in Table 1. It is clear that, in support of **H1a**, both factors contributed significantly to stated WTPs. Specifically, subjects who trained on either button platform stated significantly higher WTPs for the new platform than did those who trained on the scroll bar platform, presumably as a result of their better experience with the game in training rounds. In addition, WTPs were positively related to the experienced ease of learning ($p < .02$). Notice that our proxy for experienced ease of learning (MAX6) incorporates the recency bias in retrospective evaluations.

Table 1. Determinants of stated WTPs

Dependent Variable: Log(WTP)

Variable	*Parameter Estimate*	*SE*	*t-value*	*Pr > |t|*
Intercept	3.964	0.464	8.54	<.0001
Initial platform				
Bad buttons	0.765	0.357	2.15	0.0358
Good buttons	0.702	0.384	1.83	0.0725
MAX6	0.005	0.002	2.43	0.0177

$F(3, 63) = 4.35$, $p < .01$
$R - sq = 0.17$
Note: MAX6 = best score over the last six games during training rounds

Table 2. Effect of WTPs on Subsequent Performance

Dependent variable: Cumulative performance during money rounds

Variable	*Parameter Estimate*	*SE*	*t-value*	*Pr* > |*t*|
Intercept	1501.69	376.168	3.99	0.0002
Cumulative performance during training rounds	0.97	0.144	6.77	<.0001
Gender[a]	−370.23	139.000	−2.66	0.0098
Log(WTP)	−170.26	54.023	−3.15	0.0025

$F(3, 63) = 20.89$, $p < .0001$
Adj. $R - sq = 0.475$
Note: [a] 1 = Female and 0 = Male

To more directly examine the degree to which subjects were able to anticipate their actual performance using the combo platform we modeled each player's cumulative score during the money rounds as a function of their average score in the training rounds, their WTPs for the combo game, and gender (see Table 2). The data yield a surprising result: after controlling for training performance, the marginal effect of increasing statements of *WTP* was *negative* ($t(1, 63) = -3.15$; $p = .0025$) among those who purchased the new platform[3]. In short, at the margin those with the most optimistic estimate of how well they would do in the money rounds tended to have the lowest actual achievements. This result is consistent with the pattern of results we reported earlier, i.e. WTP was negatively correlated with raw increases in cumulative score.

Additional insight into why subjects who acquired the new platform may have underperformed relative to their *WTP*s is contained in Figure 4, which plots performance over all 30 trials for treatment versus control subjects by training condition. The figure suggests one contributing explanation for the exaggerated *WTP* estimates: while subjects who bought the new platform seem to have correctly anticipated that their performance would improve on the money trials playing with the new platform, they failed to foresee two factors that would also naturally mitigate achievable relative performance:

1. The fact that there would also be improvements in skill levels playing with the basic platform; and
2. Any potential incremental benefits of the combo version would not be immediately realized as control usage would likely alternate, at least initially, between the two options.

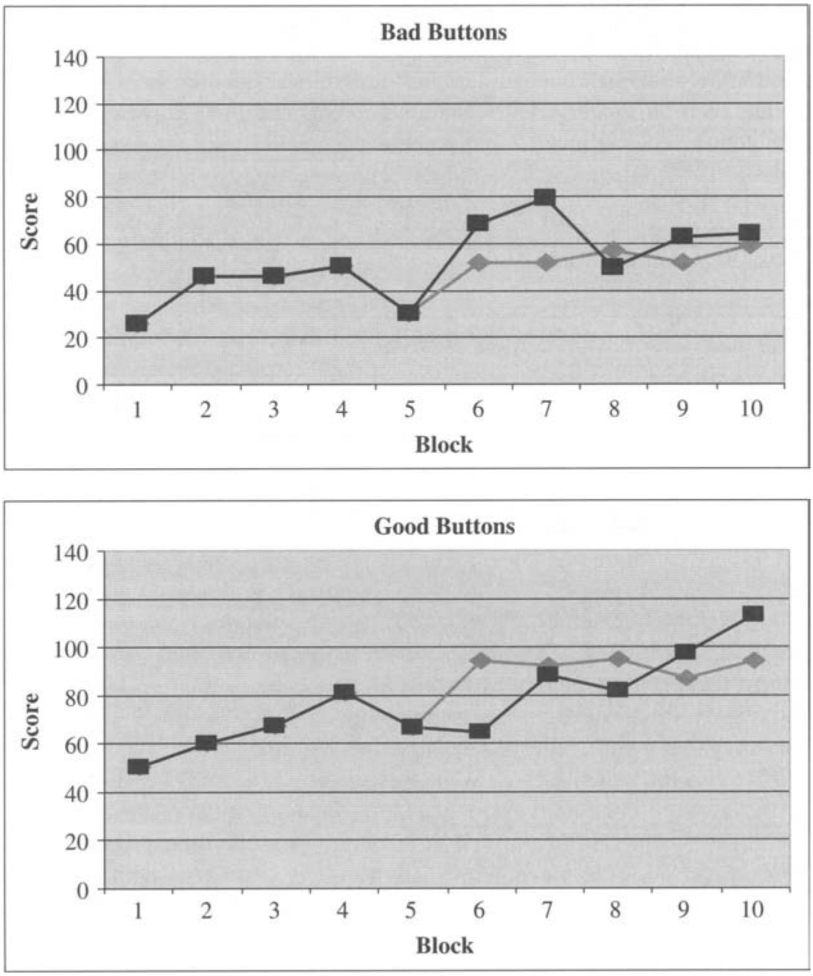


Figure 4. Performance over time by initial platform and upgrade decision.

In short, it is as if the *WTP* estimates reflected a comparison of an envisioned *asymptotic* value of the combo platform to the *current* value of the basic platform – a comparison that naively overlooks the dynamics that would govern actual relative performance during the money period.

Feature utilization. It should be emphasized, of course, that the conclusion that subjects overstated their willingness utilizes knowledge that was not in evidence at the time subjects made these assessments: the objective incremental value of the added control option. Recalling the principles of rational product adoption we discussed at the start, the apparent overvaluation of the combo device might simply be

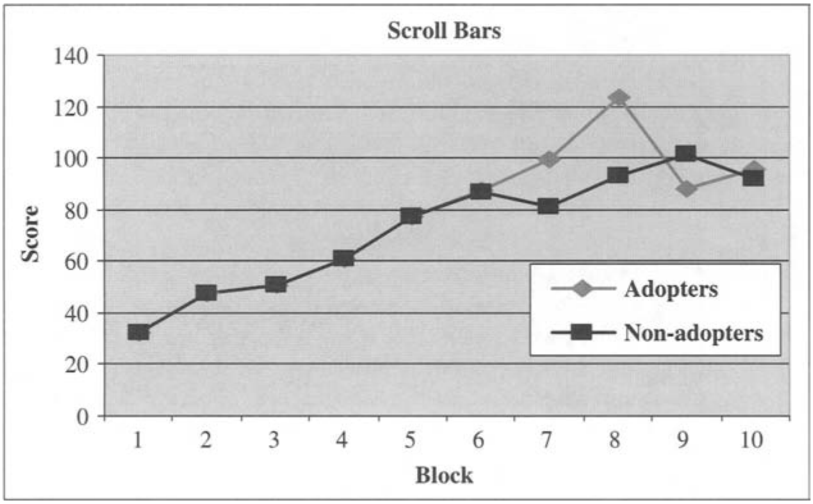


Figure 4. (cont'd)

seen as the case of rational investments in an experiment that did not pay off. That is, subjects who purchased the innovation ended up achieving levels of performance similar to those who did not simply because they discovered, after experimentation, that there was no added value.

In **H2** and **H2a**, however, we hypothesize that while subjects may well acquire the combo platform with well-meaning intentions to learn about its value, its new features will be underutilized, even in settings where there would be a real normative gain. In the current experiment such is the case of subjects who trained on the low-reliability button control. For these subjects the new availability of the scroll bar offered a very real opportunity to increase earnings, though it would require them to incur a period of learning with a control that they are likely initially to find unnatural.

In Figure 5 we plot the proportion of all control actions on the combo platform that were directed at the novel control over trials in the money period of the game. The data give strong apparent support for **H2**: although subjects paid a substantial amount – and were prepared to pay more – for the ability to at least experiment with the use of the new control, few made use of this opportunity. Specifically, during the initial three games (block 6 in Figure 5) of the money period, when utilization of the novel control should rationally have been quite high, subjects who had trained on the high-reliability buttons and the scroll bar utilized the new (reciprocal) control on average only 21% of the time, a level that diminished over time thereafter (Figures 5a and 5b). In addition – and perhaps shockingly – the data revealed 8 subjects in these two conditions who *never* utilized the new controls at all over the entire 15 games.

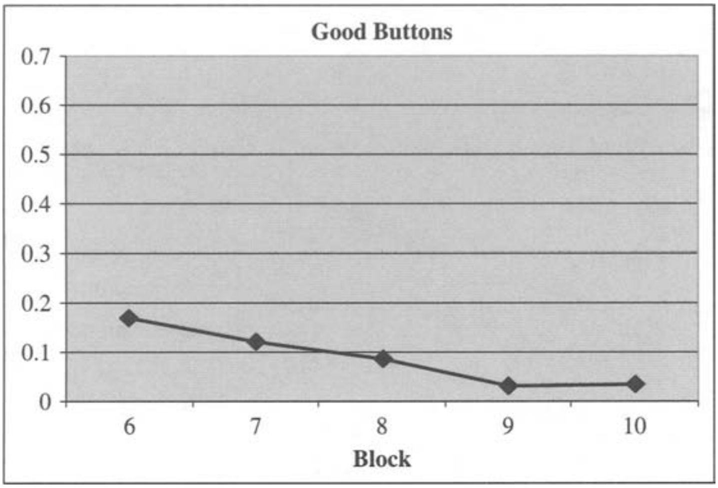


Figure 5. Utilization of new features over time by initial platform.
Figure 5a. Initial platform was Good buttons.

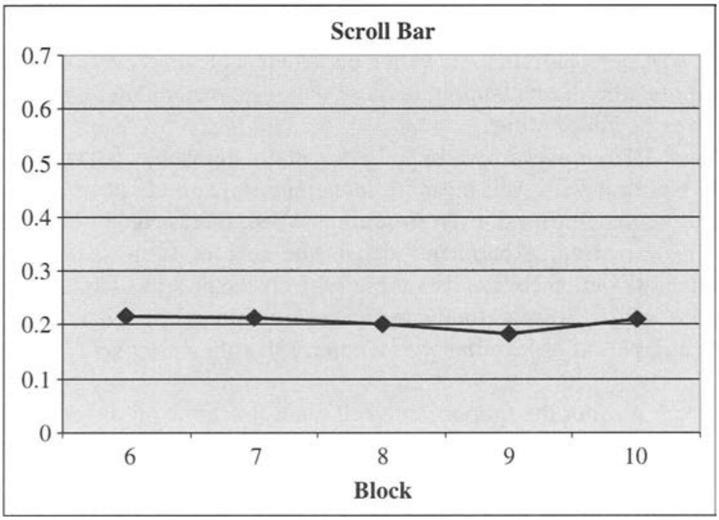


Figure 5b. Initial platform was Scroll bars

Perhaps the most compelling evidence favoring **H2** is found in Figure 5c, which plots the percentage of time subjects who had trained on the low-reliability button utilized the asymptotically superior scroll bar when given the option. On one hand, unlike those who had positive experiences in the training rounds, here we see subjects display a much higher rate of initial usage of the scroll bar, though its level

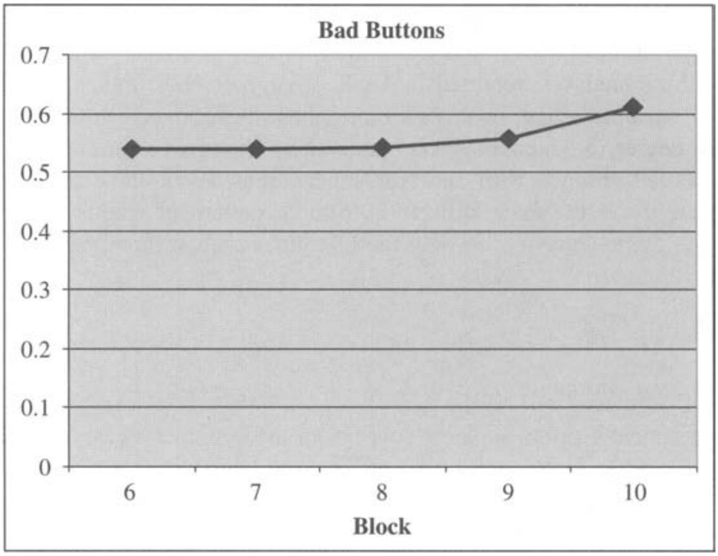


Figure 5c. Initial platform was Bad buttons.

(54%) is still below that which one would normatively prescribe if subjects were active experimenters. In addition, contrary to the normative recommendation, 3 subjects did not start experimenting with the new controls at least until after the first 3 games. On the other hand, the more disturbing feature of the data is that utilization never increased much in the task beyond this level – even though subject would have clearly benefited if it had. In essence, subjects seemed unable to abandon use of a familiar control in favor of a new one, despite the objective inferiority of the former and superiority of the latter.

As a final analysis we examined how individual differences in novel attribute utilization related to subjects' willingness-to-pay for the combo platform. This relationship is, of course, normatively positive; since *WTP* should reflect, in part, the value a subject sees in experimenting with the new control, the higher the *WTP*, the more a subject should invest in its usage, at least until its true value is established. **H2b**, however, predicts the opposite: because high *WTP* measures are theorized to be induced not by rational assessments of the value of information but rather by projected expectations of high immediate returns from the innovation (**H1a**), the more upwardly-biased this assessment, the more likely subjects will be to terminate usage after limited trials.

To test this hypothesis we estimated two models explaining the proportion of uses of the novel control for each subject over games: one that modeled usage as a function of their prior willingness-to-pay for the combo platform as departure from the basic platform, indicator variables for a subject's training platform, and game trial (Model 1 in Table 1), and another that that modeled usage as a function of game trial and three measures of their training experience as in our test of **H1a**: best performance during training, trend, and their interaction (Model 2 in Table 1). The results of these analyses, reported in Table 1, support **H2b**: rather than serving to foster new attribute usage, in this case optimistic beliefs serve to suppress it. The greater the degree to which a subject had positive experiences in the training period and/or emerged from it with superior expectations about the value of the new attribute, the lower the mean utilization. Similar pattern of results still hold if we examine the usage behaviors in only the first game or first three games.

4. EXPERIMENT 2

4.1. Motivation and Description

While Experiment 1 offers apparent support for the hypothesis of over-estimation of the value of product innovations, it leaves uncertain the degree to which the revealed levels of *WTP* accrued directly to over-forecasts of new control attribute utilization and value versus other hedonic drivers, such as a simple attraction to novelty. To resolve this issue a second group of 27 subjects[4] were recruited to replicate a version of the study reported in Experiment 1, but with one major design change. After subjects read the description of the new combo platform and before the elicitation of their willingness-to-pay they were posed with a series of questions designed to tap their beliefs of the value of the new platform. Subjects were asked to make four forecasts:

1. The likely percentage change over all 15 games if they continue to use their existing platform (positive or negative);
2. The likely percentage change over all 15 games if they switch to the new platform (positive or negative);
3. The likely percentage change in their score over the first three games compared to that which would be realized using the basic platform (positive or negative); and
4. The percentage of time during the first three games of the money rounds that they would likely utilize the new control offered by the new platform.

All four forecasts were provided by checking a box on a discrete category scale that offered a range of possible percentage responses (see Appendix).

In addition, after *WTP* measures were elicited and the mock lottery was run, subjects who received the new combo platform were asked an additional series of questions designed to elicit their reasons for setting their *WTP* levels as they did. These took the form of a series of bipolar-scaled question asking for the degree to which they agreed or disagreed with a set of possible motivations for wanting to acquire the combo platform, including expected performance increases, flexibility, aesthetics, and a desire for a change of pace. These process measures are provided in the Appendix.

Table 3. Determinants of New Control Utilization
Dependent Variable: New control usage

Variable	*Model 1*	*Model 2*
	Parameter Estimate	*Parameter Estimate*
Intercept	0.915***	0.751***
Initial platform Bad buttons Good buttons	 0.354*** −0.104***	
Log(WTP)	−0.114***	
MAX6		−0.002***
Game trial	−0.003	−0.0026
Game trial	0.001	0.0008

$F(5, 849) = 62.52, p < .0001$ $F(3, 851) = 41.67, p < .0001$
$R - sq = 0.27$ $R - sq = 0.13$
Note: MAX6 = best score over the last six games during training rounds
*** $p < .001$

Because of the limited size of the subject pool only one (rather than three) training-platform conditions was replicated: the high-reliability button – the condition where the optimism bias was most acute in Experiment 1.

4.2. Results

Mirroring the results of Experiment 1, subjects revealed a high willingness-to-pay (mean raw *WTP*s = 378, *SD* = 233) for the new game platform, with 100% of subjects adopting the new platform given the mock-lottery price. In Table 4 we report a comparison of how their actual performance and new-control utilization compared to the forecasts. The data suggest that the over-forecasts of performance and utilization provide at best a partial explanation for their high levels of willingness-to-pay. Specifically,

1. Subjects did not show excessively optimistic forecasts of their performance with the new platform; in fact, the overall trend is *under*-forecasts (see Table 3) relative to the benchmark of last six games during training rounds, especially for cumulative performance over the entire money rounds. On average, subjects

Table 4. Forecasts of Performance with New Platform and New Control Utilization[a]

	Predicted	*Actual*
WTP	378 (233)	
Performance change over existing platform implied by WTP	23% (14.7%)	
Performance change over 15 games for existing platform[b]	17% (20.1%)	25%[c] (35.1%)
Performance change over 15 games for new platform[b]	9% (18.7%)	32%* (42.7%)
Performance change over first 3 games with new platform[b]	6% (20.2%)	21% (101.5%)
New control utilization over first three games	50% (9.4%)	24%* (31.0%)

Notes:
[a] The number reported are means with standard deviations in parenthesis ($N = 27$).
[b] Percentage change, positive or negative, relative to last six games during training rounds.
[c] Mean for the control group in Experiment 1.
* Predicted is significantly different from actual ($p < .05$).

forecasted a similar percentage improvement in performance with the existing platform roughly correspond to the actual percentage change by control groups in Experiment 1 ($p > .3$). Their mean actual performance in the money rounds was 2241 (*SD* = 7113), comparable to that of the control group in Experiment 1 (mean = 2336, *SD* = 862); but
2. Subjects severely over-forecast the degree to which they would be utilizing the new control; while the average forecast rate during the first three games was 50%, actual usage was closer to 20% in the first three games. On average they used the new controls only about 10% over all 15 games.

Since subjects provided performance forecasts for both the incumbent and new platforms, these estimates can be used to derive normative *implicit WTPs*: estimates of what these assessment *should* have been had they been based solely on their comparative performance forecasts. Implicit *WTP* is thus the forecast performance for the new platform minus that for the incumbent. In Figure 6 we present a scatter plot of implicit verus actual *WTPs*. The figure yields a striking result: the absence of a systematic positive relationship between the two constructs. Indeed, the relationship

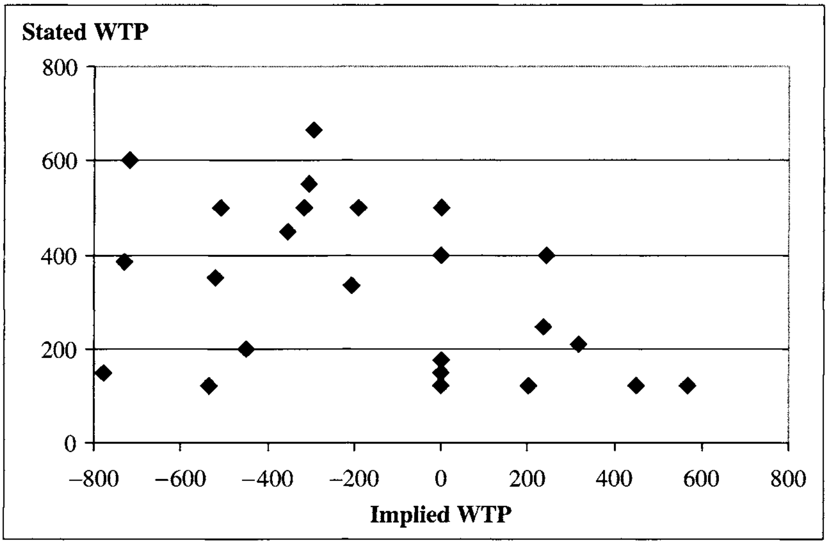


Figure 6. Scatter Plot of Implicit and Stated WTPs
The dashed line represents perfect correlation between implicit and stated WTPs.

between the two measures was nominally negative, with nearly half of subjects (13 out of 27) stating a higher or equal score forecast for the incumbent platform compared to the new, resulting in an aggregate mean implicit raw *WTP* of −154 (*SD* = 360). In contrast, recall that their average stated *WTPs* revealed only a moment later was 378 (*SD* = 223). Hence, at the very least, the data reject the idea that subjects formed assessments of WTP by contrasting forecasts of the value of the new platform with what this *would* have been had they stayed with the old one.

What were subjects' own reasons for being attracted to the new game platform? In Table 5 we summarize the degree to which subjects, on average, agreed or disagreed with each of four possible reasons for preferring to acquire the combo platform. The data suggest that if hedonic or boredom-related factors were influencing this decision, subjects were not inclined to admit to them: subjects were most inclined to agree that the decisions were motivated by the two factors that high *WTP* valuations *should* have been based on: expectations of higher performance a desire for flexibility. In contrast, subjects were less inclined to agree that attraction was motivated by the aesthetic appeal of the new platform and nominally disagreed that it was motivated by task boredom.

One interpretation of the data is thus that approached their assessments of *WTP* with normatively-correct beliefs about the factors that should drive these evaluations. Where the *WTP* assessments went awry was a tendency to overestimate the degree to which they would utilize the new control feature (the value of flexibility) and an underestimate the degree to which their performance would likely improve to

Table 5. Reasons for upgrade

Reason for upgrade	*Mean*	*SD*
New controls are useful	0.76*	1.48
New controls offer flexibility	0.48*	1.05
New platform is aesthetically more pleasing	0.28	1.24
Desire for a change of pace	−0.12	1.59

Note:
Items use a 7-point scale, anchored by −3 Disagree a lot to 3 Agree a lot.
* Mean is significantly different from 0, $p < .05$

a similar degree using the older platform. What is curious about this latter result is that subjects *did* reveal such knowledge when directly asked what theirwould performance would be using the older platform; it was simply not incorporated when WTP judgments were made.

5. DISCUSSION

One often hears it said that consumers frequently buy more technology than they can realistically make use of. We compare products by counting the number of features they offer (often without knowing what they are used for), and feel bad when we discover that our incumbent devices are no longer state-of-the-art, regardless of how adequately they may be fulfilling our needs. And this apparent bias is by no means limited to consumers; firms as well have recently been criticized as well for being prone to invest in more material and programmatic innovations that actually get implemented (Kim, Pae, Han, and Srivastava 2002). On the other hand, it also seems to be the case that firms who sell technology also worry about the possibility that consumers may walk away from technologies that are seen as *too* innovative. Hence, for example, when releasing Windows *XP* took pains to insure that its innovation would be seen by consumers as only modestly different from its old operating systems, to the point of allowing users the option eliminate new screen views if they wanted (through the "revert to classic view" command).

Yet, as pervasive as these observations may be, there has been little prior work that has formally studied the biases that characterize consumer new-technology adoption and subsequent usage decisions, and the psychology that may underlie these biases. The goal of this paper was to take a step toward gaining this knowledge by observing how a individuals made decisions whether or not to buy a new technology – an improved gaming device – in a laboratory setting where we could measure both the actual and perceived value of the technology as well as how it was utilized after purchase, and manipulate the kind of experiences with prior similar technologies that subjects had coming into the buying decision.

Central to the work was the idea that a general over-buying bias may, in fact, have a strong theoretical basis. Drawing on prior work in affective forecasting. we hypothesized that when buying new technologies consumers will usually have a difficult time anticipating how they will utilize a product after it is purchased, and will be prone to believe that the benefits of attribute innovations that are perceived now will project in a simple fashion into the future. Implicit to this over-forecast is a tendency to underestimate the impact of factors that may likely serve to diminish usage in the future that are not in evidence now, such as frustration during learning and satiation. Consequently, there is a tendency for consumers to systematically evaluate product innovations through rose-colored glasses, imagining that they will have a larger and more positive impact on the future lives than they most often will likely end up having.

The experimental data reported here provide strong apparent support for this view of new product valuation. What is notable about the current demonstration is that the evidence for the optimism bias we report was derived from a context designed to facilitate rational assessments of innovation value. Specifically, subjects were given a clearly-stated metric by which the objective value of the innovation would be assessed, there was a direct monetary penalty for overstating value (the game innovation was paid for by a point deduction), and the innovation itself was a purely functional rather than aesthetic one (a new control added to the same graphic game platform). Yet, subjects still succumbed to the same biases that we suggest may be pervasive in real markets: a tendency to overvalue prospective innovations, and then under-utilize their features upon acquisition, even for the limited purposes of experimentation (e.g., in the current task 14 out of 84 subjects across two experiments who purchased the innovation never used its added control at all).

While the current findings offer support for the hypothesized effects of product enhancements, care must obviously be taken before presuming that the findings will hold in all product-adoption settings. First, a quite natural question – one that the current work only partially addresses – is the effect of long-term learning on the enhancement bias. For example, it is natural to argue that once a consumer recognizes that they have overbought a technology they will be less inclined to do the same the next time around; i.e., they would become more astute forecasters of how they really make use of new technologies. Consistent with this, subjects who expressed the most conservative willingness-to-pay for the game innovation were those who experienced the highest learning costs with the game on which they first trained. Yet, even these subjects overvalued the new control, suggesting that the lure of innovative features may difficult to overcome. In addition, it must be recalled that in most real-world settings new technologies are bought with sufficient rarity that carry-over effects of all kinds may be quite limited.

Another important question that might be raised is that the findings of this research seem, at first blush, to cut against the grain of those who have offered evidence of an *under*valuing of new technology options due to lock-in effects, such as those reported by Johnson, Bell, and Lhose (2003; in web-site visitation patterns) and Zauberman (2003; in valuations of new search engines). The current work differs from these efforts, however, in that here we consider the case of customer valuations of new products that offer *separable enhancements* to an existing platform; that is, by buying the new product one does not have to abandon what one has already learned (i.e., start a new learning curve). Congruent with their findings, however, we observe the same type of lock-in effect once an acquisition decision is made about whether to utilize the new attribute control instead of the old one: subjects who developed a strong familiarity with one type of motion control tended to stick with it even when they paid for the option to use an alternate, and even when the new control offered an objective normative benefit. Hence, the current work suggests that when new attributes are bundled with familiar ones consumers seem unable to anticipate the effects of lock-in before hand, resulting in consistent overvaluations.

Finally, an important goal of future work would be to better resolve the psychological mechanisms that underlie consumer assessments of product novelty. In this work we show that these assessments are biased in a way that is consistent with biases found in other domains of hedonic judgment, and for some of the same apparent reasons – for example, failing to anticipate future reluctance to experiment with the new control. But account is clearly a blunt one; the actual mechanism by which consumers develop visions of the future through analogical and other forms of structured reasoning (e.g., Moreau, Lehmann, and Markham 2001) is clearly a complex one, and more thoroughly understanding it may help better resolve the empirical boundaries of assessment biases and, possible, their correction.

ACKNOWLEDGMENT

Authors are listed reverse-alphabetically. Shenghui Zhao is a doctoral candidate in Marketing at the same institution, Robert Meyer is the Gayfryd Steinberg Professor of Marketing at the Wharton School of Business, University of Pennsylvania, and Jin Han is an Associate Professor of Marketing at Singapore Management University. This work was supported by a 2003 research grant from the Wharton-Singapore Management University Research Center.

The Rationality of Consumer Decisions to Adopt and Use New Technologies: Why Are We Lured by Product Features We Never Use?

NOTES

[1] http://biz.yahoo.com/prnews/030701/sftu019a_2.html

[2] In a standard analysis these beliefs would be assumed to evolve as a Markov process given decision to utilize δ; that is, associated with z_t is a first-order cumulative conditional distribution function $G(z', z)$.

[3] Excluding those who chose not to upgrade obtained similar results.

[4] We ran two groups including a control group of combo-combo (N = 33). I left it out of the analysis since our focus is on the process measures and their relations to stated WTP.

REFERENCES

Ariely, Dan (1998). "Combining experiences over time: The effects of duration, intensity changes & on-line measurements on retrospective pain evaluations." *Journal of Behavioral Decision Making*, 11, 19–45.

Ariely, Dan, and George Loewenstein (2000). "The Importance of Duration in Ratings of, and Choices Between, Sequences of Outcomes." *Journal of Experimental Psychology: General*, 129(4), 508–523.

Becker, Gordon. M., Morris H. DeGroot, and Jacob Marschak (1964). "Measuring Utility by a Single-Response Sequential Method." *Behavioral Science*, 9(July), 226–232.

DellaVigna, Stefano, and Ulrike Malmendier (2002). "Self Control in the Market: Evidence from the Health Club Industry." *Working Paper*, Department of Economics, University of California, Berkeley.

Dhar, Ravi (1997). "Consumer Preference for a No-Choice Option." *Journal of Consumer Research*, 24(2), 215–231.

Fredrickson, B. L., and Daniel Kahneman (1993). "Duration Neglect in Retrospective Evaluations of Affective Episodes." *Journal of Personality and Social Psychology*, 65, 45–55.

Gilbert, Daniel, and J. B. Ebert (2002). "Decisions and Revisions: The Affective Forecasting of Changeable Outcomes." *Journal of Personality and Social Psychology*, 82, 502–514.

Gilbert, Daniel T., M. J. Gill, and Timothy D. Wilson (2002). "The Future is Now: Temporal Correlation and Affective Forecasting." *Organizational Behavior and Human Decision Processes*, 88, 690–700.

Gilbert, Daniel, Elizabeth C. Pinel, Timothy D. Wilson, Stephen J. Blumberg, and Thalia P. Wheatley (1998). "Immune Neglect: A Source of Durability Bias in Affective Forecasting." *Journal of Personality and Social Psychology*, 75(3), 617–638.

Gourville, John, and Dilip Soman (1998). "Payment Depreciation: The Behavioral Effect of Temporally Separating Payments." *Journal of Consumer Research*, 25(2), 160–174.

Grice, H. Paul (1975). "Logic and Conversation," in J. Cole and P. Morgan (eds), *Syntax and Semantics, Volume 3, Speech Acts*, pp. 43–58. New York: Academic Press.

Hogarth, Robin M. (1981). "Beyond Discrete Biases: Functional and Dysfunctional Aspects of Judgment Heuristics." *Psychological Bulletin*, 90, 197–217.

Johnson, Eric J., Steven Bell, Gerald L. Lhose, Jerald (2003). "Cognitive Lock-in and the Power Law of Practice." *Journal of Marketing*, 67 (April).

Kahneman, Daniel, J. Knetsch, and R. Thaler (1990). "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 98, 1325–1348.

Kahneman, Daniel, and J. Snell (1992). "Predicting Future Taste: Do People Know What they Will Like?" *Journal of Behavioral Decision Making*, 5, 1987–2000.

Kim, Namwoon, Jae H. Pae, Jin K. Han, and Rajendra K. Strivastava (2002). "Institutionalization of Technologies for Technology Markets: A Buyer-Seller Approach." *Working Paper*, Hong Kong Polytechnic University.

Klemperer, Paul (1987). "Markets With Consumer Switching Costs." *Quarterly Journal of Economics*, 102(May), 375–394.

Loewenstein, George, and D. Alder (1996). "A Bias in the Prediction of Tastes." *Economic Journal*, 105, 929–937.

Loewenstein, George, and David Schkade (1999). "Wouldn't it be Nice? Predicting Future Feelings," in Daniel Kahneman, Edward Diener, and Norbert Schwartz (eds.), *Well Being: The Foundations of Hedonic Psychology*, New York: Russell Sage Foundation, 85–105.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin (2002). "Projection Bias in Predicting Future Utility." *Working PaperE00-284*, Department of Economics, University of California, Berkeley.

Meyer, Robert J., and J. Wesley Hutchinson (1994). "Dynamic Decision Making." *Marketing Letters*, XX, 000.

Meyer, Robert J., and J. Wesley Hutchinson (2001). "Bubbling Geniuses: The Power of Everyday Reasoning in Mulitstage Decision Making", in S. Hoch and H. Kunreuther (eds), Wharton on Making Decisions", New York: John Wiley.

Meyer, Robert J., and Young Shi (1995). "Intuitive Solutions to the Armed-Bandit Problem." *Management Science*, Moreau, C. Page (2001). Donald R. Lehmann, and Arthur B. Markman, "Entrenched Knowledge Structures and Consumer Response to Really New Products." *Journal of Marketing Research* 38(February), 14–29.

Miller, Elizabeth, and Barbara E. Kahn (2003). "The Effect of Color and Flavor Names on Consumer Choices". *Working Paper*, Department of Marketing, the Wharton School of Business, University of Pennsylvania.

Norman, Donald A. (1998). *The Invisible Computer: Why Good Products can Fail, the Personal Computer is so Complex, and Information Appliances are the Answer.* Cambridge, MA: The MIT Press.

O'Donoghue, Ted, and Mathew Rabin (1999). "Doing it Now or Later." *American Economic Review*, 89, 103–124.

Read, D. and B. van Leeuwen "Predicting Hunger: the Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes*, 76, 189–205.

Rust, John (1992). "Do People Behave According to Bellman's Principle of Optimality? *Working Paper E92-10*, the Hoover Institution, Stanford University.

Shin, Jiwoong, and Dan Ariely (2003). "Keeping Doors Open: The Effect of Unavailability on Incentives to Keep Options Viable", *Working Paper*, Sloan School of Business, MIT.

Shivardi, Fabiano, and Martin Schneider, "Strategic Experimentation and Disruptive Technological Change." *Working Paper*, Department of Economics, University of Rochester.

Siegler, Robert, Judy DeLoache, and Nancy Eisenberg (2003). *How Children Develop.* New York: Worth.

Simonson, Itamar (1990). "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior." *Journal of Marketing Research*, Vol. 32, pp. 150–162.

Soman, Dilip (2003). "Prospective and Retrospective Evaluations of Experiences: How you Evaluate and Experience Depends on When you Evaluate it." *Journal of Behavioral Decision Making*, 16(1), 35–52.

Thaler, Richard (1980). "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*, 1(1), 36–90.

Thaler, Richard (1985). "Mental Accounting and Consumer Choice." *Marketing Science*, 4(Summer), 199–214.

Tversky, Amos, and Eldar Shafir (1992). "Choice under conflict: The Dynamics of Deferred Decision." *Psychological Science*, 3(6), 358–361.

Wilson, Timothy D. and Daniel Gilbert (2003). "Active Forecasting," in M. Zanna (ed.), *Advances in Experimental Social Psychology*, 35, 345–411. New York: Elsevier.

Wood, Stacy L., and John G. Lynch (2002). "Prior Knowledge and Complacency in New Product Learning." *Journal of Consumer Research*, 29(Decembert), 416–426,

Zauberman, Gal (2003). "Lock-In from an Inter-temporal Choice Perspective: The Combined Effect of the Information Cost Structure and Time Preferences." *Journal of Consumer Research*, 30 (December), in press.

APPENDIX: PROCESS MEASURES USED IN EXPERIMENT 2

Forecasts

Before you proceed, please answer the following question to your best ability and knowledge.

1. During the Practice games, you earned X points over 15 games. If you were to continue to play the SAME game, how many points would you expect to earn? (choose one)

O	O	O	O	O	O	O
50% or more LOWER			About the same			50% or more HIGHER

2. During the Practice games, you earned X points over 15 games. If you were to SWITCH to the NEW game, how many points OVER THE NEXT 15 GAMES would you expect to earn relative to your current performance?

O	O	O	O	O	O	O
50% or more LOWER			About the same			50% or more HIGHER

3. If you were to SWITCH to the NEW game, how many points OVER THE NEXT 3 GAMES would you expect to earn relative to your current performance?

O	O	O	O	O	O	O
50% or more LOWER			About the same			50% or more HIGHER

4. If you were to SWITCH to the NEW game, how much of the time would you expect to use the NEW controls?

O	O	O	O	O	O	O
Not at all						50% or more

REASONS FOR UPGRADING TO THE NEW PLATFORM

You've decided to buy the new game. Please indicate how much you agree or disagree with each of the following statements*.

1. The new controls will help me to score more points than the earlier game.
2. I got the new game because I just wanted a change.
3. The new controls offer me more flexibility in playing the game.
4. The new game is aesthetically more pleasing than the earlier game.

O	O	O	O	O	O	O
Agree a lot			Neutral			Disagree a lot

* All items used the same scale and responses were later reverse coded so that Disagree a lot = −3 and Agree a lot = +3.

This page intentionally blank

Chapter 2

A BEHAVIORAL ACCOUNTING STUDY OF STRATEGIC INTERACTION IN A TAX COMPLIANCE GAME

Chung K. Kim
Ewha Womans University, Korea

William S. Waller
University of Arizona and
Nanyang Technological University

Abstract

This paper reports an experiment on a tax compliance game based on the model of Graetz, Reinganum, and Wilde (1986). A model implication is that the audit rate, β, is insensitive to the proportion of strategic versus ethical taxpayers, ρ. Our hypotheses contrarily predict that auditors with limited rationality use ρ as a cue for adjusting β. The hypotheses assume a simple additive process: $\beta = \beta' + \beta''$, where β' depends on ρ, and β'' depends on a belief about the taxpayer's strategy. The results show positive associations between ρ and β', and between auditors' uncertainty about ρ and β'. The auditors formed incorrect beliefs about the taxpayers' responses, which affected β''. The auditors incorrectly believed that the taxpayers increased the rate of under-reporting income as ρ increased, and that the taxpayers expected a higher audit rate when the auditors faced uncertainty about ρ. The taxpayers correctly believed that β increased as ρ increased, and responded by decreasing the rate of under-reporting income.

1. INTRODUCTION

Behavioral accounting research tests hypotheses regarding the implications of realistic assumptions about human rationality for economic decisions in accounting settings (Waller, 2002). One such assumption is that decision-makers face cognitive limitations which prevent them from acting as if they maximize expected utility (Simon, 1982). As thoroughly documented in the psychology literature, cognitive limitations can lead to systematic decision biases (Connolly, Arkes, and Hammond, 2000). Drawing from psychology, behavioral accounting researchers normally test

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 35–56.
© 2005 *Springer. Printed in the Netherlands.*

their hypotheses with experimental methods that emphasize internal validity, sometimes at the expense of external validity. As in psychology, most behavioral accounting experiments examine the behavior of isolated subjects who respond to an exogenous choice or information set. This approach has the advantage of screening out many sources of noise, especially the responses of other subjects, which sharpens the focus on an individual's decision process.

Critics of behavioral accounting experiments raise two related questions. The first question concerns rationality: how should subjects without cognitive limitations act in the experimental setting? Formal theories of accounting establish stylized settings of competitive or strategic interaction, and derive the equilibrium consequences for the role of information. To maintain tractability, the formal theories assume that decision-makers maximize expected utility. This theoretical perspective prompts the question about rational behavior in a specific experimental setting, in order to evaluate if not predict the subjects' behavior. Although some behavioral accounting experiments employ a normative model or principle, most fail to relate the experimental design or evidence to any formal theory of accounting. The second question concerns external validity: do conclusions about the subjects' behavior in individual settings generalize to the aggregate economic settings that are relevant to accounting? An important mechanism in these economic settings is the competitive or strategic interaction of players with conflicting preferences, which would discipline and potentially eliminate the decision biases found in individual settings. Arbitrage traders in equity markets would take advantage of unsophisticated traders who systematically over-react to accounting disclosures. Auditees reporting income would take advantage of auditors who systematically under-react to cues indicating income manipulation.

Behavioral accounting research can address both questions by incorporating three steps into the experimental design. First, accounting experimenters can assign subjects to opposing roles, e.g., auditors and auditees, and observe their strategic interaction in game settings (Camerer, 2003). Besides addressing the question about external validity, this step allows experimenters to expand their research agenda. How does an auditor with limited rationality form a belief about the strategy choice of an auditee also with limited rationality? Second, in the tradition of experimental economics (Smith, 2000), accounting experimenters can operationalize a game-theoretic model based on the assumption that decision-makers maximize expected utility. This step addresses the question about rationality, by providing a basis for determining the strategy choices of ideal players with unlimited rationality and for evaluating the strategy choices of real players with limited rationality. Third, in the tradition of behavioral decision research (Kahneman and Tversky, 2000), accounting experimenters can manipulate a variable that, although normatively irrelevant in the game-theoretic model, is hypothesized to affect the behavior of players with limited rationality.

This study reports an experiment that used the three-step approach to examine behavior in a tax compliance game. The experiment operationalized the game-theoretic model of Graetz, Reinganum, and Wilde (1986), who provided a seminal analysis of the tax compliance problem (for literature reviews, see Andreoni, Erard, and Feinstein, 1998; Cuccia, 1994).[1] In the model, the taxpayer chooses a strategy $\{\alpha, 1-\alpha\}$ when true income is high, whereby he under-reports income with probability α and honestly reports income with probability $1-\alpha$. The auditor chooses a strategy $\{\beta, 1-\beta\}$ when reported income is low, whereby she conducts a costly audit with probability β and does not audit with probability $1-\beta$. The model assumes two taxpayer types: proportion ρ are strategic taxpayers who maximize expected wealth, and proportion $1-\rho$ are ethical taxpayers who adhere to an internalized norm for honesty. The auditor maximizes expected net revenue, i.e., tax plus fine minus audit cost. Before conducting an audit, the auditor cannot distinguish the taxpayer types. When the auditor conducts an audit and detects under-reporting, the taxpayer must pay a fine plus the tax for high true income. An implication of the model is that the optimal audit rate β^* is insensitive to an exogenous change in ρ, as long as ρ exceeds a threshold. The strategic taxpayer fully absorbs the change in ρ by adjusting the optimal rate of under-reporting income α^*.

Contrary to the model-based implication, our hypotheses predict that the auditor with limited rationality is sensitive to both the level of ρ and uncertainty about ρ. To represent limited rationality, we propose that the auditor chooses the audit rate through a simple additive process:

$$\beta = \beta' + \beta'', \tag{1}$$

where β' is a function of the factors that directly affect the auditor's choice, e.g., audit cost, and β'' is a function of the auditor's belief about the taxpayer's strategy, $E_a(\alpha)$.[2] To measure β' and β'', we performed a regression for each auditor:

$$\beta = \mathrm{A} + \mathrm{B} \cdot E_a(\alpha) + \varepsilon. \tag{2}$$

B indicates the sensitivity of the auditor's strategy choice to a change in belief about the taxpayer's strategy. For each round, we used $\{\mathrm{A} + \varepsilon\}$ as a measure of β', and $\{\mathrm{B} \cdot E_a(\alpha)\}$ as a measure of β''. We hypothesize that β' increases with increases in the level of ρ and in the auditor's uncertainty about ρ, and that β'' changes with changes in the level of ρ and in the auditor's uncertainty about ρ.[3]

The experimental procedure assigned subjects to the role of strategic taxpayer or auditor. A computer automated the role of ethical taxpayer. At the start of each round, the strategic taxpayers chose a rate for under-reporting income when true income was high, and the auditors chose an audit rate when reported income was low. Before stating their strategy choice, all subjects provided an estimate of their opponent's strategy. Over 20 rounds, the procedure randomly varied ρ among three levels (0.25, 0.50, 0.75). The procedure also manipulated the auditors' uncertainty about ρ on a between-group basis. In one group, the auditors knew the level of ρ before stating their audit rate for the round. In the other group, the auditors knew that the three levels of ρ were equally likely, but did not know the level of ρ until the end of the round. In all cases, the strategic taxpayers knew ρ, and they knew whether the auditors knew ρ. Because each level of ρ exceeded a threshold, the optimal audit rate was a constant, regardless of round-by-round variation in the level of ρ and regardless of the auditors' uncertainty about ρ.

As hypothesized, the results showed a significantly positive relationship between the level of ρ and β'. The auditors who knew the level of ρ increased β' as the proportion of potential tax cheaters increased. In addition, β' was significantly higher for the auditors with uncertainty about ρ. The results for β'' were less straightforward. The relationship between the level of ρ and β'' was positive but insignificant, indicating that the auditors' belief about how the taxpayers reacted to changes in ρ had a modest effect on the audit rate. However, β'' was significantly lower for the auditors who faced uncertainty about ρ, indicating that the auditors incorrectly believed that the taxpayers raised their estimate of the audit rate when the auditors faced uncertainty about ρ. For their part, the taxpayers were sensitive to the auditors' use of ρ as a cue for adjusting the audit rate. The taxpayers correctly believed that the audit rate increased as the level of ρ increased, and responded by decreasing the rate of under-reporting income.

2. GAME SETTING

The setting involves the interaction of a taxpayer and an auditor (Graetz et al., 1986). The population of taxpayers includes a proportion ρ of strategic taxpayers who under-report income given the right incentive, and a proportion $1 - \rho$ of ethical taxpayers who always report honestly. True income is either low or high, and reported income is either low or high. All players know the probability of high rather than low true income. Given low true income, all taxpayers report low income, because there is never an incentive to over-report income. Given high true income, ethical taxpayers report high income, but strategic taxpayers report low income with probability α, and report high income with probability $1 - \alpha$. The amount of tax is higher when reported income is high rather than low.

Given low reported income, the auditor decides whether to conduct a costly audit that reveals true income. Before conducting an audit, the auditor cannot distinguish among an honest report from an ethical taxpayer with low true income, an honest report from a strategic taxpayer with low true income, and under-reporting by a strategic taxpayer with high true income. When the auditor detects under-reporting, the taxpayer must pay a fine in addition to the tax on high income. The auditor's cost to conduct an audit is positive but less than the sum of the fine plus the tax difference for high versus low income. Given low reported income, the auditor conducts an audit with probability β, and no audit with probability $1 - \beta$. Given high reported income, there is never an audit.

Figure 1 shows the players' best-response functions, which intersect at α^* and β^*. α^* is the rate of under-reporting income at which the auditor's return from conducting an audit is zero:

$$\alpha^* = C(1 - P)/[\rho \cdot P(F + T - C)], \tag{3}$$

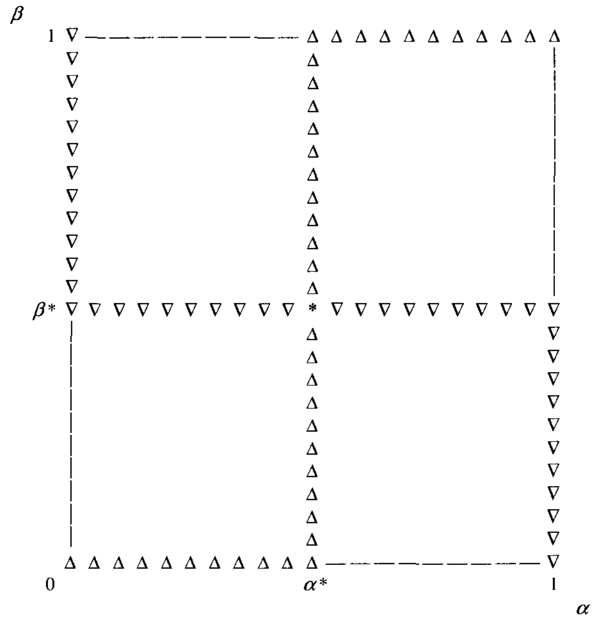


Figure 1. Best-response functions and optimal strategies.
The figure shows the best-response functions of the strategic taxpayer (∇) and auditor (Δ). The players' best-response functions intersect at α^* and β^*.

where C is the audit cost, P is the probability of high true income, F is the fine for detected under-reporting, and T is the tax difference for high versus low income. If α is less than α^*, then the auditor never conducts an audit when reported income is low. If α is more than α^*, and α^* is less than one, then the auditor always conducts an audit when reported income is low.[4] β^* is the audit rate at which the strategic taxpayer's return from under-reporting is zero:

$$\beta^* = T/[F + T]. \tag{4}$$

If β is less than β^*, then the strategic taxpayer always under-reports when true income is high. If β is more than β^*, then the strategic taxpayer never under-reports when true income is high. Equation 3 implies an inverse relationship between the level of ρ and the optimal rate of under-reporting income. Equation 4 implies that the optimal audit rate is not sensitive to the level of ρ.

The experiment fixed most of the model's parameters for each of 20 rounds. In each round, there was a 0.50 probability that true income was high (100 francs), and a 0.50 probability that true income was low (0 franc). The tax was 50 francs for reported income of 100 francs, and the tax was 0 franc for reported income of 0 franc. The audit cost was 10 francs. The fine for detected under-reporting was

50 francs. The level of ρ randomly varied round-by-round among three levels (0.25, 0.50, 0.75). The optimal audit rate was a constant, because each level of ρ exceeded a threshold (0.11).[5] Given the above parameter values, the optimal strategies in the experiment were:

$\beta^* = 0.50$, regardless of ρ,

$\alpha^* = 0.44$, when ρ was 0.25,

$\alpha^* = 0.22$, when ρ was 0.50, and

$\alpha^* = 0.15$, when ρ was 0.75.

These optimal strategies assume unlimited rationality, and provide a benchmark for evaluating the subjects' behavior. To predict the subjects' behavior, the following hypotheses assume limited rationality.[6]

3. HYPOTHESES

Our hypotheses assume that the auditor with limited rationality chooses the audit rate through an additive process:

$$\beta = \beta' + \beta''. \tag{5}$$

β' is a function of the factors that directly affect the auditor's choice:

$$\beta' = f(C, F, T, \rho). \tag{6}$$

For example, a decrease in the audit cost implies an increase in β', other things held constant. β'' is a function of the auditor's belief about the taxpayer's strategy, $E_a(\alpha)$:

$$\beta'' = g(E_a(\alpha)). \tag{7}$$

To assess $E_a(\alpha)$, the auditor assumes that the taxpayer also employs an additive process:

$$E_a(\alpha) = E_a(\alpha') + E_a(\alpha''), \tag{8}$$

where α' is based on the factors that directly affect the taxpayer's choice, and α'' is based on the taxpayer's estimate of the audit rate. $E_a(\alpha')$ reflects the auditor's belief about how the taxpayer responds to factors such as the fine for detected underreporting. $E_a(\alpha'')$ reflects the auditor's belief about the taxpayer's estimate of the audit rate. Incorporating such beliefs into the choice of β is the upper limit for strategic reasoning by the auditor with limited rationality.

Effects of ρ. The first set of hypotheses (H1-H3) predicts that the auditor with limited rationality responds to round-by-round variation in the level of ρ by adjusting the audit rate, even though ρ is normatively irrelevant to the optimal audit rate. H1 assumes that the auditor perceives ρ as a factor that directly affects her choice, so that changes in the level of ρ affect β'. When the taxpayer population includes a higher proportion of potential cheaters, the auditor increases the audit rate, even if no taxpayer changes the rate of under-reporting income. A higher audit rate is necessary, simply because there are more potential tax cheaters.

H1. When the auditor knows the level of ρ, there is a positive association between the level of ρ and β'.

H2 and H3 assume that the auditor also perceives ρ as a factor that directly affects the taxpayer's choice, so that changes in the level of ρ affect $E_a(\alpha)$ and β''. Any factor that directly affects both the auditor and taxpayer is likely to produce countervailing effects on $E_a(\alpha')$ and $E_a(\alpha'')$. Regarding $E_a(\alpha')$, the level of ρ evokes a social norm for the taxpayer to "look at what others are doing and follow the majority" (Elster, 1989, p. 56). As the proportion of potential cheaters increases, the norm dictates that the taxpayer should cheat more. This consideration implies a positive association between the level of ρ and $E_a(\alpha)$. Regarding $E_a(\alpha'')$, the taxpayer expects a higher audit rate as the level of ρ increases, consistent with H1, and responds by decreasing the rate of under-reporting. This consideration implies a negative association between the level of ρ and $E_a(\alpha)$. Because the net effect of these considerations could be either positive or negative, H2 and H3 are non-directional. H2 predicts that changes in the level of ρ induce the auditor to revise $E_a(\alpha)$, and H3 predicts that changes in the level of ρ induce the auditor to revise β''.

H2. When the auditor knows the level of ρ, changes in the level of ρ induce the auditor to revise $E_a(\alpha)$.

H3. When the auditor knows the level of ρ, changes in the level of ρ induce the auditor to revise β''.

Effects of Auditors' Uncertainty about ρ. The second set of hypotheses (H4-H6) predicts that the auditor with limited rationality responds to uncertainty about ρ by adjusting the audit rate, even though ρ is normatively irrelevant to the optimal audit rate. Beginning with Ellsberg (1961), there have been many demonstrations that individuals react to missing information about probability parameters as if they were ambiguity-averse: they prefer choices with known parameters, and desire more information about choices with unknown parameters, other things held constant (Camerer and Weber, 1992; Frisch and Baron, 1986).[7] In an experimental audit setting where auditees had an incentive to over-state their asset value, Zimbelman and Waller (1999) found that auditors increased their costly sampling to compensate for uncertainty about the asset valuation process. Similarly, H4 predicts that the auditor who faces uncertainty about ρ increases β', relative to the auditor who knows the level of ρ.

H4. An auditor's uncertainty about the level of ρ induces the auditor to increase β'.

H5 and H6 assume that the auditor also perceives her uncertainty about ρ as a factor that affects the taxpayer's choice, because of an information asymmetry. When the auditor faces uncertainty about ρ, the taxpayer has an information advantage in that the taxpayer can adjust his strategy based on the level of ρ, whereas the auditor cannot. Accordingly, the auditor's uncertainty about ρ affects $E_a(\alpha)$ and β''. These effects depend on the auditor's belief about the taxpayer's belief about the effect of the auditor's uncertainty on the audit rate. A possible taxpayer belief is that the auditor facing uncertainty always sets the audit rate as if $\rho = 0.50$, i.e., the mean of the uniform distribution for ρ. Given this belief, the taxpayer exploits the information advantage when ρ is 0.75 by increasing the rate of under-reporting, relative to the case where the auditor knows ρ. This consideration implies a positive association between the auditor's uncertainty about ρ and $E_a(\alpha)$. Another possible taxpayer belief is that the auditor facing uncertainty compensates by increasing the audit rate, consistent with H4. This consideration implies a negative association between the auditor's uncertainty about ρ and $E_a(\alpha)$. Because the net effect of these considerations could be either positive or negative, H5 and H6 are non-directional. H5 predicts that the auditor's uncertainty about ρ induces the auditor to revise $E_a(\alpha)$, and H6 predicts that the auditor's uncertainty about ρ induces the auditor to revise β''.

H5. An auditor's uncertainty about the level of ρ induces the auditor to revise $E_a(\alpha)$.

H6. An auditor's uncertainty about the level of ρ induces the auditor to revise β''.

4. METHOD

The experiment operationalized the tax game described earlier, using 80 business, science, and technology students at a major university. The data analyzed in the next section are from four experimental sessions, each with 12 strategic taxpayers and 8 auditors. Throughout the experiment, the instructions referred to strategic taxpayers as reporters and auditors as verifiers. Each session consisted of 20 rounds. The design included a within-subjects variable, i.e., round-by-round variation in ρ among three levels (0.25, 0.50, 0.75), and a between-subjects variable, i.e., the auditors' uncertainty about ρ (present, absent). In two sessions, the auditors knew the level of ρ when stating the audit rate. In the other two sessions, the auditors knew that the three levels of ρ were equally likely, but did not know the realized level of ρ until the end of the round. The procedure randomly and anonymously re-paired the auditors and taxpayers round-by-round, in order to minimize reputation formation.

We conducted the experiment with a computer network.[8] At the start of the experiment, the procedure randomly assigned subjects to the role of auditor or strategic taxpayer. The instructions told all subjects about the incentives, information, and task for each role. Each subject received US$15 for completing the experiment plus the outcome of a lottery with a prize of US$15. Each subject's chance of winning the prize was a linear function of the amount of francs earned in the 20 rounds. A different function was used for auditors and strategic taxpayers, such that expected pay was the same for each role, based on the optimal strategies stated earlier. For any subject who maximized the expected chance of winning the prize, the incentives were consistent with the assumptions of Graetz et al. (1986).

The procedure consisted of eight steps in each round. The subjects provided their responses in steps 3 and 4. The computer executed all remaining steps. In the first two rounds, the subjects had a maximum of four minutes to enter their responses. In subsequent rounds, the subjects had a maximum of 1.5 minutes to enter their responses.

Step 1. The computer determined the proportion of strategic taxpayers, ρ, for the round.[9] To vary the level of ρ without changing the number of active subjects, the procedure used *automatic reports* as a proxy for ethical taxpayers. The instructions stated that there were a number of automatic reports in addition to the reports from the 12 strategic taxpayers. The number of automatic reports varied among three equally likely values (4, 12, 36). Reported income always matched actual income in an automatic report. In contrast, reported income in a report from a strategic taxpayer could be lower than actual income, depending on the taxpayer's decision. The instructions included a schedule detailing the three possibilities (Table 1). In the no uncertainty condition, all subjects knew that all subjects received this information in Step 1. In the uncertainty condition, all subjects knew that only the taxpayers received this information in Step 1.

Step 2. The computer randomly pre-assigned all reports to the eight auditors, without disclosing the assignment to any subject. The computer assigned 6 reports to each auditor when the number of automatic reports was 36, 3 reports to each auditor when the number of automatic reports was 12, and 2 reports to each auditor when the number of automatic reports was 4. The instructions stated that the auditors would

Table 1. Schedule for Proportion of Strategic Taxpayers

	Number of reports from strategic taxpayers	*Number of automatic reports*	*Proportion of reports from strategic taxpayers*	*Chance of occurrence*
(1)	12	36	25%	1/3
(2)	12	12	50%	1/3
(3)	12	4	75%	1/3

not know whether the reports assigned to them were automatic reports or reports from the strategic taxpayers, without conducting a costly audit for each report, and that the taxpayers would not know which of the auditors had their report.

Step 3. The auditors provided responses to two questions. One question asked the auditors to assume that a taxpayer's actual income was 100 francs and to estimate the probability that the taxpayer would report income of 0 franc. This response measured $E_a(\alpha)$. The other question asked the auditors to assume that reported income was 0 franc and to state the probability of conducting an audit for this report. The latter response measured β.

Step 4. The taxpayers provided responses to two questions. One question asked the taxpayers to assume that an auditor received a report of 0 franc and to estimate the probability that the auditor would audit the report. This response measured $E_t(\beta)$. The other question asked the taxpayers to assume that their actual income was 100 francs and to state the probability that they would report income of 0 franc. The latter response measured α. Steps 3 and 4 took place concurrently.

Step 5. The computer determined whether actual income was 0 or 100 francs, separately for each taxpayer and automatic report. There was a 50% chance for each level of income.

Step 6. The computer determined the amount of reported income, separately for each taxpayer. If a taxpayer's actual income was 0 franc, the computer always produced a report of 0 franc. If a taxpayer's income was 100 francs, the computer applied the taxpayer's α response from step 4. Suppose that a taxpayer's α response was 80%. The computer implemented a chance device with a probability of 0.80 of reporting income of 0 franc and a probability of 0.20 of reporting income of 100 francs.

Step 7. The computer determined whether the auditors conducted an audit, separately for each report assigned to them. If reported income was 100 francs, the computer always produced the decision not to audit. If reported income was 0 franc, the computer applied the auditor's β response from step 3. Suppose that an auditor's β response was 25%. For each report assigned to the auditor, the computer implemented a chance device with a probability of 0.25 of conducting an audit and a probability of 0.75 of no audit.

Step 8. The computer tallied the results for the period and provided feedback to the subjects. The instructions included a schedule detailing each player's payoff under all possible scenarios (Table 2).

Each subject's computer screen displayed an information window, a history window, and a message window. The information window showed the conditions that were in effect for the round, e.g., the audit cost was 10 francs. For the auditors

Table 2. Schedule for Players' Payoffs

	Strategic taxpayer's actual income	*Strategic taxpayer's reported income*	*Does auditor audit?*	*Strategic taxpayer's payoff*	*Auditor's payoff*
(1)	0	0	no	0	0
(2)	0	0	yes	0	–10
(3)	100	100	no	50	50
(4)	100	0	no	100	0
(5)	100	0	yes	0	90

in the uncertainty condition, however, the information window did not show the proportion of strategic taxpayers. The history window showed the results for prior rounds. For each auditor, the history window showed each round's number of automatic reports, number of assigned reports, the auditor's responses, number of audited reports, total audit cost in francs, total payoff in francs, current payoff in francs per report, and cumulative average payoff in francs through the current round. For each taxpayer, the history window showed each round's number of automatic reports, the taxpayer's responses, actual and reported income, whether an audit occurred, current payoff in francs, and cumulative average payoff in francs through the current round. The message window included the two questions described earlier (step 3 for auditors, or step 4 for taxpayers) and spaces for providing responses.

5. RESULTS

A statistical problem with using multiple replications of a one-period game is serial dependence. We took several steps to reduce this problem. The experimental procedure randomly re-paired auditors and taxpayers each round, and used two random sequences of ρ over the 20 rounds. The data analysis computed each subject's mean response over multiple rounds with the same level of ρ. Using repeated-measures analysis of variance, hypothesis testing involved only three values of each auditor's responses for β and $E_a(\alpha)$. Finally, we performed additional analyses to determine whether the results were similar in the earlier rounds (i.e., the first three observations at each level of ρ) and later rounds (i.e., the last three observations at each level of ρ).

Panel A of Table 3 presents descriptive statistics for the audit rate and rate of under-reporting income. The mean audit rate was 0.48. For the no uncertainty condition, the audit rate increased from 0.34 to 0.45 to 0.51, as the level of ρ increased from 0.25 to 0.50 to 0.75. The audit rate was higher when the auditors faced uncertainty about ρ (0.53) than when the auditors knew the level of ρ (0.43). The mean

Table 3. Strategy Choice and Estimate of Opponents' Strategy – Mean (Standard Deviation)

	Auditors			*Strategic Taxpayers*		
	Uncertainty about ρ			*Uncertainty about ρ*		
	Absent	*Present*	*Over groups*	*Absent*	*Present*	*Over groups*
[A] Strategy Choice.						
$\rho = 0.25$	0.34 (0.20)	0.54 (0.26)	0.44 (0.25)	0.67 (0.24)	0.62 (0.25)	0.64 (0.25)
$\rho = 0.50$	0.45 (0.22)	0.50 (0.21)	0.47 (0.21)	0.46 (0.25)	0.50 (0.23)	0.48 (0.24)
$\rho = 0.75$	0.51 (0.21)	0.53 (0.23)	0.52 (0.22)	0.33 (0.27)	0.45 (0.24)	0.39 (0.26)
Over ρ	0.43 (0.17)	0.53 (0.23)	0.48 (0.21)	0.49 (0.20)	0.52 (0.20)	0.51 (0.20)
[B] Estimate of Opponents' Strategy.						
$\rho = 0.25$	0.63 (0.19)	0.57 (0.17)	0.60 (0.18)	0.34 (0.17)	0.41 (0.23)	0.37 (0.20)
$\rho = 0.50$	0.66 (0.11)	0.56 (0.14)	0.61 (0.13)	0.54 (0.12)	0.44 (0.21)	0.49 (0.18)
$\rho = 0.75$	0.74 (0.09)	0.58 (0.15)	0.66 (0.15)	0.68 (0.17)	0.49 (0.20)	0.59 (0.21)
Over ρ	0.68 (0.10)	0.57 (0.15)	0.63 (0.14)	0.52 (0.10)	0.45 (0.19)	0.48 (0.16)

Panel A shows means, with standard deviations in parentheses, for the auditors' audit rate and strategic taxpayers' rate of under-reporting income. Panel B shows means, and standard deviations in parentheses, for the auditors' estimate of the rate of under-reporting income and the strategic taxpayers' estimate of the audit rate. There were three levels for the proportion of strategic taxpayers, ρ (i.e., 0.25, 0.50, and 0.75), and two levels for the auditors' uncertainty about ρ (i.e., present and absent).

rate of under-reporting income was 0.51. The taxpayers' rate of under-reporting income was slightly higher when the auditors faced uncertainty about ρ (0.52) than when the auditors knew the level of ρ (0.49). For the no uncertainty condition, the rate of under-reporting decreased from 0.67 to 0.46 to 0.33, as the level of ρ increased from 0.25 to 0.50 to 0.75. For the uncertainty condition, the rate of under-reporting decreased from 0.62 to 0.50 to 0.45, as the level of ρ increased from 0.25 to 0.50 to 0.75.

Panel B of Table 3 presents descriptive statistics for the subjects' estimates of their opponent's strategy. The auditors' mean estimate of the rate of under-reporting was 0.63. For the no uncertainty condition, the auditors' estimate increased from 0.63 to 0.66 to 0.74, as the level of ρ increased from 0.25 to 0.50 to 0.75. The auditors' estimate was lower when the auditors faced uncertainty about ρ (0.57) than when the auditors knew the level of ρ (0.68). The taxpayers' mean estimate of the audit rate was 0.48. The taxpayers' estimate was lower when the auditors faced uncertainty about ρ (0.45) than when the auditors knew the level of ρ (0.52). For the no uncertainty condition, the taxpayers' estimate increased from 0.34 to 0.54 to 0.68, as the level of ρ increased from 0.25 to 0.50 to 0.75.

As a preliminary to hypothesis testing, we performed a regression for each auditor over 20 rounds:

$$\beta = A + B \cdot E_a(\alpha) + \varepsilon. \tag{9}$$

To measure β' in the tests of H1 and H4, we computed $\{A + \varepsilon\}$ for each round, and then computed the mean value for each level of ρ. To measure $E_a(\alpha)$ in the tests of H2 and H5, we used the mean estimate of the rate of under-reporting for each level of ρ. To measure β'' in the tests of H3 and H6, we computed $\{B \cdot E_a(\alpha)\}$ for each round, and then computed the mean value for each level of ρ. Although not part of hypothesis testing, we used the same approach to compute α' and α'' for the taxpayers.

Panel A of Table 4 shows the results from a series of repeated-measures analyses of variance. In each analysis, the within-subjects factor was the level of ρ (i.e., 0.25, 0.50, or 0.75), and the between-subjects factor was the auditors' uncertainty about ρ (i.e., present or absent). The dependent variables were β', $E_a(\alpha)$, β'', and β. Panel B of Table 4 shows the results from repeated-measures analyses of variance with the same dependent variables, but with a reduced data set that included the no uncertainty group. Panel C of Table 4 shows the results from one-way analyses of variance with the same dependent variables, but with a reduced data set that included cases with ρ of 0.50. The entries are F statistics. The F statistics without brackets are based on observations from all 20 rounds. The F statistics within brackets are based on observations from the earlier and later rounds, respectively.

Effects of ρ. H1 predicts that, when the auditor knows the level of ρ, there is a positive association between the level of ρ and β'. Focusing on the no uncertainty group (Panel B of Table 4), there was a significantly positive association between the level of ρ and β' ($F = 5.04$, $p = 0.013$). The results were similar for the earlier and later rounds. Taking all 20 rounds into account, the auditors in the no uncertainty

Table 4. Effects of Proportion of Strategic Taxpayers and Uncertainty on Auditors' Responses

	Independent Variable	Dependent Variable			
		β'	$E_a(\alpha)$	β''	β
[A]	ρ	1.96 {5.52**, 1.08}	4.31* {4.14*, 3.59*}	2.25 {0.94, 1.29}	3.74* {4.91**, 1.36}
	Uncertainty	9.71** {8.04**, 6.76**}	5.75* {2.45, 8.48**}	5.60* {4.55*, 5.05*}	1.82 {2.87, 0.60}
	ρ × Uncertainty	4.91** {2.19, 5.75**}	3.04 {2.10, 2.49}	1.86 {1.73, 0.87}	6.06** {3.81*, 5.42**}
[B]	ρ	5.04** {4.89**, 5.97**}	4.61* {3.95*, 3.77*}	2.17 {0.98, 1.22}	6.08** {5.69**, 4.81*}
[C]	Uncertainty	7.73** {5.77*, 6.64**}	5.00* {0.87, 8.44**}	5.97* {4.89*, 5.27*}	0.64 {0.37, 0.70}

*Significant at $p = 0.05$, two-tailed.
**Significant at $p = 0.01$, two-tailed.

Panel A shows the results from repeated-measures analyses of variance with a within-subjects factor (i.e., the proportion of strategic taxpayers was 0.25, 0.50, or 0.75), and a between-subjects factor (i.e., auditors' uncertainty about ρ was present or absent). Panel B shows the results from repeated-measures analyses of variance for the no uncertainty group only. Panel C shows the results from one-way analyses of variance for cases with ρ of 0.50. In all panels, the dependent variables are β' in the first column, $E_a(\alpha)$ in the second column, β'' in the third column, and β in the last column. The entries are F statistics including observations for all 20 rounds and, in brackets, for the first three rounds at each level of ρ and for the last three rounds at each level of ρ.

group increased β' from 0.08 to 0.15 to 0.18, as the level of ρ increased from 0.25 to 0.50 to 0.75.[10] These results support H1.

H2 predicts that, when the auditor knows the level of ρ, changes in the level of ρ induce the auditor to revise $E_a(\alpha)$. Focusing on the no uncertainty group (Panel B of Table 4), there was a significantly positive association between the level of ρ and $E_a(\alpha)$ ($F = 4.61$, $p = 0.018$). The results were similar for the earlier and later rounds. Taking all 20 rounds into account, the auditors in the no uncertainty group increased $E_a(\alpha)$ from 0.63 to 0.66 to 0.74, as the level of ρ increased from 0.25 to 0.50 to 0.75.[11] These results support H2.

H3 predicts that, when the auditor knows the level of ρ, changes in the level of ρ induce the auditor to revise β''. Focusing on the no uncertainty group (Panel B of Table 4), there was an insignificant association between the level of ρ and β'' ($F = 2.17$, $p = 0.132$). The results were similar for the earlier and later rounds.

Taking all 20 rounds into account, the auditors in the no uncertainty group increased β'' from only 0.26 to 0.29 to 0.33, as the level of ρ increased from 0.25 to 0.50 to 0.75.[12] Although the auditors in the no uncertainty group revised $E_a(\alpha)$ in response to changes in the level of ρ, consistent with H2, such belief revision did not lead to a significant change in β'', contrary to H3.

Effects of Auditors' Uncertainty about ρ. H4 predicts that an auditor's uncertainty about ρ induces the auditor to increase β'. Table 4 includes two tests of H4. Based on the entire sample, the first column in Panel A of Table 4 shows a significant effect for uncertainty ($F = 9.71$, $p = 0.004$). Based on the subsample with ρ of 0.50, the first column in Panel C of Table 4 also shows a significant effect for uncertainty ($F = 7.73$, $p = 0.009$). The results were similar for the earlier and later rounds. An advantage of the latter test is its direct comparison between auditors who faced uncertainty, i.e., a uniform distribution for ρ with a mean of 0.50, and auditors who knew that the level of ρ was 0.50. Both tests indicate that the auditors' uncertainty about ρ affected β'. For the entire sample, β' was 0.56 when the auditors faced uncertainty and 0.13 when they did not. For the subsample with ρ of 0.50, β' was 0.55 when the auditors faced uncertainty and 0.15 when they did not. These results support H4.

H5 predicts that an auditor's uncertainty about ρ induces the auditor to revise $E_a(\alpha)$. Based on the entire sample, the second column in Panel A of Table 4 shows a significant effect for uncertainty ($F = 5.75$, $p = 0.023$). Based on the subsample with ρ of 0.50, the second column in Panel C of Table 4 also shows a significant effect for uncertainty ($F = 5.00$, $p = 0.033$). These effects were stronger in the later rounds than in the earlier rounds. Both tests indicate that uncertainty about ρ induced the auditors to revise $E_a(\alpha)$. For the entire sample, $E_a(\alpha)$ was 0.57 when the auditors faced uncertainty and 0.68 when they did not. For the subsample with ρ of 0.50, $E_a(\alpha)$ was 0.56 when the auditors faced uncertainty and 0.66 when they did not. These results support H5.

H6 predicts that the auditor's uncertainty about ρ induces the auditor to revise β''. Based on the entire sample, the third column in Panel A of Table 4 shows a significant effect for uncertainty ($F = 5.60$, $p = 0.025$). Based on the subsample with ρ of 0.50, the third column of Panel C in Table 4 also shows a significant effect for uncertainty ($F = 5.97$, $p = 0.021$). The results were similar in the earlier and later rounds. Both tests indicate that uncertainty about ρ induced the auditors to revise β''. For the entire sample, β'' was -0.04 when the auditors faced uncertainty and 0.30 when they did not. For the subsample with ρ of 0.50, β'' was -0.05 when the auditors faced uncertainty and 0.29 when they did not. These results support H6.

The last column of Table 4 shows the analyses that used β as the dependent variable. Panel A shows a significant effect for the level of ρ ($F = 3.74$, $p = 0.03$), a significant interaction effect for the level of ρ and auditors' uncertainty about ρ ($F = 6.06$, $p = 0.004$), but an insignificant effect for auditors' uncertainty about ρ ($F = 1.82$, $p = 0.187$). For the no uncertainty group, Panel B shows a significant effect for the level of ρ ($F = 6.08$, $p = 0.006$). For the subsample with ρ of 0.50, Panel C shows an insignificant effect for uncertainty ($F = 0.64$, $p = 0.431$).

Taken together, the above tests indicate that the auditors responded to changes in the level of ρ, and to their uncertainty about ρ. Increases in the level of ρ, and in their uncertainty about ρ, induced a significant increase in β'. A higher level of ρ meant that there were more potential cheaters in the taxpayer population, justifying a higher audit rate. Also, the auditors increased β' to compensate for their uncertainty about ρ (cf. Zimbelman and Waller, 1999). Increases in the level of ρ induced a modest increase in β''. Although variation in the level of ρ induced the auditors to revise $E_a(\alpha)$, the impact was not strong enough to cause a significant change in β''. The auditors' uncertainty about ρ induced significant decreases in $E_a(\alpha)$ and β''. The decrease in $E_a(\alpha)$ indicated the auditors' belief that the taxpayers believed that the auditors compensated for uncertainty about ρ by increasing the audit rate. The auditors' uncertainty about ρ had countervailing effects on β' and β'', such that uncertainty had a modest effect on β.

6. STRATEGIC TAXPAYERS

Table 5 shows the results from a series of repeated-measures analyses of variance. In each analysis, the within-subjects factor was the level of ρ (i.e., 0.25, 0.50, or 0.75), and the between-subjects factor was the auditors' uncertainty about ρ (i.e., present or absent). The dependent variables were α', $E_t(\beta)$, α'', and α. The taxpayers significantly decreased their rate of under-reporting as the level of ρ increased ($F = 25.67$, $p = 0.001$, last column of Table 5). This effect depended on revisions in $E_t(\beta)$ and α''. Increases in the level of ρ induced the taxpayers to revise upward their estimates

Table 5. Effects of Proportion of Strategic Taxpayers and Uncertainty on Taxpayers' Response

Independent Variable	*Dependent Variable*			
	α'	$E_t(\beta)$	α''	α
ρ	4.88** {3.41*, 1.51}	31.15** {24.39**, 23.68**}	23.59** {17.14**, 16.36**}	25.67** {18.75**, 13.00**}
Uncertainty	0.58 {0.38, 0.83}	2.61 {1.69, 2.34}	1.45 {1.13, 1.53}	0.38 {0.66, 0.08}
$\rho \times$ Uncertainty	2.26 {0.84, 1.49}	13.00** {7.38**, 10.52**}	11.34** {5.32**, 11.88**}	2.83 {1.30, 2.20}

*Significant at $p = 0.05$, two-tailed.
**Significant at $p = 0.01$, two-tailed.

The table shows the results from repeated-measures analyses of variance with a within-subjects factor (i.e., the proportion of strategic taxpayers was 0.25, 0.50, or 0.75), and a between-subjects factor (i.e., auditors' uncertainty about ρ was present or absent). The dependent variables are α' in the first column, $E_t(\beta)$ in the second column, α'' in the third column, and α in the last column. The entries are F statistics including observations for all 20 rounds and, in brackets, for the first three rounds at each level of ρ and for the last three rounds at each level of ρ.

of the audit rate ($F = 31.15$, $p = 0.001$, second column of Table 5) and revise downward their rate of under-reporting ($F = 23.59$, $p = 0.001$, third column of Table 5). The significant interaction effects in the second and third columns of Table 5 indicate that the level of ρ had larger effects on $E_t(\beta)$ and α'' when the auditors knew the level of ρ. These results were similar for the earlier and later rounds. To a lesser extent, the level of ρ also affected α' ($F = 4.88$, $p = 0.01$, first column of Table 5). In sum, the taxpayers' beliefs were sensitive to the auditors' use of ρ as a cue for adjusting the audit rate, and the taxpayers responded by adjusting the rate of under-reporting.

7. ERRORS IN STRATEGY CHOICE

The optimal strategies from Graetz et al. (1986) provided a benchmark for evaluating the subjects' strategy choice. Panel A of Table 6 shows the mean signed error, i.e., the difference between the subjects' strategy choice and the optimal strategies stated earlier, with the standard deviation in parentheses. Panel B of Table 6 shows the mean absolute error with the standard deviation in parentheses. For the auditors, the largest mean errors occurred when the auditors knew that ρ was 0.25. The auditors decreased the audit rate with decreases in the level of ρ, and consequently under-audited when ρ was 0.25. The mean absolute error was about 0.15 to 0.20, regardless of the level of ρ or the auditors' uncertainty about ρ. For their part, the taxpayers cheated too much. The mean signed and absolute errors were about 0.20 to 0.30, regardless of the level of ρ or auditors' uncertainty about ρ.

8. ERRORS IN ESTIMATES OF OPPONENT'S STRATEGY

Panel A of Table 7 shows the mean signed error, i.e., the difference between the subjects' estimate of their opponent's strategy and their opponent's actual strategy, with the standard deviation in parentheses. Panel B of Table 7 shows the mean absolute error with the standard deviation in parentheses. For the auditors, the largest mean errors occurred when the auditors knew that ρ was 0.75. The auditors incorrectly believed that the taxpayers increased α as the level of ρ increased. On the contrary, the taxpayers decreased α as the level of ρ increased. For their part, the taxpayers correctly believed that the audit rate increased with increases in the level of ρ.

9. CONCLUSION

This paper reported a behavioral accounting experiment on strategic interaction in a tax compliance game. The experiment employed a three-step approach. First, the experiment assigned subjects to the opposing roles of auditor and strategic taxpayer. This step addressed a past criticism of behavioral accounting research: economic mechanisms such as the interaction of players with conflicting preferences discipline and potentially eliminate the decision biases found in individual settings. Second, the experiment operationalized a game-theoretic model of the tax compliance

Table 6. Errors in Strategy Choice – Mean (Standard Deviation)

	Auditors			*Strategic Taxpayers*		
	Uncertainty about ρ			*Uncertainty about ρ*		
	Absent	*Present*	*Over groups*	*Absent*	*Present Present*	*Over groups*
[A] Signed Errors.						
$\rho = 0.25$	−0.16 (0.20)	0.04 (0.26)	−0.06 (0.25)	0.23 (0.24)	0.18 (0.25)	0.20 (0.25)
$\rho = 0.50$	−0.05 (0.21)	0.01 (0.22)	−0.02) (0.21)	0.24 (0.25)	0.28 (0.23)	0.26 (0.24)
$\rho = 0.75$	0.01 (0.21)	0.03 (0.23)	0.02 (0.22)	0.18 (0.27)	0.30 (0.24)	0.24 (0.26)
Over ρ	−0.07 (0.18)	0.03 (0.23)	−0.02 (0.21)	0.22 (0.20)	0.25 (0.20)	0.23 (0.20)
[B] Absolute Errors.						
$\rho = 0.25$	0.23 (0.12)	0.20 (0.16)	0.22 (0.14)	0.29 (0.15)	0.27 (0.15)	0.28 (0.15)
$\rho = 0.50$	0.18 (0.13)	0.17 (0.13)	0.17 (0.12)	0.31 (0.15)	0.30 (0.19)	0.31 (0.17)
$\rho = 0.75$	0.16 (0.13)	0.19 (0.13)	0.18 (0.13)	0.24 (0.22)	0.32 (0.22)	0.28 (0.22)
Over ρ	0.19 (0.12)	0.19 (0.13)	0.19 (0.12)	0.25 (0.15)	0.27 (0.17)	0.26 (0.16)

Panel A (B) shows the mean and standard deviation in parentheses for the signed (absolute) error in the auditors' audit rate and strategic taxpayers' rate of under-reporting income, compared to the optimal strategies based on Graetz et al. (1986). There were three levels for the proportion of strategic taxpayers, ρ (i.e., 0.25, 0.50, and 0.75), and two levels for the auditors' uncertainty about ρ (i.e., present and absent).

problem (Graetz et al., 1986). This step addressed another past criticism of behavioral accounting research: without a formal model of strategic interaction, it is problematic to define rational behavior in the experimental setting. Third, the experiment manipulated two variables that were normatively irrelevant in the game-theoretic

Table 7. Errors in Estimate of Opponents' Strategy – Mean (Standard Deviation)

	Auditors			*Strategic Taxpayers*		
	Uncertainty about ρ			*Uncertainty about ρ*		
	Absent	*Present*	*Over groups*	*Absent*	*Present*	*Over groups*
[A] Signed Errors.						
$\rho = 0.25$	−0.04) (0.19)	−0.05) (0.17)	−0.04) (0.17)	0.00 (0.18)	−0.13) (0.22)	−0.07) (0.21)
$\rho = 0.50$	0.20 (0.11)	0.06 (0.14)	0.13 (0.14)	0.10 (0.18)	−0.06 (0.20)	0.02 (0.20)
$\rho = 0.75$	0.41 (0.09)	0.13 (0.15)	0.27 (0.19)	0.17 (0.19)	−0.04 (0.21)	0.07 (0.23)
Over ρ	0.19 (0.23)	0.05 (0.17)	0.12 (0.21)	0.09 (0.14)	−0.08 (0.19)	0.01 (0.18)
[B] Absolute Errors.						
$\rho = 0.25$	0.15 (0.12)	0.12 (0.12)	0.14 (0.12)	0.15 (0.09)	0.22 (0.14)	0.18 (0.12)
$\rho = 0.50$	0.20 (0.10)	0.11 (0.10)	0.16 (0.11)	0.17 (0.10)	0.16 (0.13)	0.17 (0.11)
$\rho = 0.75$	0.41 (0.09)	0.17 (0.10)	0.29 (0.16)	0.21 (0.15)	0.17 (0.11)	0.19 (0.13)
Over ρ	0.25 (0.16)	0.14 (0.11)	0.20 (0.15)	0.18 (0.08)	0.19 (0.11)	0.18 (0.10)

Panel A (B) shows the mean and standard deviation in parentheses for the signed (absolute) error in the auditors' estimate of the rate of under-reporting income and the strategic taxpayers' estimate of the audit rate, compared to the actual mean strategies of the opponent. There were three levels for the proportion of strategic taxpayers, ρ (i.e., 0.25, 0.50, and 0.75), and two levels for the auditors' uncertainty about ρ (i.e., present and absent).

model, i.e., the level of ρ and uncertainty about ρ, to test hypotheses about auditors' choice of the audit rate, β.

The hypotheses assumed that the auditor with limited rationality makes a strategy choice through a simple additive process: $\beta = \beta' + \beta''$, where β' depends on ρ, among other factors, and β'' depends on the auditor's belief about the taxpayer's strategy, $E_a(\alpha)$. The results showed that the auditors used ρ as a cue for adjusting the audit rate. There were significantly positive associations between the level of ρ and β', and between the auditors' uncertainty about ρ and β'. Regarding $E_a(\alpha)$ and β'', the auditors formed incorrect beliefs about the taxpayers' responses. The auditors incorrectly believed that the taxpayers increased the rate of under-reporting as the level of ρ increased, although this belief had only a modest effect on β''. Also, the auditors incorrectly believed that the taxpayers expected a higher audit rate when the auditors faced uncertainty about ρ. For their part, the taxpayers were sensitive to the auditors' use of ρ as a cue for setting the audit rate. The taxpayers correctly believed that the audit rate increased as the level of ρ increased, and responded by decreasing the rate of under-reporting income.

A key element in the experiment was the players' formation of a belief about their opponent's strategy and incorporation of this belief into their own strategy choice. Future research might further examine belief formation and its impact on strategy choice, in two ways. One way is to elaborate the subjects' task. Our experiment collected the subjects' estimates of their opponent's strategy. A future experiment additionally might collect the subjects' estimates of their opponent's estimate of their own strategy. For example, the auditors might estimate $E_a(E_t(\beta))$ as well as $E_a(\alpha)$. The extra data would clarify how the subjects make belief-based adjustments in their strategy choice. The other way is to simplify the subjects' task. Our experiment required the taxpayers to estimate the audit rate and incorporate the estimate into their strategy choice. A future experiment might use a sequence that (1) requires the auditors state the audit rate, (2) informs the taxpayers about the audit rate, and (3) requires the taxpayers to state the rate of under-reporting income. This sequence simplifies the taxpayers' reasoning process by replacing $E_t(\beta)$ with β, and simplifies the auditors' reasoning process by eliminating the need to estimate $E_a(E_t(\beta))$. Consequences of this sequence might be less cheating by the taxpayers and more accurate estimates of α by the auditors, relative to the levels in our experiment. Such evidence could inform tax policy makers who decide whether to pre-announce the audit rate.

NOTES

[1] Other experiments on tax compliance include Alm, Jackson, and McKee (1992), Beck, Davis, and Jung (1991, 1992), Boylan and Sprinkle (2001), Collins and Plumlee (1991), Friedland, Maital, and Rutenberg (1978), Kim (2002), and Moser, Evans, and Kim (1995). These experiments examined the effects on taxpayer compliance of independent variables such as the tax rate, audit rate, penalty for under-reporting, and public good "payback" to the taxpayer. None of these experiments adopted the three-step approach of our experiment, none operationalized the Graetz et al. (1986) model, and none examined the effects of the proportion of strategic versus ethical taxpayers.

[2] Equation 1 is comparable to an anchoring-and-adjustment process, where β' is an anchor that is based on the auditor's payoffs and other parameters of the setting, before considering the taxpayer's strategy, and β'' is an adjustment that is based on the auditor's belief about the taxpayer's strategy.

[3] As elaborated below, the hypotheses for β'' are non-directional, because of countervailing considerations. The net effect depends on how the auditor weighs these considerations.

[4] When α^* is greater than one, it never pays to conduct an audit.

[5] To compute the threshold, enter all parameter values except ρ into Equation 3, set the equation equal to one, and solve for ρ.

[6] Our hypotheses focus on the auditor. However, we also collected and analyzed observations of the taxpayers' choice of α and belief about the audit rate, $E_t(\beta)$.

[7] There are many theoretical and operational definitions of ambiguity in the decision literature. Ellsberg (1961, p. 657) defined ambiguity as the "quality depending on the amount, type, reliability, and 'unanimity' of information, giving rise to one's degree of 'confidence' in an estimate of relative likelihoods." Camerer (1995, p. 645) more simply defined ambiguity as "not knowing relevant information that could be known." Our experiment manipulated whether the auditors knew the level of ρ or not. In the latter case, the auditors knew the uniform distribution for ρ. This manipulation is consistent with the above definitions, but not with other operational definitions in the literature that left unstated the distribution for the probability parameter. To avoid ambiguity on this point, we refer the auditor's uncertainty about ρ, rather than ambiguity about ρ.

[8] Contact the first author (acckim@ewha.ac.kr) for a copy of the instructions.

[9] When designing the experiment, we pre-determined two random sequences with 20 values of ρ drawn from a uniform distribution over 0.25, 0.50, and 0.75. We used each random sequence in two of the four sessions. When implementing step 1, the computer used one of the pre-determined random sequences.

[10] Panel A of Table 4 shows that, taking all 20 rounds into account, there was a significant interaction between ρ and uncertainty ($F = 4.91, p = 0.011$). The interaction effect was stronger in the later rounds. The auditors used ρ as a cue for adjusting β' when they knew ρ.

[11] Panel A of Table 4 shows that, taking all 20 rounds into account, there was a significant main effect for ρ ($F = 4.31$, $p = 0.018$), and a marginally significant interaction between ρ and uncertainty ($F = 3.04$, $p = 0.055$). The level of ρ affected $E_a(\alpha)$.

[12] Panel A of Table 4 shows that, taking all 20 rounds into account, there was an insignificant main effect for ρ ($F = 2.25$, $p = 0.114$) and an insignificant interaction between ρ and uncertainty ($F = 1.86$, $p = 0.164$).

REFERENCES

Alm, J., B. Jackson, and M. McKee. (March 1992). "Estimating the Determinants of Taxpayer Compliance with Experimental Data." *National Tax Journal*, 107–114.

Andreoni, J., B. Erard, and J. Feinstein. (June 1998) . "Tax Compliance." *Journal of Economic Literature*, 818–860.

Beck, P., J. Davis, and W. Jung. (July 1991). "Experimental Evidence on Taxpayer Reporting Under Uncertainty." *The Accounting Review*, 535–58.

——, ——, and ——. (Fall 1992). "Experimental Evidence on an Economic Model of Taxpayer Aggression under Strategic and Nonstrategic Audits." *Contemprary Accounting Research*, 86–112.

Boylan, S. and G. Sprinkle. (Spring 2001). "Experimental Evidence on the Relation between Tax Rates and Compliance: The Effect of Earned vs. Endowed Income." *Journal of the American Taxation Association*, 75–90.

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton University Press: Princeton, NJ).

——. Individual Decision Making. (1995). In J. Kagel and A. Roth (eds.), *Handbook of Experimental Economics* (Princeton, NJ: Princeton University Press).

—— and M. Weber. (December 1992). "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity." *Journal of Risk and Uncertainty*, 325–370.

Collins, J. and D. Plumlee. (July 1991). "The Taxpayer's Labor and Reporting Decision: The Effect of Audit Schemes." *The Accounting Review*, 559–76.

Connolly, T., H. Arkes, and K. (2000). "Hammond." *Judgment and Decision Making: An Interdisciplinary Reader* (Cambridge University Press: Cambridge, UK).

Cuccia, A. (January 1994). "The Economics of Tax Compliance: What Do We Know and Where Do We Go?" *Journal of Accounting Literature*, 81–116.

Ellsberg, D. (November 1961). "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics*, 643–669.

Elster, J. (1989). *Nuts and Bolts for the Social Sciences* (Cambridge University Press: Cambridge, UK).

Friedland, N., S. Maital, and A. Rutenberg. (August 1978). A Simulation Study of Tax Evasion. *Journal of Public Economics*, 107–16.

Frisch, D. and J. Baron. (September 1988). "Ambiguity and Rationality." *Journal of Behavioral Decision Making*, 149–157.

Graetz, M., J. Reinganum, and L. Wilde. (Spring 1986). "The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement." *Journal of Law, Economics, and Organization*, 1–32.

Kahneman, D. and A. Tversky (eds.) (2000). *Choices, Values, and Frames* (Cambridge University Press: Cambridge, UK).

Kim, C. (2002). "Does Fairness Matter in Tax Reporting Behavior?" *Journal of Economic Psychology* , 771–785.

Moser, D., J. Evans III, and C. Kim. (October 1995). "The Effects of Horizontal and Exchange Inequity on Tax Reporting Decisions." *The Accounting Review*, 619–34.

Simon, H. (1982). *Models of Bounded Rationality: Behavioral Economics and Business Organization* (MIT Press: Cambridge, MA).

Smith, V. (2000). *Bargaining and Market Behavior* (Cambridge University Press: Cambridge, UK).

Waller, W. (2002). "Behavioral Accounting Experiments in Market and Game Settings." In R. Zwick and A. Rapoport (eds.), *Experimental Business Research* (Kluwer: Boston).

Zimbelman, M. and W. Waller. (Supplement 1999). "An Experimental Investigation of Auditor-Auditee Interaction under Ambiguity." *Journal of Accounting Research*, 135–155.

Chapter 3

INFORMATION DISTRIBUTION AND ATTITUDES TOWARD RISK IN AN EXPERIMENTAL MARKET OF RISKY ASSETS

David Bodoff
Hong Kong University of Science and Technology

Hugo Levevq
Reims Management School

Hongtao Zhang
Hong Kong University of Science and Technology

Abstract

Policies such as the SEC's Fair Disclosure Rule, and technologies such as SEC EDGAR, aim to disseminate corporate disclosures to a wider audience. In this study, we adopt an experimental approach to measure whether this wider disclosure is beneficial to investors. Analytical predictions based on theories of non-revealing and full-revealing prices can be summarized as "None > All > Half". A laboratory study was conducted to test these predictions. Subjects' preference for the fraction of informed traders can be summarized as "Half > None > All", i.e., investors most favor a situation where a random half of investors are informed. We explore two possible explanations for the contradiction of the theoretical predictions, either that the analytical models fail to predict market behavior, or that they succeed in predicting market behavior but nevertheless fail to predict subjects' preferences for the different sets of risk faced in each market. We ultimately adopt the second approach, and propose that subjects have different attitudes toward different sources of risk, a phenomenon which traditional analytical models do not consider.

1. INTRODUCTION

The Security and Exchange Commission's EDGAR on the Internet ("EOI") system delivers corporate disclosure documents to anyone with access to the Internet. This free

This research was supported by the Centre for Experimental Business Research, grant HKUST-3 Experiential based teaching for networked economy, and the Research Grants Council of Hong Kong, projects no. HKUST6056/01H and HKUST6213/04H.

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 57–79.
© 2005 *Springer. Printed in the Netherlands.*

system, which offers day-delayed information, also spawned an industry in disclosure document re-sellers such as EDGAR-online.com, which offers non-delayed access to the EDGAR data on a subscription basis. Prior to EOI, which is based on Internet technology, only a small number of institutional investors could afford to pay the high fees for the electronic EDGAR data feed, which was managed by Mead Data Central. Earlier still, the paper documents were only available to a handful of institutions whose employee sat in the SEC reading room to await publication of new disclosure documents. The historical progression has thus been from a very small number of investors with access to the information, to nearly universal access. Another initiative called the Fair Disclosure Rule (FD) shares a similar goal of leveling the playing field by making everyone informed, instead of only a lucky few.

While EOI has been a great success in achieving its goal of universal access, research in economics and finance points to unanswered questions about the effects of such a system on investors' welfare. If we model EOI as a shift in the fraction of informed traders, then a number of analytical models can be applied, which show sensitive and/or ambiguous effects on welfare. For example, the literature of partially revealing rational expectations indicates a variety of interacting effects from a change in the fraction of informed traders. These effects include signal risk, endowment risk, and other effects whose net result is ambiguous and highly sensitive to the details of the model. The analytical literature thus challenges any easy assumptions about the benefits of EOI, if we are to consider welfare, rather than merely access, as the measure of success.

But while the analytical literature offers insights and raises questions, it falls short of providing any real answers, in the absence of empirical evidence. We therefore turned to the experimental literature to help assess the benefit of EOI. We found that there is not a lot of experimental work on the effects of information on investors' welfare. The experiments we found, e.g., Bloomfield and O'hara (1999), measure the profits ("welfare") of insiders as compared with outsiders. Our focus, on the other hand, is not to compare the profits of insiders against those of outsiders, but to assess the effect of EOI, by comparing all traders' welfare before EOI against welfare after EOI. If λ denotes the fraction of informed traders, our question is, under which level of λ is welfare greatest?

When assessing whether welfare improves as a result of EOI, we must first clarify two points: Whose welfare are we measuring, and how do we define welfare? On the question of whose welfare we are measuring, we may be interested in insiders, or outsiders, or an average trader. In this paper, the identity of the informed traders is determined randomly in each trading round, so that all traders are identical ex ante. We thus focus on the welfare of an average trader. The rationale is that a trader is not normally fixed as a perennial outsider or an insider for all his/her trading life. Rather, a trader is a relative insider with respect to some stocks, and on some days, and a relative outsider with respect to other stocks and/or on other days. This choice allows us to focus on before-versus-after EOI, rather than on the difference between insiders and outsiders. Our question is, is an average trader better off in the pre-EOI world, in which he/she is an insider sometimes and an outsider at other times, or is he/she better off in a world where every trader is informed every time, as is the goal of EOI?

Regarding the second question, the definition of welfare, in an exchange economy it is pointless to define welfare in terms of expected profit of an average trader, because the experimental treatments do not affect the expected profit of an average trader. Rather, the policy choice of $\lambda = \{0, \frac{1}{2}, 1\}$ affects an average investor (only) though the set of risks he/she faces. We therefore want to study the ex ante expected *utility* (EU) – as opposed to profit – of an average trader, under the various levels of λ. A utility-maximizing trader ought to prefer the condition with the higher ex ante EU. We directly test this (rather than trying to elicit a numeric EU or certainty equivalent) using a self-reported pairwise preference, with proper economic incentives to motivate honest responses.

To summarize: The treatment is the fraction λ of informed traders, all traders are identical, and we study which level of λ our subjects prefer.

To make predictions, we solve our experimental market using market-clearing prices for both Non-Revealing (NR) and Fully-Revealing prices. Both approaches predict that an average risk-averse investor would prefer that no investors are informed rather than all are informed; and NR theory further predicts that an average investor would prefer a situation in which all investors are informed rather than half the investors are informed. These predictions can be summarized as "**None > All > Half**", i.e., $\lambda = 0$ is preferred to $\lambda = 1$, etc. This prediction holds for both partial and complete information.

Our results are as follows: Subjects' trading behavior and equilibrium prices generally conform to predictions and to previously reported results. But subjects' preference for the fraction of informed traders with partial information can be summarized as "**Half >** {indifferent between **None, All**}. With complete information, preferences can be summarized as "**None > Half > All**".

The main result from the point of view of public policy is that subjects always preferred "Half" to "All". This is the most relevant to policy, because the status quo has some non-zero fraction of informed traders, and policy initiatives such as FD and EOI aim to push the fraction closer to $\lambda = 1$. It turns out that, at least in our experimental conditions, an average trader prefers the status quo of $\lambda = \frac{1}{2}$. This result adds an empirical component to the ongoing debate regarding the motivation for policies that aim to level the playing field by making everyone an informed trader. In Bodoff, Levecq et al. (2003) we analyze all six of subjects' pairwise preferences – three levels of λ, with perfect or partial information – which were the main target of our experiment.

In this paper, we focus on one phenomenon that emerged unexpectedly in our results, and which we think deserves additional attention. We found that during trading, subjects acted in the ways we'd expect from risk-averse traders in a market for risky assets, but that they nevertheless preferred a treatment condition that is more risky over one that is less risky. Specifically, trading patterns in all treatment conditions conformed well to risk-averse behavior, lying somewhere in between the patterns predicted by fully revealing and non-revealing expectations for risk-averse traders. But ultimately, when asked whether they prefer to participate in an additional round with $\lambda = {}^1\!/\!_2$ or $\lambda = 1$, subjects often preferred $\lambda = {}^1\!/\!_2$, even though it is the riskier choice (we will discuss this more carefully below). In the remainder, we will primarily focus on this choice between $\lambda = {}^1\!/\!_2$ and $\lambda = 1$.

Our data thus suggests that for deciding some questions, our subjects' preferences are determined by something other than a traditional calculation of ex ante expected utility. We conjecture that subjects' preferences depend as well on the *source*, or *kind* of risk. Methodological questions arise when it comes to measuring the amount of risk under different treatments, and characterizing *kinds* or *sources* of risk that may be important in determining preferences. Further work will be required to target and validate this phenomenon in a "clean" experiment that is specifically designed for this purpose. But we could not find any other explanation for our subjects' preferences. We therefore explore this phenomenon in the context of our market trading data, and we deal with the methodological questions as they arise.

2. LITERATURE REVIEW

It is helpful to distinguish the relevant literature for our purposes, from much of the literature on insider trading. Research on insider trading usually considers a small fraction of insiders, whose presence we may wish to ban. That is, we consider a move from λ slightly > 0, to $\lambda = 0$. In our case, the situation is essentially reversed. The information is *supposed* to be public. We are considering a change to $\lambda = 1$.

In analytical models, asymmetric information, as is the case with $\lambda = {}^1\!/\!_2$, has a variety of effects on ex ante EU. On the one hand, differences in traders' beliefs may lead to additional trading opportunities (Indjejikian 1991), when two potential trading partners both benefit from the trade. On the other hand, such trades from asymmetric information may result in traders taking extreme positions rather than sharing risk equally (Diamond 1985), and this is harmful for ex ante EU of an average trader assuming traders have identical utility functions, since ex ante EU for identical traders is maximized with equal holdings. Finally, as explained by Lev and Ohlson (1982), Diamond (1985), Leland (1992) and others, outsiders may act cautiously to avoid being taken advantage of, thereby reducing liquidity which is bad for all traders' ex ante EU.

When asking whether ex ante EU is greater for $\lambda = {}^1\!/\!_2$ or $\lambda = 1$ we first need to specify whose EU we are talking about. If $0 < \lambda < 1$, then there are insiders and outsiders. Some models calculate separately the EU of insiders and of outsiders. Then, when comparing levels of λ, there are two separate questions: Do those who would be outsiders under $\lambda = {}^1\!/\!_2$ benefit from the switch to $\lambda = 1$, and separately, do those who would be insiders under $\lambda = {}^1\!/\!_2$ benefit from the switch to $\lambda = 1$? This is the approach taken in (Alles and Lundholm 1993) who find that even the would-be outsiders might not benefit from the switch to $\lambda = 1$. It is also the approach taken in Ausubel (1990) who find the opposite – that even would-be insiders might benefit from the switch to $\lambda = 1$, as a result of increased liquidity and other reduced risks.

As already mentioned, in this paper we instead measure the ex ante EU of an average trader, where each trader in the case of $\lambda = {}^1\!/\!_2$ stands an equal probability of being informed or uninformed in each round. The case of an average trader is similar to the analysis in Diamond (1985), except that the treatment in that paper involves release of a second information signal, rather than a widening of λ for the original information, as in our case. Our analysis in Bodoff and Zhang (2003) took Diamond's model and explored the effect of an exogenous change in λ for the original information signal. We found that a switch to $\lambda = 1$ may harm even would-be outsiders. In summary, the analytical literature shows ambiguous and sensitive effects from a policy of releasing information to all, rather than to some.

In reviewing the *experimental* literature, we found very few papers that consider the welfare effects of policies that affect the fraction of informed traders. In fact, we found no papers that attempt to experimentally measure ex ante EU or its surrogate. This is because most experimental studies focus either on theoretical price predictions, or else on comparing insiders' profits to those of outsiders. For these purposes, there is no need to study EU. But in our study of the supposed benefits of EOI, we were less interested in comparing informed traders to uninformed traders, and more interested in comparing (everyone) pre-EOI to post-EOI, which we model as $\lambda = 1/2$ and $\lambda = 1$, respectively. An average trader in our exchange economy faces the same expected profits under any treatment, so that profits is not an interesting statistic when comparing different levels of lambda. Our comparison of ex ante EU focuses instead on the different set of risks that a trader faces under $\lambda = 1/2$ and $\lambda = 1$, as reflected in ex ante EU which we abstract and elicit as a rank preference. We wanted to know, in a live market, with subjects' naturally occurring utility functions, does an average trader benefit from a policy that widens the fraction λ of informed traders? The main result of subjects' preferences for a policy of $\lambda = 1/2$ has been reported elsewhere. Our focus in this paper is on one apparent anomaly in results, which we now explain.

This shift in focus is depicted in figures 1-2. Analytical models assume a utility function and utility-maximizing traders, and from these assumptions calculate equilibrium prices and holdings. These are the predictions that are most often tested in experimental papers. Figure 1 depicts this.

(Analytical) Utility → Equilibrium Prices/Holdings

↕ tested

(Experimental) Equilibrium Prices/Holdings

Figure 1. Traditional Empirical Tests.

But analytical models take the analysis one step further. From the equilibrium predictions, the analytical model works "backwards" to calculate ex ante EU. This same process – i.e., assume utility maximizers, calculate equilibrium, calculate ex ante EU – can be repeated for different disclosure policies, and this allows the analytical models to compare the ex ante EU under different policies. This extra step, in which EU is calculated, had not been previously tested experimentally. Theoretically, traders' EU follows mathematically from the market equilibrium. But we do not know whether subjects' subjective EU follows these predictions. This is what we tested in our previously reported experiment. We found that subjects prefer the policy with the lower analytical EU. Figure 2 depicts this newer experimental focus.

(Analytical) Utility → Equilibrium Prices/Holdings → ex ante EU

↕ Tested, prediction not supported

(Experimental) Equilibrium Prices/Holdings → subjective EU (preferences)

Figure 2. Our Experimental Focus in Bodoff et al. (2003).

The main result of those experiments was that subjects did not prefer the policy ($\lambda = 1$) with the higher ex ante EU, but preferred instead $\lambda = \frac{1}{2}$, with its lower ex ante EU. That is, they did not prefer the policy that the model predicts.

2.1. Investigating an Anomaly

In this paper we explore the question, did subjects not prefer the predicted policy because the model failed to accurately predict the market equilibriums under the two policies? Or, were the predictions of market equilibrium upheld, and subjects did not prefer the policy with the higher mathematically-calculated expected utility? This would be an anomaly, because utility-maximizing subjects ought to prefer the condition with the higher ex ante EU. Figure 3 depicts this question, which is the focus of this paper.

(Analytical) Utility → Equilibrium Prices/Holdings → ex ante EU

↕ HERE? ↕ OR HERE?

(Experimental) Equilibrium Prices/Holdings → subjective EU (preferences)

Figure 3. Where did model predictions go wrong?

The detailed analysis in section 5 below will show the following: The market Equilibrium was generally well predicted by the theory. Using a variety of calculations, we find that, in accordance with the analytical model, the ex ante EU of an average trader in our actual market was lower in the condition $\lambda = \frac{1}{2}$ than with $\lambda = 1$. And still, subjects expressed a preference for condition $\lambda = \frac{1}{2}$ over $\lambda = 1$. In terms of figure 3, this means that the predictions held up well regarding market behavior (predicted equilibriums), but still faltered on the rightmost side, when it came to predicting subjects' preferences (predicted ex ante EU). This is the anomaly that we explore in this paper.

To summarize, our experiment previously reported in Bodoff, Levecq, et al. (2003) extended the range of predictions to be tested experimentally, as shown in figure 2. We tested both price/holdings equilibrium predictions, as well as predictions of EU. The anomaly we explore in the remainder of this paper is that the experimental data fairly well supports the price/holdings equilibrium predictions, but subjects' preference for the different treatments do not conform to any obvious calculation of EU *based* on those equilibriums. Instead, subjects prefer $\lambda = {}^1\!/_2$, which is the more risky treatment according to a traditional calculation of ex ante EU.

3. EXPERIMENTAL DESIGN

We ran 12 sessions, each consisting of a series of 18 independent trading periods in which 12 subjects/investors competed to buy and/or sell a fictitious security. The subjects were undergraduate business students in Hong Kong University of Science and Technology with no prior trading experience. To test for cultural effects, half of the sessions were duplicated with undergraduate business students in Baruch College, New York, and no important differences were found. Results reported here are limited to the Hong Kong group, for which we completed the full set of sessions. Each session lasted 90 minutes. Subjects were guaranteed HK$160 (approximately US $20) for their participation, but could earn significantly more based on their performance. The experimental currency was fantasy-franc (FF).

At the beginning of a trading period, a randomly selected half (i.e., six) of the investors received two units of the security (the endowed investors), and the other half received nothing. In order to conduct their trading activities, all investors were allowed a line of credit of FF1,000, which had to be repaid at the end of the trading period. The treatment conditions were $\lambda = 1/2$, and $\lambda = 1$. In a market with $\lambda = 1/2$, three of the six endowed investors, and three of the six unendowed investors, were randomly selected to receive inside information at the beginning of that period. This was done by giving an envelope to all investors, with some of the envelopes containing information and others containing no information. Thus, in each period of treatment $\lambda = 1/2$ there were three investors of each of the following four categories, randomly assigned and anonymous:

- endowed and informed,
- endowed and uninformed,
- unendowed and informed, and
- unendowed and uninformed.

In a period when all investors are informed, there are only two categories: six "endowed" investors and six "unendowed" investors.

3.1. Payoffs

The security traded in our markets paid a state-dependant dividend at the end of each trading period. In a significant departure from most experimental designs, we did not try to induce trading between subjects by separating them into different types, facing different payoff structures. Instead, trading is naturally motivated by the difference

Table 1. Possible Payoffs of the Stock

	X	*Y*	*Z*
Payoff	0	70	100

in endowments combined with non-neutral and heterogeneous risk preferences. All investors face the same state-dependent payoff schedule (see Table 1) if they held the security until the end of the trading period. For example, a subject who received two securities at the beginning of the trading period, sold one security for FF60 during trading, and received FF70 for her remaining share after state Y was randomly determined, would earn FF130 in that period.

Each subject kept the proceeds accumulated at the end of each trading period. These were accumulated until the end of the experiment, at which time the investors were paid in real Hong Kong dollars.

The state of nature was determined by drawing a ball from an "urn" containing 12 labeled balls, 4 balls representing each state of nature.

3.2. Treatment Variables

Each trading period was characterized by the "quality of information" – partial or complete – released to investors, and the proportion of informed investors.

- Information was either partial or complete. If partial, the number of possible states of the world was reduced to 2. The information was either "not 0 in this round", "not 70 in this round", or "not 100 in this round". If complete, the information revealed the final state for that round, 0, 70, or 100.
- The proportion of informed investor was either zero ($\lambda = 0$), half ($\lambda = 1/2$) or all ($\lambda = 1$).

(note: the description of the experiment reflects the fact that our experiment included three levels of λ: $\lambda = 0$, $\lambda = 1/2$, and $\lambda = 1$. The focus in this paper is on the pair $\lambda = 1/2$, and $\lambda = 1$).

Table 2 lists the five possible market conditions for a trading period. The table should be viewed column-wise, as we do not predict or measure comparisons across markets, only across levels of λ within one market, i.e., column.

The identity of informed investors, and the states of nature (payoff) were random and independent from period to period, with no guarantee of balance. Subjects were made aware of this. In conditions #3 and #4, the experimenters needed to draw a ball from the urn before the experiment, in order to know what information to give to the informed investors (half or all). In condition #1 and #2, the experimenters similarly needed to determine, before the experiment, which state of nature would *not* occur in that period, in order to give that information to the informed investors.

Table 2. Five Trading Conditions

Informed Investors	*Market A Partial Information*	*Market B Complete Information*
None $\lambda = 0$	Condition #0	Condition #0
Half $\lambda = \frac{1}{2}$	Condition #1	Condition #3
All $\lambda = 1$	Condition #2	Condition #4

In this case, when it was time to determine the state of nature after the trading round, the 4 balls representing the disqualified state were conspicuously removed from the urn.

3.3. Dependent Variable

We want to investigate the effects on an average investor's ex ante EU, of a policy (technology) that requires (facilitates) an exogenous change in the fraction λ of informed investors. How can we *experimentally measure* ex ante EU for the two levels of λ?

In this study, we solve (avoid) numerous methodological issues by abstracting the measurement of ex ante EU: After exposing subjects to markets with two different levels of λ, we ask subjects to specify which of two levels of λ they prefer. The prediction is that utility-maximizerrs should prefer the condition with the higher analytically-derived ex ante EU.

To ensure honest responses, we proceed as follows. As shown in Table 3, each session begins with 3 training rounds. The remaining 15 rounds are organized in sequences of five trading rounds: two rounds each of two different levels of λ, then a vote and one extra round. For example, session 1 rounds 4–7 is followed by a vote, and round 8 depends on that vote. To continue the example, session 1 rounds 4–7 had two rounds of condition 0 (none-informed) and two rounds of condition 2 (all-informed with complete information). After the conclusion of these four rounds to familiarize subjects with two different levels of λ, there is a vote. We announce that there will be one additional round to complete the set of five, and that the information condition to be used in that extra round will depend on the group's majority preference. For example, following rounds 4–7, subjects are asked to vote anonymously whether they would prefer condition #0 or #2 in period 8. In the additional period as with all periods, subjects will face randomly assigned endowments, and if $\lambda = \frac{1}{2}$, they will face random assignment to be either informed or uninformed. Also as usual, profits will be added to their totals. Thus, when they vote, subjects do not know what their endowments will be in the extra period, and in the case of a vote favoring $\lambda = \frac{1}{2}$, they do not know whether they will be informed or uninformed in that period.

Table 3. Information Condition for Each Trading Period

Experiment session	*Trading Period*									
	1–3 Training	*4–7*	*Preference elicitation*	*8*	*9–12*	*Preference elicitation*	*13*	*14–17*	*Preference elicitation*	*18*
1	0–2–1	2–2–0–0	0 vs 2	0 or 2	1–1–2–2	2 vs 1	1 or 2	0–0–1–1	0 vs 1	0 or 1
2	0–2–1	0–2–0–2	0 vs 2	0 or 2	2–1–2–1	2 vs 1	1 or 2	1–0–1–0	0 vs 1	0 or 1
3	0–2–1	0–0–1–1	0 vs 1	0 or 1	2–2–0–0	0 vs 2	0 or 2	1–1–2–2	2 vs 1	1 or 2
4	0–2–1	1–0–1–0	0 vs 1	0 or 1	0–2–0–2	0 vs 2	0 or 2	2–1–2–1	2 vs 1	1,or 2
5	0–2–1	1–1–2–2	2 vs 1	1 or 2	0–0–1–1	0 vs 1	0 or 1	2–2–0–0	0 vs 2	0 or 2
6	0–2–1	2–1–2–1	2 vs 1	1 or 2	1–0–1–0	0 vs 1	0 or 1	0–2–0–2	0 vs 2	0 or 2
7	0–4–3	4–4–0–0	0 vs 4	0 or 4	3–3–4–4	4 vs 3	3 or 4	0–0–3–3	0 vs 3	0 or 3
8	0–4–3	0–4–0–4	0 vs 4	0 or 4	4–3–4–3	4 vs 3	3 or 4	3–0–3–0	0 vs 3	0 or 3
9	0–4–3	0–0–3–3	0 vs 3	0 or 3	4–4–0–0	0 vs 4	0 or 4	3–3–4–4	4 vs 3	3 or 4
10	0–4–3	3–0–3–0	0 vs 3	0 or 3	0–4–0–4	0 vs 4	0 or 4	4–3–4–3	4 vs 3	3 or 4
11	0–4–3	3–3–4–4	4 vs 3	3 or 4	0–0–3–3	0 vs 3	0 or 3	4–4–0–0	0 vs 4	0 or 4
12	0–4–3	4–3–4–3	4 vs 3	3 or 4	3–0–3–0	0 vs 3	0 or 3	0–4–0–4	0 vs 4	0 or 4

In this manner, the preference is ex ante, and subjects are motivated to vote for the information condition they actually prefer. We rotated the sequence to ensure that each pair-wise comparison appears both towards the beginning and towards the end of some sessions. We also rotated the order that the two λ values were presented in a four-period sequence. All this was done in order to completely isolate the treatment variable, so that the recorded preferences can only reflect the treatment itself.

There are six pairs of market condition comparisons:

Market A'
0 vs. 1 – None informed vs. half informed with Partial information
0 vs. 2 – None informed vs. all informed with Partial information
1 vs. 2 – Half informed vs. all informed with Partial information

Market B
0 vs. 3 – None informed vs. half informed with Complete information
0 vs. 4 – None informed vs. all informed with Complete information
3 vs. 4 – Half informed vs. all informed with Complete information

Each experimental session involves different levels of λ for only one information quality, partial or complete. No attempt was made to predict or elicit preferences across Markets.

3.4. Trading Procedure

The trading institution is an electronic double auction with limit order book. Each subject is assigned an identifier, and conducts trading from a computer terminal. Although all subjects are in the same room, trading is anonymous: subjects do not know who posts an order or whom they are trading with. Subjects can submit limit orders to buy and sell, as well as market orders when a counterpart limit order exist. Subjects can cancel outstanding limit orders. Each order is for one single unit of security. The trading period ends after five minutes, or after one minute without any trade, whichever comes first. At the end of the trading period, the final state of the world is determined, and payoffs are computed.

The user interface appears as shown in Figures 4 and 5. While Figure 4 shows the market-wide summary screen that is projected for all investors to see, Figure 5 shows an example of an individual investor's private screen, which shows his/her own holdings, available cash, current orders, and profits to date.

4. MODELS AND PREDICTIONS

We assume a negative exponential utility function and solve a market-clearing equilibrium for the specific conditions of our experimental market. Figures 6–7 show the calculations of ex ante EU for an average trader, under treatment conditions $\lambda = \{0, \frac{1}{2}, 1\}$, for a market where information is partial (figure 6) or complete

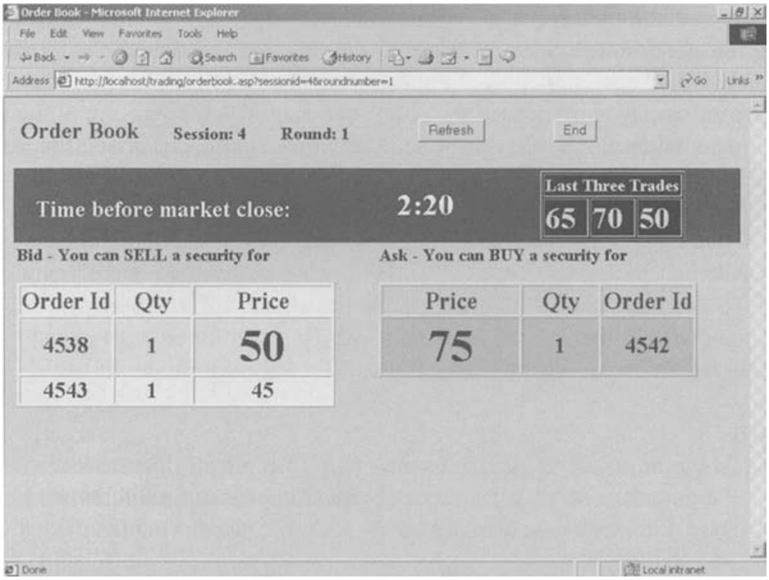


Figure 4. Screen Seen by All Investors.

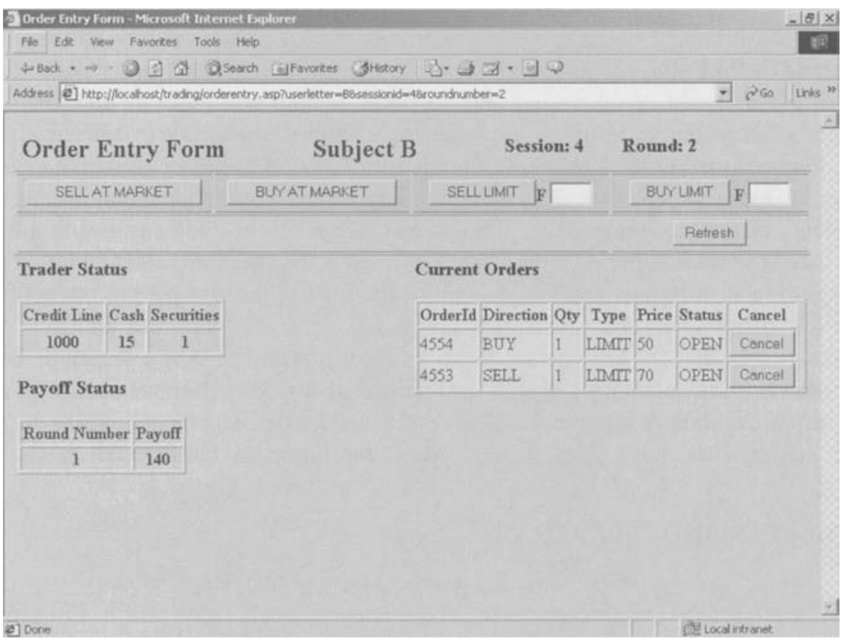


Figure 5. An Individual Investor's Screen.

No info

Trader Type	Final Holding	Net sold	Utility
Endowed	1	1	-0.8341
Unendowed	1	-1	-0.9975
			-0.9158
	Price =	53.688	

Partial info for All	Info = not X			Info = not Y			Info = not Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Endowed	1	1	-0.7540	1	1	-0.8553	1	1	-0.8944
Unendowed	1	-1	-0.9997	1	-1	-0.9966	1	-1	-0.9983
			-0.8768			-0.9259			-0.9464
	Price =	84.625		Price =	45.852		Price =	32.963	-0.9164

Partial info for Half	Info = not X			Info = not Y			Info = not Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Uninformed/Endowed	0	2	-0.7552	2	0	-0.8583	2	0	-0.8959
Uninformed/Unendowed	0	0	-1.0000	2	-2	-1.0161	2	-2	-1.0608
Informed/Endowed	2	0	-0.7542	0	2	-0.8446	0	2	-0.8446
Informed/Unendowed	2	-2	-0.9988	0	0	-1.0000	0	0	-1.0000
			-0.8770			-0.9298			-0.9503
	Price =	84.251		Price =	50.656		Price =	50.656	-0.9190

Figure 6. Market A Partial Information.

No info

Trader Type	Final Holding	Net sold	Utility
Endowed	1	1	-0.8341
Unendowed	1	-1	-0.9975
			-0.9158
	Price =	53.688	

Complete info for All	Info = X			Info = Y			Info = Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Endowed	2		-1.0000	2		-0.7919	2		-0.7165
Unendowed	0		-1.0000	0		-1.0000	0		-1.0000
			-1.0000			-0.8959			-0.8583
There is no profitable trade.									-0.9181

Complete info for Half	Info = X			Info = Y			Info = Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Uninformed/Endowed	2	0	-1.0000	2		-0.7919	2		-0.7165
Uninformed/Unendowed	2	-2	-1.1839	0		-1.0000	0		-1.0000
Informed/Endowed	0	2	-0.8446	2		-0.7919	2		-0.7165
Informed/Unendowed	0	0	-1.0000	0		-1.0000	0		-1.0000
			-1.0071			-0.8959			-0.8583
	Price =	50.656		Price =	70		Price =	100	-0.9205
				There is no profitable trade.			There is no profitable trade.		

Figure 7. Market B Partial Information.

(figure 7). "Ex ante" means before the state of nature is known and before the trader's endowment is known. In the market with partial information, it also means before the information signal is determined, i.e. before it is determined whether informed traders will be told "Not 0" or "Not 70" or "Not 100". In addition, in the case of $\lambda = {}^1/_2$, ex ante means before a trader knows if he/she will be informed. Further explanation of the equilibrium calculations is found in Bodoff, Levecq, et al. (2003).

For both non-revealing (NRE) and fully revealing (FRE) prices, the solved model predicts that an average trader's ex ante EU for the three treatment conditions would be ordered as:

$$\lambda = 0 > \lambda = 1 > \lambda = 1/2$$

That is, a utility maximizer should most prefer $\lambda = 0$, then $\lambda = 1$, then $\lambda = 1/2$. In our experiment, the dependent variable of interest is actually the (order of) ex ante EU, rather than equilibrium prices or holdings per se. This prediction holds for both partial and complete information. For the case of partial information, we can show that the results generalize to other distributions of payoffs. In all calculations, we use the utility function $-\exp(-w/r)$ with r = FF 600. In (Bodoff, Levecq et al. 2003) we explain that r = FF 600 is a reasonable choice and that predictions do not change for any r > FF 600.

We will be focusing primarily on the unexpected preference for $\lambda = 1/2$. More specifically, we will be investigating how subjects came to prefer this most "risky" option. We therefore call particular attention to aspects of the calculations for $\lambda = 1/2$ which make this treatment most risky and to have the lowest EU of all the treatments.

First, it will be immediately noticed from the analytical predictions that there are four rows under the condition $\lambda = 1/2$, corresponding to the four combinations of a trader's being endowed/un-endowed, and informed/uninformed. The additional chance element of informed/uninformed represents an extra source of uncertainty in the case of $\lambda = 1/2$. All else being equal, such an additional source of uncertainty is detrimental to a risk-averse trader. Moreover, "all else" is not equal. The profits of the informed are predicted to be greater than those of the uninformed, and this unevenness further exacerbates the detrimental effect of the extra layer of uncertainty. Finally, though more subtly, the effect of information is not only that informed profit more than uninformed, but it also affects prices and holdings in such a way that each of the four categories – uninformed/endowed, uninformed/unendowed, informed/endowed, and informed/undendowed – has a unique expectation under each possible signal.

5. RESULTS

Table 4. Preferences in Market A with Partial Information

	$\lambda = 0$	$\lambda = 1$	$\lambda = 1/2$**	$\lambda = 1$	$\lambda = 1/2$**	$\lambda = 0$
#Votes	41	31	51	21	49	23

Table 5. Preferences in Market B with Complete Information

	$\lambda = 0$**	$\lambda = 1$	$\lambda = 1/2$**	$\lambda = 1$	$\lambda = 1/2$	$\lambda = 0$*
#Votes	62	10	60	12	28	44

Tables 4 and 5 show the number of votes for each condition in the three pairwise comparisons in both markets. A single (double) asterisk indicates a one-tailed (two-tailed) 95% statistically significant preference using a Chi-square test of the difference from a 50–50 vote. The one-tailed test is used only if the result is in the predicted direction.

Subjects' actual preferences:
In Market A, $(\lambda = 1/2) > (\lambda = 0)$ *and* $(\lambda = 1)$, *indifferent between* $(\lambda = 0)$ *and* $(\lambda = 1)$.
In Market B, $(\lambda = 0) > (\lambda = 1/2) > (\lambda = 1)$.

The basic result is that contrary to the model's predictions, subjects always preferred $\lambda = 1/2$ over $\lambda = 1$.

5.1. Do Subjects' Preferences Coincide with EU of Each Market?

In this section we undertake to explain and explore why our subjects preferred the condition ($\lambda = 1/2$) over ($\lambda = 1$) in both markets (see figure 3 above). There are basically two sorts of explanations. The first possible direction is that the predicted prices and equilibrium holdings did not materialize as predicted. That is, although a risk-averse subject *would* dislike the *analytically predicted* equilibrium of treatment ($\lambda = 1/2$), the actual equilibrium did not materialize as predicted. The second direction is that the market behavior did unfold in a manner that makes ($\lambda = 1/2$) the most risky option with the lower EU as predicted, but that subjects' preferences were determined by something other than a traditional calculation of EU.

We find that the predicted patterns of trade and equilibrium holdings correspond to predictions. The reader is referred to Bodoff, Levecq et al. (2003) for details and statistical tests. For example, the average final holding of the unendowed traders was .62 shares in the case of Market A, $\lambda = 1$, while endowed traders held 1.32 in equilibrium. This is not the completely equal risk-sharing predicted by the model, but clearly the trade was in the direction of risk-sharing, as predicted by the models that assume risk-averse traders. The pattern of equilibrium prices across conditions also followed the pattern of predictions for partially revealing rational expectations. The one element in the data that did not correspond to predictions, was that the absolute level of prices was consistently above the (posterior, where information was available) expected value of the stock.

Having found that in general the predicted patterns of trade did materialize as predicted, we turn to specifically confirm whether the specific (detrimental) effects of condition ($\lambda = 1/2$) materialized as predicted. The reason for this focus is that we want to understand how and why subjects preferred this condition, which is predicted to be most risky and least favored by risk-averse traders. In section 2 we identified the possible detrimental effects of asymmetric information. These effects operate through asymmetric profits of informed versus uninformed traders, which imbalance is detrimental ex ante to an average risk-averse trader. Table 6 compares the profits of the informed against those of the uninformed in our market.

Table 6. NRE Predicted (Actual) Profits, in FF$

Market A
Partial information

Half informed	*Informed*	*Not informed*	*Number of obs*	Anova *F*	*P-value*
Not X	$85.75 ($89.69)	$84.25 ($87.23)	156	0.03	0.86
Not Y	$50.65 ($52.64)	$49.35 ($47.36)	168	0.09	0.76
Not Z	**$50.65 ($52.31)**	**$19.35 ($17.69)**	**72**	**2.39**	**0.13**

Market B
Complete information

Half informed	*Informed*	*Not informed*	*Number of obs*	*F*	*P-value*
X	**$50.65 ($23.44)**	**$−50.65 ($−23.44)**	**96**	**5.51**	**0.02**
Y	$70.00 ($72.18)	$70.00 ($67.82)	156	0.14	0.71
Z	$100.00 ($102.55)	$100.00 ($97.45)	120	0.08	0.78

Table 6 shows that the informed do indeed benefit at the expense of the uninformed. More specifically, these unequal profits occur in the case of bad news, but not in the case of good news, as predicted in the solved model. The table shows NRE-predicted versus actual profits (for FRE, there is no difference between insiders and outsiders) and provides ANOVA tests for the differences in actual profits between informed and uninformed (the ANOVA is one-way, treating Markets A and B separately, as the effects of information are specifically and differently modeled in the two cases). In the case of not-Z of Market A, the difference in average profits was 52.31 versus 17.69. However due to chance, there was a relatively smaller number of data points for that condition. Combined with the relatively high variance, this large difference of 52.31 versus 17.69 did not achieve statistical significance. In State X of Market B, the difference was between −23 and +23 and was statistically significant. The experimental results lie somewhere between NRE and FRE predictions, which is consistent with previously reported experimental results that support a theory of partially revealing prices. In summary, the basic detrimental effect of (λ = ½) did materialize as predicted. This begins make us wonder why subjects preferred this condition.

We have thus far only compared the profits of informed versus uninformed, because this difference is the biggest part of the detrimental (to EU) effects of λ = ½.

However, the information has other subtle effects on price and holdings, that ultimately affect not only informed versus uninformed, but individually affect all four categories of informed/uninformed with endowed/unendowed. We therefore present a more complete analysis of utility, given the experimental markets' equilibrium prices and holdings for both $\lambda = {}^1/_2$ and $\lambda = 1$.

Figures 8–9 calculate actual utility for each treatment and every trader, using our experimental markets' *actual* equilibrium *prices and holdings* under every market condition. The question is, can we understand subjects' preferences in light of the actual equilibriums of each market? Figures 8–9 calculate actual ex ante EU for each treatment, using the actualized (as opposed to analytically predicted) equilibrium

No info

Trader Type	Final Holding	Net sold	Utility
Endowed	1.27	0.73	-0.8158
Unendowed	0.73	-0.73	-1.0202
			-0.9180
	Price =	72.000	

Partial info for All

	Info = not X			Info = not Y			Info = not Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Endowed	1.38	0.62	-0.7491	1.17	0.83	-0.8307	1.23	0.77	-0.8774
Unendowed	0.62	-0.62	-1.0063	0.83	-0.83	-1.0262	0.77	-0.77	-1.0178
			-0.8777			-0.9285			-0.9476
	Price =	91.000		Price =	67.000		Price =	48.000	-0.9179

Partial info for Half

	Info = not X			Info = not Y			Info = not Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Uninformed/Endowed	1.05	0.95	-0.7523	1.1	0.9	-0.8262	1.72	0.28	-0.8861
Uninformed/Unendowed	0.67	-0.67	-1.0013	0.83	-0.83	-1.0291	0.83	-0.83	-1.0293
Informed/Endowed	1.21	0.79	-0.7526	1.02	0.98	-0.8236	1.06	0.94	-0.8641
Informed/Unendowed	1.08	-1.08	-1.0022	1.05	-1.05	-1.0378	0.39	-0.39	-1.0133
			-0.8771			-0.9291			-0.9482
	Price =	86.000		Price =	69.000		Price =	55.000	-0.9181

Figure 8. Market A Partial Information.

No info

Trader Type	Final Holding	Net sold	Utility
Endowed	1.27	0.73	-0.8188
Unendowed	0.73	-0.73	-1.0164
			-0.9176
	Price =	69.000	

Complete info for All

	Info = X			Info = Y			Info = Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Endowed	2		-1.0000	2		-0.7919	2		-0.7165
Unendowed	0		-1.0000	0		-1.0000	0		-1.0000
			-1.0000			-0.8959			-0.8583
									-0.9181

There is no profitable trade.

Complete info for Half

	Info = X			Info = Y			Info = Z		
Trader Type	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility	Final Holding	Net sold	Utility
Uninformed/Endowed	1.67	0.33	-0.7784	1.64	0.36	-0.8432	1.3	0.7	-0.8273
Uninformed/Unendowed	1.46	-1.46	-0.8690	0.51	-0.51	-1.0198	1.3	-1.3	-1.1545
Informed/Endowed	0.88	1.12	-0.8396	1.26	0.74	-0.8284	1.77	0.23	-0.8726
Informed/Unendowed	0	0	-1.0000	0.59	-0.59	-1.0231	0.77	-0.77	-1.0881
			-0.8717			-0.9286			-0.9856
	Price =	27.000		Price =	72		Price =	100	-0.9287
				There is no profitable trade.			There is no profitable trade.		

Figure 9. Market B Complete Information.

prices and holdings. For ($\lambda = 0$) a trader faces one random event aside from the state of nature, i.e., his/her endowment. Therefore, the figure separately calculates the expected utility ("expected" over the three possible states of nature) of an endowed trader and an unendowed trader, then averages these. The result is ex ante expected utility for the actual experimental markets, where "ex ante" means before a trader in that treatment knows if he/she will be endowed. The only unusual aspect of these calculations is that they use our actual market equilibrium prices and holdings, rather than the predicted values. Continuing to ($\lambda = 1$), a trader in this condition faces two random events, i.e. endowment and the content of the information signal. The figure thus shows six cells whose expected (over the possible states of nature) utility is calculated based on empirical equilibrium, then averaged to get ex ante EU, where "ex ante" means before the trader knows if he/she will be endowed and before the information signal is determined. For ($\lambda = 1/2$) a trader faces those same chance events plus also the chance of being an insider or an outsider. The figure thus shows 12 cells whose ex ante EU is calculated in a corresponding manner.

The main result of figures 8–9 is that ($\lambda = 1/2$) has the lowest EU.

Figures 10–11 take this analysis one step further, by considering actual average profits under each condition. By considering actual profits, figures 10–11 are in effect considering the full path and price of all trades, as well as the actual distribution of information signals. This is different from the analysis of figures 8–9 which only considered the actualized *equilibriums*. Once again, we find that ($\lambda = 1/2$) has the worst ex ante EU.

We have thus arrived at the opposite end of the spectrum, in the continuum from analytically predicted equilibrium to actual realizations of our market. On the basis of all this data, we return to discuss where we think the analytical predictions went wrong, whether in the estimation that ($\lambda = 1/2$) is most risky, or whether in assuming

No info

Trader Type	Final Profit	Utility	
Endowed	121.3	-0.816959	
Unendowed	-6.6	-1.011061	
			-0.9140

Partial info for All

	Info = Not X			Info = Not Y			Info = Not Z		
Trader Type	Final Profit	Utility		Final Profit	Utility		Final Profit	Utility	
Endowed	162.77	-0.7624		106.2	-0.83778		131.28	-0.803482	
Unendowed	-6.4	-1.010724		-46.2	-1.080042		8.7	-0.985605	
			-0.8866			-0.9589			-0.8945
									-0.9133

Partial info for Half

	Info = Not X			Info = Not Y			Info = Not Z		
Trader Type	Final Profit	Utility		Final Profit	Utility		Final Profit	Utility	
Uninformed/Endowed	168.33	-0.755368		112.3	-0.829306		65.6	-0.896432	
Uninformed/Unendowed	9.7	-0.983963		-36.8	-1.063253		-43.22	-1.074691	
Informed/Endowed	174.5	-0.74764		121.3	-0.816959		97.1	-0.850583	
Informed/Unendowed	8.9	-0.985276		-36.9	-1.06343		-26.1	-1.04446	
			-0.8681			-0.9432			-0.9665
									-0.9259

Figure 10. Market A Partial Information, ex ante EU based on Actual Total Profits.

No info

Trader Type	Final Profit	Utility	
Endowed	-6.7	-1.011229	
Unendowed	121.3	-0.816959	
			-0.9141

Complete info for All

	Info = X			Info = Y			Info = Z		
Trader Type	Final Profit	Utility		Final Profit	Utility		Final Profit	Utility	
Endowed	0	-1		140	-0.79189		200	-0.716531	
Unendowed	0	-1		0	-1		0	-1	
			-1.0000			-0.8959			-0.8583
									-0.9181

Complete info for Half

	Info = X			Info = Y			Info = Z		
Trader Type	Final Profit	Utility		Final Profit	Utility		Final Profit	Utility	
Uninformed/Endowed	26.38	-0.956986		136.77	-0.796164		194.71	-0.722877	
Uninformed/Unendowed	-72	-1.127497		-0.7	-1.001167		0.238	-0.999603	
Informed/Endowed	45.66	-0.926724		143.59	-0.787166		202.57	-0.713469	
Informed/Unendowed	0	-1		0.33	-0.99945		2.476	-0.995882	
			-1.0028			-0.8960			-0.8580
									-0.9189

Figure 11. Market B Complete Information, ex ante EU based on Actual Total Profits.

that subjects' preferences would directly shadow the order of ex ante EU for each condition.

5.2. *Discussion of Anomaly*

We know that the analytical model did not properly predict subjects' preferences regarding λ. In the previous section we have supplied evidence that helps explain where the models went wrong. Did the model fail to accurately predict equilibrium prices and holdings? Or did the model accurately predict prices, holdings, and E? In this latter case, subjects were not preferring the conditions with the higher EU.

The evidence from the previous section shows that in the two markets, with two approaches to measurement, in every case the empirically based ex ante EU was lower for ($\lambda = {}^1\!/\!_2$) than for ($\lambda = 1$). That is, not only was the analytically predicted ex ante EU lower for ($\lambda = {}^1\!/\!_2$) than for ($\lambda = 1$), but so was the empirically based ex ante EU. We have thus found nothing in the empirical behavior of our markets that could explain subjects' preference for ($\lambda = {}^1\!/\!_2$) in terms of ex ante EU. Of course, this analysis is limited especially by our choice of utility functions. There may of course exist another – or an arbitrary – utility function according to which our subjects' preferences were rational. But adopting the negative exponential, which is used in many papers that make policy prescriptions, both the analytical predictions and every conceivable way of calculating EU based on the empirical market, all point to lower EU for the case of ($\lambda = {}^1\!/\!_2$). What, then, could explain subjects' preference for ($\lambda = {}^1\!/\!_2$)?

6. KINDS OF RISK

Our approach to explain these results is influenced by a style of discussion that is found in analytical papers of the welfare effects of information. Authors of analytical models will often analyse formula they've derived for ex ante EU, and identify each

Table 7. Synthesis of Kinds of Risk

	Knowledge	*Trading Opportunities and Risk-Sharing*	*Excess Profits*
$\lambda = 0$			
$\lambda = 1$	Lower risk vis-à-vis states of nature	Signal Risk Endowment Risk	
$\lambda = 1/2$	Lower risk vis-à-vis states of nature, if informed	(Some) Signal Risk (Some) Endowment Risk Defensiveness of outsiders Uneven final holdings	Insiders profit at expense of outsiders

component of the formula as signifying a certain kind of risk, such as signal risk or endowment risk. These associations are intended as intuitive explanations for the components of a traditional, if complex formula for ex ante EU. For example, the full expression of ex ante EU in Indjejikian (1991) includes terms that the author associates with "endowment risk", i.e., the risk of being stuck with a non-optimal endowment as a result of too much common information which reduces opportunities for trade. As another example, Alles and Lundholm (1993) analyze their equilibrium model with an emphasis on "signal risk", through which information affects prices. Such explanations are not offered in lieu of a traditional calculation of ex ante EU, but as intuitive explanations for the reasons some conditions end up with lowed ex ante EU than others.

We propose to take this analysis one step further. We propose that subjects actually have different attitudes toward these different kinds of risk. We believe that our experimental data is most parsimoniously explained by this approach.

Table 7 categorizes and summarizes the various effects that information is modeled to have on ex ante EU in a variety of models that are based on partially revealing rational expectations (PRE). PRE has been shown experimentally to make superior predictions to either non-revealing (NRE) or fully-revealing (FRE) prices; indeed, our own experimental data lies somewhere in between NRE and FRE theories.

We propose that our subjects are indeed risk averse but that their preferences for the different treatments depend partly on their attitude toward the different kinds of risk that characterize the different treatments.

Taking all our experimental results, we came to understand that all else being equal, our subjects value having information, but not if it means being stuck with endowments or otherwise prevented from trading. For these reasons, they dislike the treatment condition $\lambda = 1$. In terms of table 7, they like knowledge (don't like "ignorance risk"), but dislike still more "endowment risk" and "signal risk". On the other hand, our subjects appear to actually welcome the risk of being an insider or outsider.

An experiment that is specifically designed with this phenomenon in mind, would arrange for two very different markets, with two different sets of risks, in which the equilibriums were identical. Then, subjects' preferences for one market over the other could only be explained in terms of differing attitudes to different sources of risk. In our experiment, the markets with $\lambda = 1/2$ and $\lambda = 1$ don't yield identical equilibriums. Instead, we have a result that shows that by a traditional calculation, the market with $\lambda = 1$ is *at least as good as* (not identical to) the market with $\lambda = 1/2$. Yet still, our subjects strongly prefer $\lambda = 1/2$ over $\lambda = 1$. Based on this, we have come to believe that our subjects' preferences must reflect something more than a dry calculation of EU. We propose that our subjects' preferences depend not only on the numeric risk associated with an equilibrium, but with the process through which those positions were attained. More specifically still, it appears that subjects strongly dislike an imbalance that is the result of being stuck with unequal endowment, but actually like information asymmetries, even though they invariably lead to equally uneven equilibriums.

There are in fact a number of possible explanations for why subjects might view favorably (the asymmetries that result from) the risk of being assigned to an insider versus outsider. For one thing, the gamble of being an insider/outsider is not as extreme as the gamble that arises from the state of nature. So, a model of utility such as Conlisk (1993) that considers local risk-seeking with global risk-aversion may be able to account for our subjects' preferences. We say "may", because while these psychological effects have been demonstrated in lotteries, we are not aware of any equilibrium model of trade that begins with such utility functions.

Another possible explanation is to differentiate not between local and global, but between different kinds of risk. In this interpretation, subjects may be risk averse vis-à-vis the state of nature, but not averse to the ongoing chance of being an insider sometimes and an outsider at other times. This may be related to the question of risk *perception* Gaba and Viscusi (1998). A trader may not perceive the risk that comes from being an outsider relative other traders, because of hubris. That is, a trader may feel that if she is an outsider in any given round, then she will be clever enough to hold back until the information is revealed; while if she is an insider, she can always hope that there will be some outsider who is not so careful, who will yield up excess profits to the insider. In the terminology of Gaba and Viscusi (1998), a trader may indeed be risk averse, but may not perceive that there is risk to him/her in the 50–50 allocation to insider versus outsider. Another possible explanation is that subjects exaggerate their chances of being informed, even though they are told the chances are 50–50. This "unrealistic optimism" (Weinstein 1980), a phenomenon related to self-positivity bias, is not hubris of skill, but over-confidence regarding *chance* events. Other possibilities relate to the question of "sources" of uncertainty (Fox and Weber 2002).

6.1. Summary and Conclusion

Analytical models of equilibrium, such as rational expectations models, have been successful in predicting equilibrium prices in experimental markets of risky assets.

In previous work, we explored whether such models are also useful in their other predictions regarding welfare in the sense of ex ante expected utility. We previously found that they are not, i.e. subjects did not prefer the predicted market condition. In this paper, we explored the tension between the correct predictions of equilibrium and the incorrect predictions of subjects' preferences. In analytical models, predictions of EU follow by definition from the equilibrium prices, so it would be expected that if a theory properly characterizes the equilibrium, then it will properly predict ex ante EU. But this is apparently not the case, which suggests an anomaly. If market equilibriums were perfectly accurate, then the anomaly would be total. Because predictions of market equilibriums are not perfect, we explored the possibility that perhaps subjects' preferences were consistent with the expected utility of the actual market equilibriums, if not with the analytically predicted market equilibrium. We found that they still were not.

Numerous methodological questions must be addressed. But we conjecture that subjects' preferences depend on something other than ex ante EU as traditionally calculated. One implication of our finding is that in order to justifiably refer to an analytical model in support of welfare-related policy choices, it may be insufficient to test a model's predictions of equilibrium market prices. Rather, it may be necessary to separately test the model's welfare predictions, since the experimental welfare results do not automatically mirror the experimental price results, as is the case with analytical models, where the two are related by definition.

ACKNOWLEDGMENT

This research was supported by the Centre for Experimental Business Research, grant HKUST-3 "Experiential based teaching for networked economy", and the Research Grants Council of Hong Kong, project no. HKUST6056/01H.

NOTE

[1] There are certain conditions in which vigilant traders in a rich market can insure themselves against this negative effect, if they know the information is going to be released.

REFERENCES

Alles, M. and R. Lundholm (1993). "On the Optimality of Pubilc Signals in the Presence of Private Information." *The Accounting Review*, 68(1), 93–112.

Ausubel, L. M. (1990). "Insider Trading in a Rational Expectations Economy." *The American Economic Review*, 80(5), 1022–1041.

Bloomfield, R. and M. O'hara (1999). "Market Transparency: Who Wins and Who Loses?" *The Review of Financial Studies*, 12(1), 5–35.

Bodoff, D., H. Levecq, H. Zhang. (2004). "The Welfare Effect of Wider Information Distribution In An Experimental Market Of Risky Assets." Forthcoming in Journal of Experimental Economics

Bodoff, D. and H. Zhang (2003). "Corporate Disclosure Information: When More is Less." *Decision Support Systems*, 35, 455–466.

Conlisk, J. (1993). "The Utility of Gambling." *Journal of Risk and Uncertainty*, 6, 255–275.

Diamond, D. W. (1985). "Optimal Release of Information by Firms." *The Journal of Finance*, 40(4), 1071–1094.

Fox, C. R. and M. Weber (2002). "Ambiguity Aversion, Comparative Ignorance, and Decision Context." *Organizational Behavior and Human Decision Processes*, 88(1), 476–498.

Gaba, A. and W. K. Viscusi (1998). "Differences in Subjective Risk Thresholds: Worker Groups as an Example." *Management Science*, 44(6), 801–811.

Indjejikian, R. J. (1991). "The Impact of Costly Information Interpretation on Firm Disclosure Decisions." *Journal of Accounting Research*, 29(2), 277–301.

Leland, H. E. (1992). "Insider Trading: Should it be Prohibited?" *The Journal of Political Economy*, 100(4), 859–887.

Lev, B. and J. A. Ohlson (1982). "Market-Based Empirical Research in Accounting: A Review, Interpretation, and Extension." *Journal of Accounting Research*, 20(Supplement), 249–322.

Weinstein, N. D. (1980). "Unrealistic Optimism about Future Life Events." *Journal of Personality and Social Psychology*, 39, 806–820.

This page intentionally blank

Chapter 4

EFFECTS OF IDIOSYNCRATIC INVESTMENTS IN COLLABORATIVE NETWORKS: AN EXPERIMENTAL ANALYSIS

Wilfred Amaldoss
Duke University

Amnon Rapoport
University of Arizona
and
Hong Kong University of Science and Technology

Abstract

There has been an increase in the incidence of firms who collaborate to develop and market new products. Partners in these collaborations often make investments that are idiosyncratic to the collaboration and have limited value outside the scope of the alliance. Amaldoss and Rapoport (2003) reports that the joint investment of a network does not decrease if the number of partners increases. Specifically, if the investments are only partially idiosyncratic, the joint investment increases as a network grows in size. If they are fully idiosyncratic, then the joint investments are predicted to be independent of the number of partners. In this paper, we test these predictions in a laboratory setting. The experimental results support the qualitative predictions of the model, and adaptive learning accounts for the investment patterns of our subjects over time. We also extend the model to consider competitions among more than two networks.

1. INTRODUCTION

Collaborations are becoming more popular especially in high technology industries. For example, we see Motorola, IBM, and Apple working together on the PowerPC chip, while Microsoft and Intel innovate on the Wintel standard. In part this growth in collaborations has been encouraged by the National Cooperative Research Act (1984) and the National Cooperative Research and Production Act (1993) that provide antitrust exemption coverage for partnering firms. On examining the *Federal Register* that documents new product alliances, Rindfleish and Moorman (2000) report that 242 alliances were formed during the period 1/1/89 to 3/15/95. Yet there

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 81–112.
© 2005 *Springer. Printed in the Netherlands.*

are only a few studies that investigate new product alliances (e.g., Dutta and Weiss 1997, Robertson and Gatignon 1998). Additionally, much of the literature on strategic alliances is focused on dyadic inter-firm relationships. However, our understanding of the factors that govern competition between larger networks of firms is quite limited (Nohria 1992, Gomes-Casseres 1994, Gulati 1998, Achrol and Kotler 1999).[1]

By being part of an alliance, a firm is likely to gain access to a larger resource base for developing and marketing new products. Yet, it is commonly believed that larger networks are more likely to fail (Gomes-Casseras 1994, p. 9).[2] There is a basis for this belief: As the number of partners in a network increases, its partners may commit fewer resources for the joint endeavor. In agreement with conventional wisdom, the large Mips network is a failure in the microprocessor market, while the smaller PowerPC network is more successful.[3] On the other hand, biotechnology firms such as Genentech and Amgen are successful in developing innovative products by working in collaboration with their network of partners. Also, the Star Alliance, an eleven-member airline network established in 1997, is growing stronger in the face of competition. These observations raise the question: What is the effect of the *number of partners in a network* on fostering collaboration?[4]

In light of the seemingly contradictory empirical evidence, the likely effect of network size on fostering collaboration is not clear. There may be circumstances when the positive effect of network size may override the free-riding problems, making it optimal for firms to forge larger networks. It is possible that the *type of investment* required for the joint effort may moderate the effect of network size. Some joint R&D endeavors call for idiosyncratic investments, because such investments might have limited value outside the scope of the network. Others call for investments that might be recoverable – at least in part – if the joint effort fails. The free-riding problems are probably more acute when the investments are nonrecoverable, as the danger of losing such investments is greater.

Amaldoss and Rapoport (2003) investigate how the joint investments of network partners are influenced by the type of investment. They focus on networks that do not involve equity arrangements: network partners are independent profit-maximizing firms linked together by a common technology platform or product.[5] Contrary to some intuition, their game-theoretic model implies that if the investments are recoverable, the joint investment increases as the number of partners in a network increases. However, if the investments are nonrecoverable, the joint investment in product development does not depend on network size.

In this paper we test the key predictions of Amaldoss and Rapoport (2003) in a laboratory setting, as there is sparse field data on strategic behavior of partners in networks (Gomes-Casseras 1994). In particular, we report the results of an experiment designed to examine the effect of type of investment (recoverable vs. nonrecoverable) and network size (n = 2 vs. n = 4) on fostering innovation. Our results show that when the investments are recoverable, individual network partners invest less, as the network size increases. In keeping with the model prediction, the joint investment of the network increases as the networks grows in size. When the investments are nonrecoverable, theory predicts that the joint investment should not be influenced by network size. We find that the joint investment made by our subjects does not change in a significant way, as the network size increases.

How do naïve subjects come to make investments that are consistent with equilibrium predictions? Our experimental design is intended to answer this question by using iterated games. We find that the investment decisions were mostly guided by previously received payoffs, and that adaptive learning mechanism can account for the major trends over multiple iterations of the stage game. This additional analysis offers greater confidence in the equilibrium prediction, as financially motivated agents can learn to behave in a fashion that conforms to the model predictions.

The rest of the paper is organized as follows: Section 2 outlines Amaldoss and Rapoport (2003) model of competition between networks and highlights the effect of idiosyncratic investments on collaboration. Section 3 presents and discusses a laboratory test of the model. Section 4 explores whether adaptive learning mechanisms can account for the investment patterns observed in the laboratory. Section 5 summarizes the major findings and concludes by briefly discussing the managerial implications of this study.

2. THEORETICAL ANALYSIS

A Model of Inter-network Competition. Consider two networks that are competitively developing a new product, and index them by $i = \{1,2\}$. There are n_i partners in each network. Partner j in network i places value v_{ij} on the winning the competition. This value depends on the benefits it derives from the public good created by the joint endeavor. The size of the benefit could potentially be related to factors such as pre-existing product portfolio, distribution network, and brand-name reputation. The value of v_{ij} is common knowledge. Further, the distribution of v_{ij} is symmetric across networks, though partners within a network could potentially be asymmetric, that is $v_{ik} \neq v_{il}\ \forall\ k \neq l$. Each partner is assumed to invest part of its endowment, denoted by y_{ij} for the joint endeavor. Denote the investment of each network partner by x_{ij} $(0 \leq x_{ij} \leq y_{ij})$, and the joint investment of network i by $X_i = \sum_{j=1}^{n_i} x_{ij}$.

The investments of network partners may be translated into idiosyncratic or network-specific assets that may have limited value outside the scope of the network. For instance, these investments could be in a particular site that is closer to the partners and cannot be easily moved, in physical assets such as equipment and machinery that are specially required for the collaboration, or in human capital to develop special skills required for the project (see Williamson 1983, p. 526 for a detailed discussion on *asset-specificity*). The model allows a fraction of the investments made by each network partner, denoted by α $(0 \leq \alpha \leq 1)$, to be specific for the joint project, and hence that part of the investment is nonrecoverable. In other words, if a network loses the competition, then each of its partners can recover a fraction of its investments, that is $(1 - \alpha)x_{ij}$. In particular, if $\alpha = 1$, the investments are specific for the network and are completely nonrecoverable, whereas if $\alpha = 0$, they are completely recoverable.

In such an inter-network competition, partners face two major sources of uncertainty. The very process of winning the competition could be uncertain because of technological uncertainties that, in turn, create *outcome uncertainty*. Additionally, firms often find it difficult to *ex ante* specify the performance of their partners or monitor their performance. This creates opportunities for firms to engage in behavior that emasculate the spirit of the contract without violating the letter of the contract. Partners in these networks are not sure of the likely behavior of both partners and competitors (e.g., Kogut 1988, Pisano, Russo and Teece 1988). This kind of uncertainty is referred to as *strategic uncertainty*.

Following the literature on rent-seeking, the model assumes that the winner is determined probabilistically (Tullock 1967, 1980; see Nitzan 1994 for a review).[6] Investing more increases the probability of winning, but does not guarantee success. The probability of network i winning the competition, $\Pr(i)$, is determined by the relative utility of its product. Specifically, $\Pr(i) = U(i)/(U(1) + U(2))$, $i = 1, 2$. As the utility of the product developed by network i depends on how the resources are pooled in the network, we assume that $U(i) = X_i = \sum_{j=1}^{n_i} x_{ij}$. Thus, $\Pr(i)$ is a special case of Tullock's contest success function, $\Pr(i) = X_i^R/(X_1^R + X_2^R)$ with $R = 1$, which is very common in the literature on rent-seeking.[7] The network partners make their decisions simultaneously so that they cannot condition their behavior on the decisions of either their partners or their competitors. The inability to observe the actions of partners and competitors creates strategic uncertainty. Thus, the model captures the essence of the two principal uncertainties faced by partners in a network: outcome and strategic uncertainties.

Firm j in network i is assumed to be risk neutral and maximize its expected payoff:

$$\pi_{ij} = y_{ij} - x_{ij} + \Pr(i)v_{ij} + (1 - \alpha)x_{ij}(1 - \Pr(i)).$$

Thus, the inter-network competition is modeled as a non-cooperative n-person game with complete information, where the values of α, n_i, y_{ij}, and v_{ij} are commonly known. The symmetric pure strategy Nash Equilibrium of this game is used to study the effect of idiosyncratic investments on collaborations.

Theoretical Results. Conventional wisdom of alliance managers suggests that as the number of partners in a network increases, the under-investment problem heightens and, consequently, the joint investment in the collaboration decreases. Proposition 1 below runs counter to this intuition.

Proposition 1. If the investments are partially recoverable ($0 \leq \alpha < 1$), then the collective investment of partners in a network increases as the network grows in size ($\frac{\partial X_i}{\partial n_i} > 0$). But if the investments are nonrecoverable ($\alpha = 1$), then the joint investment made by a network does not depend on its size ($\frac{\partial X_i}{\partial n_i} = 0$).

Proof: See Amaldoss and Rapoport (2003).

Proposition 1 suggests that if investments are partially recoverable then joint investment of a network rises as the size of the network increases. To see this, note that if the investments are fully recoverable ($\alpha = 0$), then the joint investment of a network is $X_i = \dfrac{\bar{v}_i}{2 + \dfrac{1}{n_i}}$, where $\bar{v}_i$ is the mean benefit that partners in network i gain from winning the inter-network competition. However, note that $x_{ij} = \dfrac{v_{ij}}{1 + 2n_i}$, suggesting that as the size of a network increases, the *individual* investment of each partner decreases: $\dfrac{\partial x_{ij}}{\partial n_i} < 0$. Why does the joint investment increase as a network grows in size? This is because the incremental investment from additional members more than offsets the total decrement in the investment of current members. Then, dividing x_{i1} by x_{i2}, we obtain $\dfrac{x_{i1}}{x_{i2}} = \dfrac{v_{i1}}{v_{i2}}$, suggesting that the investments of any two partners in a network are proportional to the relative values they place on winning the competition.

To illustrate the first implication of Proposition 1, consider the case where $n_1 = n_2 = 2$ and $v_{ij} = 80$. In equilibrium, if investments are fully recoverable ($\alpha = 0$), then each network should jointly invest $X_1 = X_2 = 32$, and correspondingly each partner in these networks should invest $x_{ij} = 16$. If the size of each network is changed from $n = 2$ to $n = 4$, then in equilibrium collective investment rises further from 32 to $X_1 = X_2 = 35.556$, though the individual investment of each partner sharply declines from 16 to $x_{ij} = 8.889$.

The second implication of Proposition 1 is that, if the investment is nonrecoverable, then the number of partners in a network does not influence the joint investment in the network. This result is consistent with the findings in rent-seeking literature where individuals rather than groups compete. But this finding, too, is not in agreement with conventional wisdom of alliance managers. However, note that $x_{ij} = \dfrac{v_{ij}}{4n_i}$. This implies that, though the network size does not affect the joint investment, it decreases the investments made by individual partners. That is, $\dfrac{\partial x_{ij}}{\partial n_i} < 0$. What is driving this result? We note that as the number of network partners increases, the total decrease in investment from current partners is precisely offset by the incremental investments from the new members.

To appreciate the second implication of Proposition 1, note that if $n_1 = n_2 = 2$, $v_{ij} = 80$ and $\alpha = 1$, then in equilibrium each network should invest 20 ($X_1 = X_2 = 20$), while each partner, in turn, should invest 10 ($x_{ij} = 10$). If the network increases in size from $n = 2$ to $n = 4$, then the predicted joint investment still remains $X_1 = X_2 = 20$, but the individual investment of each partner decreases from 10 to 5. Note that the joint investment is one fourth of the value each partner places on winning the competition. If the number of partners is sufficiently large, then the joint investment of a network tends to be approximately half of that when the assets are completely recoverable ($\alpha = 0$).

Model Extension. We now extend our model to competitions involving more than two networks ($N > 2$).

PROPOSITION 2. If the investments are recoverable ($0 \leq \alpha < 1$), then the joint investment made by a network increases as the number of competing networks increases ($\frac{\partial X_i}{\partial N} > 0$). If the investments are nonrecoverable ($\alpha = 1$), however, the joint investment decreases as the number of competing networks increase ($\frac{\partial X_i}{\partial N} < 0$).

Proof: See Appendix.

The intuition for this result is as follows. As the number of competing networks increases, networks are encouraged to make more competitive investments, and this has a positive impact on the joint investment. At the same time, an increase in the number of competing networks decreases the probability of a network winning the competition. This reduction in the probability of winning has a negative impact on the joint investment, if the investments are nonrecoverable. Further, this negative impact dominates the positive impact, if the investments are nonrecoverable.

To appreciate the result, consider the situation where the investments are fully recoverable. We find that $X_i = \frac{\bar{v}_i}{\frac{1}{n_i} + N\left(\frac{1}{N-1}\right)}$, where $\bar{v}_i$ is the mean value that partners in network i place on winning the competition. Note that $\frac{\partial X_i}{\partial N} > 0$. Consistent with Proposition 1, the joint investment increases as the network size (n_i) increases, though the individual investment of a partner decreases. That is $\frac{\partial X_i}{\partial n_i} > 0$ but $\frac{\partial x_{ij}}{\partial n_i} < 0$.

If investments are completely nonrecoverable, then we find that $\frac{\partial X_i}{\partial N} < 0$. As in Proposition 1, we find that the joint investment of a network does not depend on its size ($\frac{\partial X_i}{\partial n_i} = 0$). But the investment of an individual partner in a network decreases, as the size of the network increases ($\frac{\partial x_{ij}}{\partial n_i} < 0$).

Discussion. The focus of the model and its equilibrium analysis is on the supply of a product by a network of firms. But a firm could well enjoy demand-side economies of scale when it becomes part of a network: positive consumption network externalities might arise as the number of consumers increases (e.g., Katz and Shapiro 1985). For instance, the utility derived from a good might increase as the number of other consumers who use the same product increases. In our formulation, one can accommodate potential consumption externalities by allowing the benefit obtained on winning the competition to increase with network size. That is, one can let $v_{ij}(n_i)$ be the benefit that partner j in network i earns on winning the competition, and $v'_{ij}(n_i) > 0$. Amaldoss and Rapoport (2003) show that allowing for consumption externality does not change the qualitative implications Proposition 1. They also show that Proposition 1 is robust to letting the market size be endogenous to the model.

3. EMPIRICAL INVESTIGATION

The major obstacle in testing the implications of Proposition 1 is insufficient field data on the strategic behavior of partners in networks (Gomes-Casseras 1994). It is difficult to obtain detailed information on the strategic behavior of network partners in a field setting. Hence, we decided to test the model in a laboratory.

It is not clear whether the model predictions can survive a laboratory test. Subjects are not expected to solve for the equilibrium and choose their investments accordingly. Rather, their decisions are likely to be guided by some simple heuristics. For instance, subjects might invest less as the number of partners in a network increases. If this under-investment problem is acute, then the joint investment is more likely to decrease as a network increases in size, contrary to the model predictions. Further, the investment decisions of subjects might not be sensitive to the recoverability level of investments as predicted by theory. It is an empirical question whether subjects can learn to behave in a fashion that is consistent with equilibrium play. The literature shows that sometimes they do and in other times they do not. Hence, we designed an experiment to test some of the more counterintuitive predictions. In particular, we tested how network size and type of investment influence the joint investments of new product development alliances.

3.1. Laboratory Test

Subjects. The subjects were mostly business school students recruited through advertisements and class announcements promising monetary reward contingent on their performance in a decision-making experiment. In addition to their earnings in the experiment, the subjects were paid a show-up fee of \$5. All transactions were in an experimental currency called "francs". At the end of the experiment, the cumulative payoffs were converted to US dollars. Each subject earned \$12–20.

Experimental Design. We conducted a 2×2 between-subjects factorial design with two levels of investment nonrecoverability, namely $\alpha \in \{0,1\}$, and two levels of network size, namely $n \in \{2,4\}$. The number of competing networks remained fixed in our experiment ($N = 2$). Sixteen subjects participated in each session for a total of $16 \times 4 = 64$ subjects.

Procedure. Upon arriving at the laboratory, the subjects were randomly seated in 16 computer booths. They were then asked to read the instructions (see Appendix). After reading the instructions, the subjects participated in five practice trials to familiarize themselves with the task. Questions asked during these practice trials were fully answered. Communication between the subjects during the course of the experiment was prohibited.

On each trial, the 16 subjects in the session were randomly divided into smaller sets of either two or four subjects depending on the network size. The network size varied across experimental sessions, but remained unchanged within a session. Each of these networks was set to compete with another network. The random assignment schedule ensured that each subject was networked on each trial with a different set of subjects. The subjects had no way of knowing the identity of their partners or competitors on any given trial. Therefore, reputation effects were minimized.

At the commencement of the experiment, each subject was informed of the network size. Subjects were also informed of the extent to which their investments could not be recovered in the event that their network lost the competition, $\alpha \in \{0,1\}$. Both the values of n and α remained unchanged throughout the session. At the beginning of each trial, each subject was endowed with $y_{ij} = 24$ francs in all the experimental conditions. The individual reward for winning the competition was also fixed throughout the experiment at $v_{ij} = 80$ francs.

Each subject had then to decide how much capital to invest in the product to be jointly developed by his/her network. Subjects could invest any amount including zero, provided the investment did not exceed the endowed capital, that is $0 \le x_{ij} \le y_{ij}$. After all the subjects made their investments privately and anonymously, the computer assessed the joint investments made by the two competing networks. The winning network was determined probabilistically. At the end of each trial, subjects were informed of the total investments made by the winning and losing networks, the probability of their network winning the competition, the winning network, their payoff for the trial, and their cumulative payoff. The subjects were provided with paper and pencil to help them record the outcomes of past decisions, if they wished to do so. The stage game was played repeatedly for 150 times, except in one treatment. Due to hardware problems, subjects in networks with two partners and recoverable investment completed only 120 trials. At the end of the experiment, subjects were paid according to their cumulative earnings in the experiment, debriefed, and dismissed. Each session lasted between 90 and 120 minutes.

Results. In presenting the results, we first discuss how closely the individual investments of network partners follow the qualitative and the point predictions of the model. Then, we compare the predicted and actual joint investments of networks.

Mean Investment of Individual Partner. Table 1 presents the mean individual investment made in a network along with the corresponding equilibrium prediction. The mean joint investments appear in parentheses.

Table 1 Mean Investment of Individual Partners and Networks (by Network Size and Type of Investment)

Type of Investment	*Network Size*	*Investment*	
		Actual Investment	*Equilibrium Prediction*
Recoverable Investment ($\alpha = 0$)	$n = 2$	19.041 (38.082)	16 (32)
	$n = 4$	11.786 (47.145)	8.889 (35.556)
Nonrecoverable Investment ($\alpha = 1$)	$n = 2$	11.624 (23.248)	10 (20)
	$n = 4$	7.089 (28.357)	5 (20)

Note: The individual investment is the average investment computed across the 16 subjects in each cell. The joint investment of a network is indicated within *parentheses.*

Qualitative Predictions. The equilibrium solution implies two qualitative predictions. First, as the network size increases individual investments should decrease. Second, individual partners should invest more, as the recoverability of their investment increases. To test whether the experimental results support these predictions, we conducted a two-way ANOVA with two between-subject factors (network size and type of investment). For this analysis, we used the mean individual investments made by the 16 subjects in each of the 4 treatments.

Both hypotheses were supported. In particular, as predicted by theory, the main effect of network size was highly significant ($F_{(1,60)} = 22.74$, $p < 0.0001$). Table 1 shows that the mean individual investment of subjects decreased as their network size increased. On the average, partners in networks with two and four members invested 19.041 and 11.786 francs respectively, when the investments were recoverable. Similarly, partners in networks with two and four members invested 11.624, and 7.089 francs, respectively, when the investments were nonrecoverable. As predicted by theory, the main effect of type of investment was significant ($F_{(1,60)} = 24.01$, $p < 0.0001$), with subjects investing less when the investments were nonrecoverable.

Point Predictions. Table 1 shows that the mean investments of individual subjects are consistent with the point predictions of the model when investments are nonrecoverable. But the mean investments are in general higher than the point prediction when the investments are fully recoverable. Our model implies that risk-neutral individuals in networks with two partners should invest 10 francs if the investments are nonrecoverable. In actuality, the subjects invested on average 11.624 francs. We cannot reject the null hypothesis that the predicted and actual investments are same ($t = 1.12$, $p > 0.27$). Subjects in networks with four partners invested on the average 7.089 francs. This investment is not significantly different from the predicted 5 francs ($t = 1.38$, $p > 0.18$). However, when the investments were fully recoverable, subjects invested significantly more than the point predictions. In equilibrium, risk-neutral subjects in networks with two partners should invest 16 francs. In actuality, our subjects invested 19.041 francs ($t = 3.898$, $p < 0.01$). Subjects in networks with four partners invested 11.786 francs, which again is significantly higher than the predicted investment of 8.889 francs ($t = 3.617$, $p < 0.01$). It is useful to note that the tendency to overinvest is weaker in our experiments compared to those reported in the rent seeking literature (e.g., Milner and Pratt 1989 and 1991, Shogren and Baik 1991, Potters et al. 1999, Onculer and Croson 2003). Our subjects might have invested closer to the equilibrium prediction because we provided them greater opportunity to learn from experience, avoided some of the pitfalls in the earlier experimental designs, and finally the desire to free ride could have dampened the enthusiasm to invest more.[8] We will probe the behavioral consequences of greater opportunity to learn later in the section on learning.

Individual Differences. Figure 1A displays the frequency distribution of the mean individual investment when the investments are recoverable. The mean individual investment ranges all the way from 2.8 to 23.8 francs, exhibiting considerable individual differences not accounted for by the equilibrium solution. Although there are substantial individual differences in the mean investments, the distribution of mean investments is, in general, in keeping with the theory: Subjects in networks with two partners tend to invest more than those in networks with four partners. Figure 1B exhibits the frequency distribution of the mean individual investment when the investments were nonrecoverable. Again, we notice considerable differences among subjects. As before, subjects in networks with two partners invested more than subjects in networks with four partners.

Mean Joint Investment of Networks. Table 1 reports (in parentheses) the mean joint investments of networks. In equilibrium, the joint investment of networks should increase as network size increases, if the investments are fully recoverable. But if the investments are not recoverable, then the joint investment should not be affected by network size. To test for the effect of network size, we divided the first 120 trials into 12 blocks of 10 trials, and then conducted an ANOVA with the joint investments as the dependent variable.

Our subjects increased their joint investments, as predicted by theory, when the investments were recoverable. The main effect of network size was significant ($F_{(1,22)} = 68.19$, $p < 0.0001$). Theory predicts that joint investments should not be affected by network size, when the investments were not recoverable. We note that

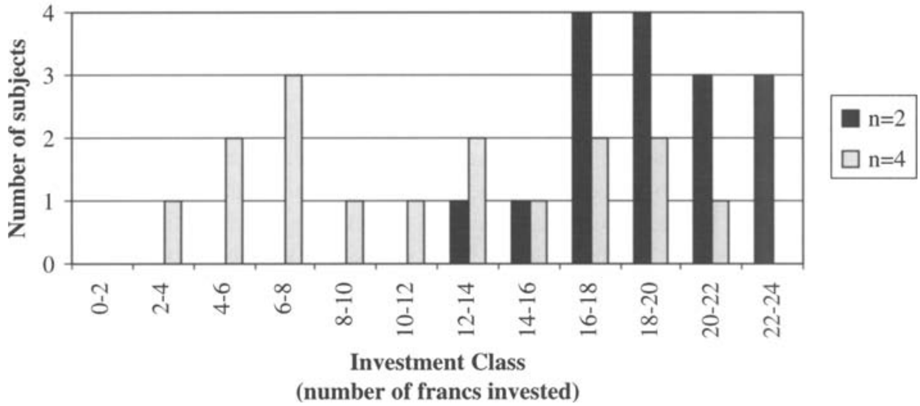


Figure 1A Recoverable Investment
Distribution of the Mean Investments of Subjects
Note: The equilibrium prediction for networks with two and four partners is 16 and 8.889 francs, respectively.

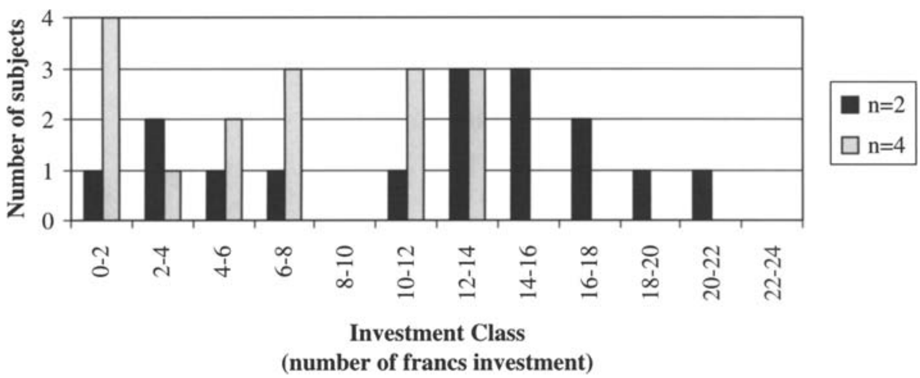


Figure 1B Nonrecoverable Investment
Distribution of the Mean Investments of Subjects
Note: The equilibrium prediction for networks with two and four partners is 10 and 5 francs, respectively.

the effect of network size was not significant ($F_{(1,22)} = 5.35$, $p = 0.03$). On probing further, we find that the effect of network size is marginally significant in the first three quarters of the trials of the game ($p < 0.07$), but not significant in the last quarter of the experiment ($p > 0.40$). Hence, the results show that individuals in networks with nonrecoverable investments moved toward equilibrium behavior in the last three blocks of trials. This finding calls for a more detailed analysis of the learning process.

4. ADAPTIVE LEARNING

To detect trends in investment across trials, we divided the trials into 12 blocks of 10 trials and then tested for block effects. We performed an ANOVA with two between-subject factors (Network Size and Type of Investment) and one within-subject factor (Block) with repeated measures. The main effect of block was significant, implying that subjects changed their mean investments across trials ($F_{(11,660)}$ = 4.97, $p < 0.0001$). Further, the interaction effect of block and type of investment was significant, suggesting that the changes in investments over blocks varied with the type of investment ($F_{(11,660)}$ = 3.02, $p < 0.001$). On probing further, we find that the block effect was significant at all networks except in networks with four partners and recoverable investment.[9] Figure 2 presents the trends in the investment patterns.

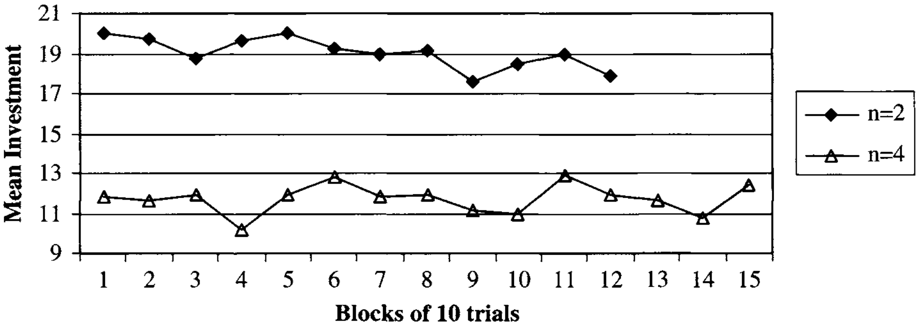


Figure 2A Recoverable Investment
Trends in the Mean Investment of Individual Partners
Note: The equilibrium prediction for networks with two and four partners is 16 and 8.889 francs, respectively. For $n = 2$, we collected data for only 120 trials.

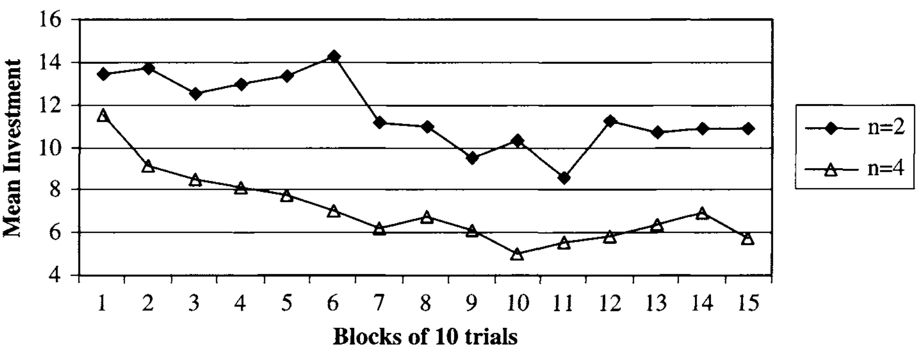


Figure 2B Nonrecoverable Investment
Trends in the Mean Investment of Individual Partners
Note: The equilibrium prediction for networks with two and four partners is 10 and 5 francs, respectively.

The trend analysis shows that subjects *learned* to play the game over the multiple iterations of the one-period game. Learning in games by subjects and learning in field by network managers are probably guided by different processes and based on different information sources. However, in both instances past experience is likely to inform future decisions. In the laboratory, subjects can potentially identify strategies that worked well in the past, and also evaluate new strategies based on their knowledge of the past behavior of competitors and collaborators. In the field, managers can identify strategies that worked well in the past in the context of their firm as well as learn from the best practices of industry leaders.

The goal of this section is to determine whether the significant trends in the investments in three of the four conditions can be accounted for by adaptive learning mechanisms. Learning in games is a rapidly growing field of research, and several classes of learning models have been proposed. We focus on the "Experience-Weighted Attraction" (EWA) learning model of Camerer and Ho (1999) for two principal reasons:

1. The EWA model allows for three different learning effects: actual, simulation, and declining effects. The first principle, which is shared by all reinforcement-based models, states that success increases the choice probability of chosen strategy. The second principle asserts that unchosen strategies, which would have yielded high payoffs, are more likely to be chosen in the future. The third principle states that with experience, players move to reduce discrepancies between actual and foregone payoffs.
2. The EWA model has been shown to be successful in accounting for the behavior of subjects in varied experimental settings. It is reported that the hybrid EWA model outperforms reinforcement and belief learning in 31 data sets spanning a dozen different games (Camerer, Ho and Chong 2002, Camerer, Hsia, and Ho, 2002).

It is not our intent to compare the EWA model with alternative learning models as this comparison is beyond the scope of the present paper. Rather, our goal is to use the EWA model to account for the expenditure patterns reported in Section 3.

The EWA Learning Model. At the heart of the EWA model are two parameters that are updated after each round of play. The first parameter, denoted by $N(t)$, is interpreted as the number of observations-equivalents of past experience. The second parameter, denoted by $A_{ij}^{x_m}(t)$, is interpreted as the attraction of investing x_m units of capital by player j in network i after round t. The initial values of these two parameters are denoted by $N(0)$ and $A_{ij}^{x_m}(0)$. For updating $N(t)$, the model assumes that $N(t) = \rho N(t-1) + 1$, $t \geq 1$, where the parameter $\rho (0 \leq \rho \leq 1)$ is the rate of depreciation. While updating the attraction of a strategy, expected payoffs corresponding to unchosen strategies are given a weight of δ, whereas the payoffs pertaining to chosen strategies are given an additional weight of $1 - \delta$. The parameter δ measures the relative weight given to foregone payoffs, compared to actual payoffs, in updating attractions. It can be interpreted as a kind of "imagination" of foregone payoffs. Previous attractions are depreciated by another parameter $\phi(0 \le \phi \le 1)$. The decay rate ϕ reflects a combination of forgetting and the degree to which players recognize that other players are adapting and thereby place lower weight on the history of the game. If ϕ is lower, players discard old observations more quickly, and are responsive to the most recent observations.

Earlier we denoted the investment of player j in network i by x_{ij}. Our subjects could invest any amount ranging from 0 to 24 francs. To be consistent with Camerer and Ho (1999), we denote strategy of investing $x_{ij} = m$ by $s_{ij}^{x_m}$. In order to make the strategy space discrete, we rounded the investments to the closest integer, and thus $s_{ij}^{x_m} \in \{0, 1, 2 \ldots 24\}$. The attraction of strategy $s_{ij}^{x_m}$, namely $A_{ij}^{x_m}(t)$, is a weighted average of the payoff for period t and the previous attraction $A_{ij}^{x_m}(t-1)$:

$$A_{ij}^{x_m}(t) = \frac{\phi N(t-1)A_{ij}^{x_m}(t-1) + [\delta + (1-\delta)I(S_{ij}^{x_m}, s_{ij}(t))]\pi_{ij}(S_{ij}^{x_m}, s_{-ij}(t))}{N(t)}.$$

In the expression above, $I(s_{ij}^{x_m}, s_{ij}(t))$ is an indicator variable that is equal to 1 if $s_{ij}^{x_m} = s_{ij}(t)$, or 0 if $s_{ij}^{x_m} \neq s_{ij}(t)$. Thus, if player j in network i plays strategy $s_{ij}^{x_m}$ on trial t, then the payoff $\pi_{ij}(s_{ij}^{x_m}, s_{-ij}(t))$ is added to the attraction of strategy x_m. But if partner j in network i did not play strategy $s_{ij}^{x_m}$ on trial t, then only $\delta.\pi_{ij}(s_{ij}^{x_m}, s_{-ij}(t))$ is added to the attraction of strategy x_m.

The probability of player j in network i investing x_m on trial $t+1$ is given by the function:

$$p_{ij}^{x_m}(t+1) = \frac{e^{\lambda A_{ij}^{x_m}(t)}}{\sum_{l=0}^{24} e^{\lambda A_{ij}^{x_h}(t)}},$$

where the parameter λ measures sensitivity of the players to attractions. The parameter λ can also be interpreted as a measure of noise in the strategy choice process. This concludes our brief description of the EWA model.

Results. The maximum likelihood method was used to estimate the model parameters. Table 2 reports the parameter estimates and goodness of fit statistics for the EWA model and its special cases – reinforcement-based and belief-based learning. We conducted a separate set of analyses for each session of the six treatments. The left panel reports the results for networks with recoverable investments, whereas the right panel reports the corresponding results for networks with nonrecoverable investment. The reported parameters are significant at $\alpha = 0.05$. We briefly summarize our results below.

Overall Model Fit. Table 2 presents the log-likelihood (*LL*), Akaike Information Criterion (*AIC*), Bayesian Information Criterion (*BIC*), pseuodo-R^2 (ρ^2), and χ^2 statistic

Table 2 Learning Model Parameter Estimates

Parameter	*Recoverable Investment*						*Nonrecoverable Investment*					
	Network size = 2			*Network size = 4*			*Network size = 2*			*Network size = 4*		
	EWA	*Reinforcement-based*	*Belief-based*	*EWA*	*Reinforcement-based*	*Belief-based*	*EWA*	*Reinforcement-based*	*Belief-based*	*EWA*	*Reinforcement-based*	*Belief-based*
δ	0.251	0.000	1.000	0.000	0.000	1.000	0.000	0.000	1.000	0.282	0.000	1.000
ρ	0.063	0.000	0.480	0.000	0.000	0.499	0.542	0.000	0.485	0.000	0.000	0.484
ϕ	0.843	0.895	0.480	0.986	0.986	0.499	0.396	0.864	0.485	0.888	0.874	0.484
λ	0.014	0.013	0.019	0.003	0.003	0.003	0.016	0.017	0.016	0.016	0.015	0.016
A0(0)	226.776	672.564	140.731	692.424	692.424	672.011	44.002	196.530	42.586	116.456	117.638	30.735^{ns}
A1(0)	0.000	0.000	0.000	0.000	0.000	0.000	51.744	130.933	39.748	0.000	0.000	0.000
A2(0)	0.000	0.000	0.000	238.298	238.298	223.825	0.000	0.000	0.000	28.771^{ns}	0.000	0.000
A3(0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A4(0)	0.000	0.000	0.000	489.274	489.274	474.648	0.000	0.000	0.000	131.453	134.331	0.000
A5(0)	173.474	551.485	0.000	423.102	423.102	383.661	0.000	0.000	0.000	30.498^{ns}	0.000	0.000
A6(0)	0.000	0.000	0.000	333.977	333.977	307.845	0.000	0.000	0.000	148.298	155.965	0.000
A7(0)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
A8(0)	143.820	482.647	0.000	422.833	422.833	394.184	0.000	0.000	0.000	147.141	139.778	93.104
A9(0)	137.089	462.242	0.000	407.097	407.097	376.591	0.000	0.000	0.000	0.000	0.000	0.000
A10(0)	171.092	606.808	0.000	644.950	644.950	588.940	0.000	0.000	0.000	151.135	151.139	115.797
A11(0)	172.688	555.214	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2 {cot'd)

A12(0)	194.341	612.179	0.000	283.244	283.244	274.082	40.117	138.873	30.443	180.625	173.279	135.436
A13(0)	0.000	0.000	0.000	0.000	0.000	0.000	10.969^{ns}	69.561	14.400^{ns}	0.000	0.000	0.000
A14(0)	169.718	583.006	0.000	291.024	291.024	286.682	0.000	0.000	0.000	109.592	131.735	0.000
A15(0)	207.963	641.871	89.993	366.232	366.232	361.737	47.958	140.789	38.320	111.806	115.393	91.407
A16(0)	203.759	633.714	0.000	0.000	0.000	0.000	25.854	39.697	23.078^{ns}	0.000	0.000	0.000
A17(0)	224.977	710.796	126.975	0.000	0.000	0.000	7.729^{ns}	45.636	21.750^{ns}	0.000	0.000	0.000
A18(0)	246.961	746.356	157.213	0.000	0.000	0.000	45.572	150.587	37.969	84.138	72.454	70.912
A19(0)	220.896	672.797	108.963	0.000	0.000	0.000	23.041	33.854^{ns}	20.570^{ns}	0.000	0.000	0.000
A20(0)	252.194	767.886	176.690	499.768	499.768	514.933	44.128	169.641	33.898	164.705	145.826	156.075
A21(0)	251.445	755.121	156.342	0.000	0.000	0.000	7.487^{ns}	0.000	21.853^{ns}	0.000	0.000	0.000
A22(0)	240.428	744.265	144.100	0.000	0.000	0.000	32.418	98.210	29.783	0.000	0.000	0.000
A23(0)	171.960	542.068	0.000	0.000	0.000	0.000	27.378	114.137	36.707	0.000	0.000	0.000
A24(0)	239.932	717.239	140.846	711.631	711.631	735.682	67.106	182.073	57.542	126.110	126.017	130.512
N(0)	2.360	1.000	1.899	1.000	1.000	2.064	4.432	1.000	4.483	0.987	1.000	1.882
Log-Likelihood	−4494.713	−4496.528	−4627.354	−6569.245	−6569.245	−6740.055	−4654.758	−4667.738	−5174.321	−4191.733	−4200.801	−4384.963
AIC	−4524.713	−4523.528	−4655.354	−6599.245	−6596.245	−6768.055	−4684.758	−4694.738	−5202.321	−4221.733	−4227.801	−4412.963
BIC	−4608.114	−4598.589	−4733.195	−6685.993	−6674.318	−6849.020	−4771.506	−4772.812	−5283.286	−4308.481	−4305.875	−4493.929
ρ^2	0.273	0.272	0.251	0.150	0.150	0.128	0.397	0.396	0.330	0.457	0.456	0.432
χ^2		3.632	265.282		0.000	341.620		25.812	1038.978		18.138	386.462
(p-value, d.o.f)		(0.304,3)	(0.000,2)		(1.000,3)	(0.000,2)		(0.000,3)	(0.000,2)		(0.000,3)	(0.000,2)

Note: n.s. = not significant at alpha = 0.05.

for model comparisons.[10] The hybrid EWA model outperforms its special cases – the reinforcement and belief models – in tracking the investment behavior of the subjects in three of the four experimental sessions. For brevity, hereafter our discussion will focus exclusively on EWA model and not its special cases.

To assess the predictive accuracy of the EWA model, we compared the actual mean individual investment against the EWA prediction. The EWA predictions were obtained by averaging the predicted investment of the 16 subjects in each experimental session. When the investments were nonrecoverable, subjects in networks with two and four partners invested 11.624 and 7.089 francs (see Table 1). According to the EWA model, these subjects should have invested 13.964, and 7.885 francs. We cannot reject the null hypothesis that the EWA prediction and actual investment are the same for networks with two ($t = 1.12, p > 0.2$) and four ($t = 0.52$, $p > 0.5$) partners. When the investments are recoverable, subjects in networks with two and four partners actually invested 19.041 and 11.786 francs (see Table 1), but the corresponding EWA predictions are 16.602 and 18.737 francs. The difference between the actual and EWA predicted investments is not significant in networks with two partners, but significant in networks with four partners (n = 2: t = 1.6, $p > 0.10$; n = 4: $t = 3.4$, $p < 0.01$).

Interpretation of the Parameter Values. If the estimated value of δ is zero, then we can infer that the choice of strategies in these games was not influenced by expected payoffs. Rather, the investment decisions were based on the payoffs earned in previous trials. We observe that $\delta = 0$ in three of the four sessions. Clearly, we can reject the hypothesis that the strategy choices of our subjects were *only* influenced by their beliefs formed by observing the history of other's choices. This inference is validated by the Akaike Information Criterion (*AIC*), Bayesian Information Criterion (*BIC*), pseuodo-R^2 (ρ^2), and χ^2 statistic reported for model comparisons, which suggest that a belief-based model fits the data poorly.

If $\phi = \rho$, then the model implies that the strategy choices of our subjects were guided by the average performance of past strategies. But if $\phi > \rho$, then it implies that the investment decisions were influenced by a weighted average of cumulative performance and average performance of past strategies. We observe that in three of the four cases $\phi > \rho$.

The parameter λ measures the level of noise in the choice of strategies. Our estimates fall in the range $0.003 \leq \lambda \leq 0.016$, implying that the learning process was noisy. The level of noise is higher than that reported in Rapoport and Amaldoss (2000). But our experiments differ from this earlier study in an important way – in our inter-network competitions the winner was determined probabilistically, whereas in the earlier study the winner was determined determinstically. Hence, feedback to our subjects on past actions is inherently more noisy, and this probably accounts for the lower values of parameter λ.

Pre-game Disposition. The estimates of the parameters $A^0(0)$, $A^1(0)$, . . . , $A^{24}(0)$ indicate the predisposition of the subjects to invest 0, 1, . . . , 24 francs, respectively, before the first trial of the experiment. These pre-game disposition parameters are reported in the Table 2. We observe that the predisposition of our subjects is not sharply focused around the equilibrium solution, implying that our subjects were not able to arrive at these equilibrium solutions through sheer introspection. However, subjects in networks with two partners were predisposed to invest more than subjects in networks with four partners. We draw a similar pattern of inferences for networks where the investments were nonrecoverable. In general, the strength of pre-game disposition as evidenced in the empirical estimates of $N(0)$ is less than five observations-equivalent in all the four sessions. This implies that the predisposition of our subjects was weak, though qualitatively consistent with the model predictions.

Discussion. Our analysis suggests that previously received payoffs, rather than foregone payoffs, guided the investment decisions of our subjects. Further, the predisposition of our subjects, though not sharply defined, is qualitatively consistent with the equilibrium prediction. In general, the EWA model accounts for the major behavioral regularities observed in our experiments, though it tends to over-predict actual investments. The adaptive learning analysis suggests that naïve players can learn from experience, and move toward equilibrium behavior without any knowledge of the equilibrium solution. This raises hope that financially motivated agents in networks might potentially learn from experience and slowly begin to behave in a fashion that is consistent with theory.

5. SUMMARY AND CONCLUSIONS

Our experimental research was motivated by a desire to better understand the emerging phenomenon of networks, rather than individual firms, developing new products. In contrast to common belief of alliance managers, Amaldoss and Rapoport (2003) show that the joint investment of network partners does not decrease as a network grows in size. Specifically, if the investments are recoverable the joint investment should increase as the network size increases. But if they are not, then joint investment should not change with network size. On extending the theoretical model to investigate competition among a large number of networks ($N > 2$), we find that the effect of number of competing networks on joint investment depends on whether or not the investments are recoverable. If the investments are recoverable, it exerts a positive effect, but if they are not, it has a negative impact.

We conducted a laboratory experiment primarily because of the absence of reliable empirical data. A major advantage of the experimental method is the precise implementation of the model's assumptions, and an opportunity to closely scrutinize the decisions of the players. Yet it is useful to be careful in generalizing these results to field settings. Additional experimental investigation is required to replicate the present findings. Nevertheless, certain consistent patterns of behavior have emerged that support our claim about the effect of network size and type of investment on the joint investment of network partners. We find that joint investment increases as network size increases when investment is recoverable. But joint investment does not change significantly with increase in network size when investments are nonrecoverable. We also note that there is a trend toward equilibrium behavior over multiple iterations of the stage game even when reputation effects are minimized, and that the EWA model tracks the investment patterns of the subjects.

It is paradoxical to observe that managers have several apprehensions about joint development of products while the incidence of these collaborations is surging ahead. For instance, collaborations already account for 6–15% of the market value of typical firms, and it is expected to grow to 16–25% of the company value in five years, according to the consulting firm Accenture (Kalmbach and Roussel 1999). Our experimental investigations help allay some of the misapprehensions of jointly developing new products. First, the under-investment problems associated with networks might decline, as the recoverability of investment increases. Second, the incremental inputs from a new partner might outweigh the potential free-riding problem in networks, where the investments are partly recoverable.

ACKNOWLEDGMENT

This research has been supported by grants from the National Science Foundation (NSF grant SES-089182/SES-0114138).

NOTES

1 Bensaou and Anderson (1999) provide empirical evidence on the motivation for making idiosyncratic investments in buyer-supplier relations, and Jap (1999) studies the collaboration process of buyer-supplier dyads. Amaldoss et al. (2000) investigate the effect of profit-sharing arrangement and type of alliance on the investment behavior of *two*-member alliances. Wathne and Heide (2000) outline active and passive opportunistic behavior in interfirm relationships, discuss governance strategies that can potentially handle such behavior.

2 We refer to alliances with more than two partners as *networks*.

3 Mips Computer Systems, a 1987 semiconductor start-up, developed a network of semiconductor partners, so that it could challenge firms like Intel. Partners in the Mips network included among others NEC, Sony, DEC, Daewoo, AT&T, Nixdorf, Silicon Graphics, Toshiba, Olivetti, Kuboto. The Mips network promoted the Advanced Computing Environment (ACE). However, the ACE initiative died within a year, and Mips Computer system was acquired soon by Silicon Graphics. It is reported that the number of partners in the Mips network was a potential cause for its downfall (see Gomes-Casseras 1994, page 7). However, the PowerPC triad – IBM, Motorola, and Apple – was successful in developing a powerful microprocessor.

4 Sometimes being part of a larger network might confer an additional benefit: positive consumption network externalities might arise as the number of consumers increases (e.g., Katz and Shapiro 1985). Later we discuss the implications of consumption externalities in the context of our model.

5 Some networks involve equity ownership and shared managerial control or the creation of a new organization (e.g., GM-Suzuki joint venture). Others do not involve any equity arrangements, and each partner in the network earns profit from selling its part of the technology platform (e.g., Wintel). These independent profit maximizing firms are linked together only by the common platform or purpose they

share. It is estimated that non-equity arrangements account for 50% of all collaborative arrangements across all industries (Zagnoli 1987).

[6] In general, there are three broad classes of process by which the winner could be determined: auctions (e.g., Dasgupta and Stiglitz 1980, Gilbert and Newberry 1982, Katz and Shapiro 1985), stochastic racing (e.g., Lee and Wilde 1980, Reinganum 1981 and 1982), and contest models (e.g., Tullock 1967, Hartwick 1982, and Rogerson 1982). The auction models are deterministic and focus only on strategic uncertainty. The stochastic racing models and contest models allow for both outcome and strategic uncertainty.

[7] When $R > 1$, an increase in investment induces more than proportionate rise in the probability of winning the competition. When $R = \infty$, if a network invests ε more than its competitors then it will win the competition.

[8] Some of the common pitfalls are using a sequential, rather than simultaneous, decision making (e.g., Millner and Pratt 1989), using fixed pairing instead of random pairing (see Shogren and Baik 1991), placing binding budget constraints (e.g., Davis and Reilly 1998).

[9] When the investments are recoverable the block effect is significant only in networks with two partners ($n = 2$: $F_{(11,1800)} = 2.43$, $p < 0.01$; $n = 4$: $F_{(11,1800)} = 1.56$, $p = 0.103$). When investments were nonrecoverable, the effect of block was significant in all the networks ($n = 2$: $F_{(11,1800)} = 6.99$, $p < 0.0001$; $n = 4$: $F_{(11,1800)} = 14.08$, $p < 0.0001$).

[10] $AIC = LL - k$, and $BIC = LL - (k/2) \times log(M)$, where k is the number of degrees of freedom and M is the sample size. The pseudo-R^2 (ρ^2) is the difference between the AIC measure and the log-likelihood of a model of random choices, normalized by the random model log-likelihood.

REFERENCES

Achrol, R. and P. Kotler (1999). "Marketing in the Network Economy." *Journal of Marketing*, 63, 146–163.

Amaldoss, W., R. Meyer, J. Raju, and A. Rapoport (2000). "Collaborating to Compete." *Marketing Science*, 19, 2, 105–126.

Amaldoss, W. and A. Rapoport (2003). "An Analysis of Inter-network Competition." Working Paper, Department of Marketing, Duke University.

Bensaou, M. and E. Anderson (1999). "Buyer-Supplier Relations in Industrial Markets: When Do Buyers Risk Making Idiosyncratic Investments?" *Organization Science*, 10, 460–481.

Bleeke, J., and D. Ernst (1991). "The Way to Win Cross Border Alliances." *Harvard Business Review*, 69, 127–135.

Camerer, C., and T. Ho (1999). "Experience-weighted Attraction Learning in Normal Form Games." *Econometrica*, 87, 827–874.

Camerer, C., T. Ho, and J. Chong (2002). "Experience-weighted Attraction Learning and Strategic Teaching Games." *Journal of Economic Theory*, 104, 137–188.

Camerer, C., D. Hsia, and T. Ho (2002). "EWA learning in bilateral call markets." In R. Zwick and A. Rapoport (Eds.). *Experimental Business Research*. New York: Kluwer.

Davis, D. D. and Reilly, R. J. (1998). "Do too many cooks always spoil the stew? An experimental analysis of rent-seeking and the role of strategic buyer." *Public Choice*, 95, 89–115.

Dasgupta, P., and J. Stiglitz (1980). "Uncertainty, Industrial Structure, and the Speed of R&D." *Bell Journal of Economics*, 11, 1–28.

Dutta, S., and A. M. Weiss (1997). "The Relationship Between a Firm's Level of Technological Innovativeness and Its Pattern of Partnership Agreements." *Management Science*, 43, 343–356.

Gilbert, R. J., and D. M. Newbery (1982). "Preemptive Patenting and the Persistence of Monopoly." *American Economic Review*, 72, 514–526.

Gomes-Casseres, B. (1994). "Group Versus Group: How Alliance Networks Compete." *Harvard Business Review*, 72, 62–74.

Gulati, R. (1998). "Alliances and Networks." *Strategic Management Journal*, 19, 293–317.

Hartwick, J. M. (1982). "Efficient Prizes in Prototype Development Contests." *Economic Letters*, 10, 375–379.

Jap, S. D. (1999). "Pie-Expansion Efforts: Collaboration Processes in Buyer-Supplier Relationships." *Journal of Marketing Research*, 36, 461–475.

Kalmbach, C. Jr., and C. Roussel (1999). Dispelling the Myths of Alliances, Outlook, Special edition published by Accenture, October, 1–57.

Katz, M. L., and C. Shapiro (1985). "Network Externalities, Competition and Compatibility." *American Economic Review*, 75, 424–440.

Kogut, B. (1988). "Joint Ventures: Theoretical and Empirical Perspectives." *Strategic Management Journal*, 9, 319–332.

Lee, T., and L. L. Wilde (1980). "Market Structure and Innovation: A Reformulation." *Quarterly Journal of Economics*, 94, 429–436.

Millner E. L. and Pratt, M. D. (1989). "An experimental investigation of efficient rent seeking." *Public Choice*, 62, 139–151.

Millner, E. L. and Pratt, M. D. (1991). "Risk aversion and rent seeking: An extension and some experimental evidence." *Public Choice*, 69, 81–92.

Nitzan, S. (1994). "Modelling rent-seeking contests." *European Journal of Political Economy*, 10, 41–60.

Nohria, N. (1992). "Introduction: Is the Network Perspective a Useful Way of Studying Organizations?" In N. Nohria and R. Eccles (eds.), *Networks and Organizations: Structure, Form and Action*, Harvard Business School Press, Boston MA.

Onculer, A. and R. Croson (2003). Rent-Seeking for Risky Rent: A Model and Experimental Investigation, Working Paper, INSEAD, France.

Pisano, G. P., M. V. Russo, and D. J. Teece (1988). "Joint Ventures and Collaborative Agreements in the Telecommunications Equipment Industry." In D. Mowery (Ed.), *International Collaborative Ventures in U.S. Manufacturing*. Ballinger, Cambridge, MA, pp. 23–70.

Potters, J., de Vries, C., and van Winden, G. (1999). "An experimental examination of rational rent-seeking." *European Journal of Political Economy*, 14, 783–800.

Rapoport, A., and W. Amaldoss (2000). "Mixed strategies and iterative elimination of strongly dominated strategies: an experimental investigation of states of knowledge." *Journal of Economic Behavior & Organization*, 42, 483–521.

Reinganum, J. F. (1981). "Dynamic Games of Innovation." *Journal of Economic Theory*, 25, 21–41.

Reinganum, J. F. (1982). "A Dynamic Game of R&D: Patent Protection and Competitive Behavior." *Econometrica*, 50, 671–688.

Rindfleisch, A., and C. Moorman (2001). "The Acquisition and Utilization of Information in New Product Alliances: A strength of Ties Perspective." *Journal of Marketing*, 65 (April), 1–18.

Robertson T. S., and H. Gatignon (1998). "Technology Development Mode: A Transaction Cost Conceptualization." *Strategic Management Journal*, 19, 515–531.

Rogerson, W. P. (1982). "The Social Costs of Regulation and Monopoly: A Game-Theoretic Analysis." *Bell Journal of Economics*, 13, 391–401.

Shogren, J. F. and Baik, K. H. (1991). "Reexamining efficient rent seeking in laboratory markets." *Public Choice*, 69, 69–79.

Tullock, G. (1967). "The welfare costs of tariffs, monopolies and theft." *Western Economic Journal*, 5, 224–232.

Tullock, G. (1980). "Efficient rent-seeking." In J. M. Buchanan, R. D. Tollison, and G. Tullock (Eds.), *Toward a Theory of the Rent-Seeking Society*, College Station, TX: Texas A&M University Press. pp. 97 - 112.

Wathne, K. H., and J. B. Heide. (2000). "Opportunism in Interfirm Relationships: Forms. Outcomes, and Solutions." *Journal of Marketing*, 64, 36–51.

Williamson, O. E. (1983). "Credible Commitments: Using Hostages to Support Exchange." *American Economic Review*, 73, 519–540.

Zagnoli, P. (1987). "Inter-firm Agreements as Bilateral Transactions," Paper delivered at the Conference on "New Technology and New Intermediaries: Competition, Intervention and Cooperation in Europe, America, and Asia," Center for European Studies, Stanford University, June 4–6.

APPENDIX

Section 1.1: Effect of Competition among N networks

Proposition 2: *a) If the investments are recoverable, then the joint investment made by a network increases as the number of competing networks increases* ($\frac{\partial X_i}{\partial N} > 0$). *b) But if the investments are nonrecoverable then the joint investment decreases as the number of competing networks increase* ($\frac{\partial X_i}{\partial N} < 0$).

Outline of the proof. We prove in Lemma 1 that $X_i = \dfrac{\sum_{j=1}^{n_i} v_{ij}}{1 - \alpha + n_i N\left(\alpha + \frac{1}{N-1}\right)}$ $\forall \alpha \in [0, 1)$. Then in Lemma 2 we show that $X_i = \dfrac{v}{N\left(\frac{1}{N-1}\right)}$, if $\alpha = 1$. Later using these Lemmas, we prove the two claims in the proposition.

Lemma 1: $X_i = \dfrac{\sum_{j=1}^{n_i} v_{ij}}{1 - \alpha + n_i N\left(\alpha + \frac{1}{N-1}\right)}$ $\forall \alpha \in [0, 1)$

Proof: The joint investment of network i is, as given by

$$X_i = \sum_{j=1}^{n_i} x_{ij}. \qquad \text{A1}$$

The total investment made by all the N networks engaged in the competition is

$$X = \sum_{i=1}^{N} X_i. \qquad \text{A2}$$

Now, the probability of network i winning the inter-network competition is

$$\Pr(i) = \frac{X_i}{X}. \qquad \text{A3}$$

The payoff partner 1 in network i earns by investing x_{i1} is:

$$\pi_{i1} = y_{i1} - x_{i1} + \Pr(i).v_{i1} + (1 - \alpha)x_{1i}(1 - \Pr(i)). \qquad \text{A4}$$

First, we replace X_i with $X_{i1} + \sum_{j=2}^{n_i} x_{ij}$ and differentiate π_{i1} with respect to x_{i1}. Then, we set it to zero ($\frac{\partial \pi_{i1}}{\partial x_{i1}} = 0$) and solve for x_{i1}. By summing the investments of all the partners in network i we obtain X_i.

$$\frac{\partial \pi_{i1}}{\partial x_{i1}} = 0 \qquad \text{A5}$$

$$-1 - \frac{v_{i1}X_i}{X^2} + \frac{v_{i1}}{X} + (1-\alpha)x_{i1}\left(\frac{X_i}{X^2} - \frac{1}{X}\right) + (1-\alpha)\left(1 - \frac{X_i}{X}\right) = 0 \qquad \text{A6}$$

$$x_{i1} = \frac{v_{i1}}{1-\alpha} - \frac{X}{1-\alpha}\left(\alpha + \frac{X_i}{X - X_i}\right). \qquad \text{A7}$$

Summing the investments of all the n_i partners in network i, we get

$$X_i = \frac{\sum_{i=1}^{n_i} v_{i1}}{1-\alpha} - \frac{n_i X}{1-\alpha}\left(\alpha + \frac{X_i}{X - X_i}\right). \qquad \text{A8}$$

As the networks are symmetric, the aggregate investment of all the N networks is $X = NX_i$. Substituting this in the above expression, we have:

$$X_i = \frac{\sum_{j=1}^{n_i} v_{ij}}{1-\alpha} - \frac{n_i N X_i}{1-\alpha}\left(\alpha + \frac{1}{N-1}\right) \qquad \text{A9}$$

$$X_i\left(1 - \alpha + n_i N\left(\alpha + \frac{1}{N-1}\right)\right) = \sum_{j=1}^{n_i} v_{ij} \qquad \text{A10}$$

Hence,

$$X_i = \frac{\sum_{j=1}^{n_i} v_{ij}}{1 - \alpha + n_i N\left(\alpha + \frac{1}{N-1}\right)}. \qquad \text{A11}$$

Lemma 2: $X_i = \dfrac{v}{N\left(\dfrac{1}{N-1}\right)}$, if $\alpha = 1$.

Proof: The joint investment of network i (A1), the total investment made by all the N networks engaged in the competition (A2), the probability of network i winning the competition (A3). Setting $v_{ij} = v$ and $\alpha = 1$ in A4, we obtain the expected payoff for partner 1 in network i on x_{i1}.

$$\pi_{i1} = y_{i1} - x_{i1} + \Pr(i).v \qquad \text{A12}$$

We differentiate π_{i1} with respect to x_{i1}, set it to zero $(\frac{\partial \pi_{i1}}{\partial x_{i1}} = 0)$, and then solve for X_i:

$$\frac{\partial \pi_{i1}}{\partial x_{i1}} = 0 \qquad \text{A13}$$

$$-1 - \frac{vX_i}{X^2} + \frac{v}{X} = 0. \qquad \text{A14}$$

As the networks are symmetric, the aggregate investment of all the N networks is $X = NX_i$. Substituting this in A14 we get

$$-1 - \frac{vX_i}{(NX_i)^2} + \frac{v}{NX_i} = 0. \qquad \text{A15}$$

On solving for X_i, we obtain

$$X_i = \frac{v}{N\left(1 + \dfrac{1}{N-1}\right)}. \qquad \text{A16}$$

Claim (a) of the Proposition 2: If the investments are recoverable, then the joint investment made by a network increases as the number of competing networks (N) increases: $\dfrac{\partial X}{\partial N} > 0$.

Proof: Setting $\alpha = 0$ in A11, we obtain:

$$X_i = \frac{\sum_{j=1}^{n_i} v_{ij}}{1 + n_i N\left(\dfrac{1}{N-1}\right)}. \qquad \text{A17}$$

$$X_i = \frac{\bar{v}_i}{\frac{1}{n_i} + N\left(\frac{1}{N-1}\right)} \qquad \text{A18}$$

$$\frac{\partial X_i}{\partial N} = \frac{n\bar{v}_i(N-1)}{N + n_i N - 1}. \qquad \text{A19}$$

Hence,

$$\frac{\partial X_i}{\partial N} > 0, \text{ as } N > 2. \qquad \text{A20}$$

This proves claim (a).

Claim (b) of the Proposition 2. If the investments are nonrecoverable, then the joint investment made by a network decreases as the number of competing networks (N) increases: $\frac{\partial X}{\partial N} < 0$.

Proof: We have from Lemma 2 (A16):

$$X_i = \frac{v}{N\left(1 + \frac{1}{N-1}\right)}.$$

$$\frac{\partial X_i}{\partial N} = \frac{v(N-2)}{N^3} \qquad \text{A21}$$

Therefore,

$$\frac{\partial X_i}{\partial N} < 0, \text{ as } N > 2. \qquad \text{A22}$$

This proves claim (b).

Section 1.2: Instructions for Subjects

We have included in this appendix the instructions for two treatments: one where the investments are nonrecoverable ($\alpha = 1$), the other where investments are recoverable ($\alpha = 0$).

Instructions for network size = 2, and nonrecoverable investment

"You will participate today in a decision making experiment concerning competition between two networks. Each network is comprised of two firms. You will be asked to make many decisions in the course of this decision making experiment, and you will be paid depending on your decisions and the decisions of other subjects. A research foundation has contributed funds to support this research.

In the present experiment, you are asked to represent a partner in the network. Three other subjects participate in this experiment. One represents the firm that joins you in the network for jointly developing a product, whereas the other two subjects represent firms in the competing network.

The experiment involves many trials. At the beginning of each trial, each subject will be provided with some investment capital and then asked how much of it he/she wishes to invest in developing the new product of his/her network. The four subjects will be provided with the same amount of investment capital, and asked to make independent investment decisions.

Rules of the Investment Game. The rules of this investment game are quite simple.

1) The winner of the competition between the networks will be determined *in proportion* to the *total investments* of the two networks. In particular, the probability of your network winning the competition will be determined as follows:

$$\textit{probability of your network winning} = \frac{\textit{Total investment in your network's product}}{\textit{Total investment in your network's product} + \textit{Total investment in the competing network's product}}$$

2) The capital invested by partners in the winning network is consumed in developing the product and thus completely nonrecoverable. Members of the losing network can potentially recover some of their investments. In this game, the investments of the losing network are completely nonrecoverable (In other words, the non-recoverability factor for members of the losing network is 1.) Therefore, once invested the money is *lost* irrespective of the outcome of the competition.
3) Each member of the winning network will receive a fixed reward. The reward does not depend on the relative investments made by each member of the winning network. The fixed reward represents profit that each firm earns from producing and marketing the new product. Each member of the losing network receives nothing.

Experimental Procedure.
As discussed above, there are *two networks*, and each network is comprised of *two partners*. At the beginning of each trial each subject will be given the same investment capital, which will remain unchanged from trial to trial. The investment capital will be stated in terms of a laboratory currency called "francs". At the end of the experiment your earnings will be converted to US dollars.

Once each player is provided with the investment capital, he or she must decide how much to invest in developing his/her network's new product. You may invest any number of francs (including zero), provided your investment does not exceed your endowment (investment capital allotted for the trial). After all the four players have made their investment decisions, privately and anonymously, the computer will compute the total investment made by each of the competing networks. The probability that your network wins the competition will be computed from the equation shown above. Then the computer will *randomly* choose a number between 0 and 1 to determine the winning network. If this random number falls in the interval between 0 and the probability that your network wins the competition, then your network will be declared the winner. Otherwise, if the random number is greater than your network's probability of winning, the competing network will win the competition. Members of the winning network will receive a reward of known size (in francs), whereas members of the losing network will receive nothing.

Note that if you invest more you increase the probability of winning the competition. However, the same reasoning applies to all the firms. Moreover, investing more does not guarantee winning the competition, as the winning network is determined *probabilistically*.

Computation of Individual Payoffs. The individual payoffs for a trial will be computed as follows:

Payoff to a member of the *winning network* = endowment for the trial
– investment made by the firm in the trial + reward for winning the competition.

Payoff to a member of the *losing network* = endowment for the trial
– (non-recoverability × investment made by the firm in the trial).

Outcome Information. At the end of each trial, the computer will display the following information on the computer screens:

1) The total investments made by the winning and losing networks,
2) The network winning the competition,
3) Your payoff for the trial,
4) Your cumulative earnings.

It is important to note that *only you know your investment decision.* This decision is made anonymously.

Network Membership. Network membership will vary randomly from trial to trial. *On each trial you will be paired with a different person in this room, and both of you will compete as a network against another new network of two players.* Therefore, you will never know the identity of your partner on any give trial.

The following two examples are provided to help you understand how your payoff is computed at the end of each trial.

Example 1. Suppose that the capital endowed to each subject at the beginning of a trial is 24 francs, and the reward for winning the competition is 80 francs to each member of the winning network. Also suppose that you invest 20 francs and your partner invests 15 francs in developing your network's product. Thus, the total investment in your network's product is 35 francs. Assume that the competing network invests a total of 15 francs in the development of its product. Then,

$$\textit{probability of your network winning} = \frac{\textit{Total investment in your network's product}}{\textit{Total investment in your network's product} + \textit{Total investment in the competing network's product}}$$
$$= \frac{35}{35+15} = \frac{35}{50} = 0.7$$

Suppose that the random number randomly chosen by the computer is 0.589. As this random number is *smaller* than 0.7, your network is the winner. Therefore, each member of your network receives a reward of 80 francs.

Your payoff for this trial will be:
Your payoff = endowment − your investment + reward = 24 − 20 + 80 = 84 francs.
Your partner's payoff = endowment − your partner's investment + reward
= 24 − 15 + 80 = 89 francs.

Example 2. Suppose that the capital endowed to each subject at the beginning of a trial is 24 francs, and the reward for winning the competition is 80 francs to each member of the winning network. Also suppose that you invest 20 francs and your partner invests 15 francs in the developing your network's product. Thus, as in Example 1, the total investment in your network's product is 35 francs. Assume that the competing network invests a total of 15 francs in its product. Therefore, exactly as before,

$$\textit{probability of your network winning} = \frac{35}{35+15} = \frac{35}{50} = 0.7.$$

Suppose that the random number randomly chosen by the computer is 0.856. As this random number is *larger* than 0.7, your network loses the competition. Further, the investments are completely non-recoverable, and so the non-recoverability factor is 1. In other words, members of the losing network cannot recover any part of their investments.

Your payoff in this trial will be:
Your payoff = endowment – (non-recoverability factor × your investment) + reward = 24 − (1*20) + 0 = 4 francs.
Your partner's payoff = endowment – (non-recoverability factor × your partner's investment) + reward = 24 − (1*15) + 0 = 9 francs.

This concludes the description of the decision task. Paper and pencil are placed beside the computer terminal in case you wish to record the investments made in your network and in the competing network.

At the end of the experiment, your accumulated payoff will be converted to US dollars at the rate of 1000 francs = 2 dollars. You will be asked to sign a receipt for the money, and complete a brief questionnaire before leaving the lab. We are required to retain some biographical information about you, as we are paying you for participating in this experiment. However, during the course of this experiment you will remain anonymous. If you have any questions, please raise your hand and the supervisor will assist you.

After all the participants have understood the instructions, we will start the experiment. In order to help you become familiar with the decision task, you will participate in five practice trials, and then the actual 150 trials.

Instructions for network size = 2, recoverable investment

You will participate today in a decision making experiment concerning competition between two networks. Each network is comprised of two firms. You will be asked to make many decisions in the course of this decision making experiment, and you will be paid depending on your decisions and the decisions of other subjects. A research foundation has contributed funds to support this research. In the present experiment, you are asked to represent a partner in the network. Three other subjects participate in this experiment. One represents another firm that joins you in the network for jointly developing a product, whereas the other two subjects represent firms in the competing network.

The experiment involves many trials. At the beginning of each trial, each subject will be provided with some investment capital and then asked how much of it he/she wishes to invest in developing the new product of his/her network. The four subjects will be provided with the same amount of investment capital, and asked to make independent investment decisions.

Rules of the Investment Game. The rules of this investment game are quite simple.

1) The winner of the competition between the networks will be determined *in proportion* to the *total investments* of the two networks. In particular, the probability of your network winning the competition will be determined as follows:

$$\textit{probability of your network winning} = \frac{\textit{Total investment in your network's product}}{\textit{Total investment in your network's product} + \textit{Total investment in the competing network's product}}$$

2) The capital invested by partners in the winning network is consumed in developing the product and thus *completely nonrecoverable*. Members of the losing

network can potentially recover their investments. In this game, the investments of the losing network are *completely recoverable* (In other words, the non-recoverability factor for members of the losing network is 0).

3) Each member of the winning network will receive a fixed reward. The reward does not depend on the relative investments made by each member of the winning network. The fixed reward represents profit that each firm earns from producing and marketing the new product. Each member of the losing network receives no reward, but gets to take back their investments.

Experimental Procedure

As discussed above, there are *two networks*, and each network is comprised of *two partners*. At the beginning of each trial each subject will be given the same investment capital, which will remain unchanged from trial to trial. The investment capital will be stated in terms of a laboratory currency called "francs". At the end of the experiment your earnings will be converted to US dollars.

Once each player is provided with the investment capital, he or she must decide how much to invest in developing his/her network's new product. You may invest any number of francs (including zero), provided your investment does not exceed your endowment (investment capital allotted for the trial). After all the four players have made their investment decisions, privately and anonymously, the computer will compute the total investment made by each of the competing networks. The probability that your network wins the competition will be computed from the equation shown above. Then the computer will *randomly* choose a number between 0 and 1 to determine the winning network. If this random number falls in the interval between 0 and the probability that your network wins the competition, then your network will be declared the winner. Otherwise, if the random number is greater than your network's probability of winning, the competing network will win the competition. Members of the winning network will receive a reward of known size (in francs), whereas members of the losing network will receive no reward.

Note that if you invest more you increase the probability of winning the competition. However, the same reasoning applies to all the firms. Moreover, investing more does not guarantee winning the competition, as the winning network is determined *probabilistically*.

Computation of Individual Payoffs. The individual payoffs for a trial will be computed as follows:

Payoff to a member of the *winning network* = endowment for the trial
  – investment made by the firm in the trial + reward for winning the competition.

Payoff to a member of the *losing network* = endowment for the trial
  – (non-recoverability × investment made by the firm in the trial).

Outcome Information. At the end of each trial, the computer will display the following information on the computer screens:

1) The total investments made by the winning and losing networks,
2) The network winning the competition,
3) Your payoff for the trial,
4) Your cumulative earnings.

It is important to note that *only you know your investment decision*. This decision is made anonymously.

Network Membership. Network membership will vary randomly from trial to trial. *On each trial you will be paired with a different person in this room, and both of you will compete as a group against another new group of two players.* Therefore, you will never know the identity of your partner on any give trial.

The following two examples are provided to help you understand how your payoff is computed at the end of each trial.

Example 1. Suppose that the capital endowed to each subject at the beginning of a trial is 24 francs, and the reward for winning the competition is 80 francs to each member of the winning network. Also suppose that you invest 20 francs and your partner invests 15 francs in developing your network's product. Thus, the total investment in your network's product is 35 francs. Assume that the competing network invests a total of 15 francs in the development of its product. Then,

$$\textit{probability of your network winning} = \frac{\textit{Total investment in your network's product}}{\textit{Total investment in your network's product} + \textit{Total investment in the competing network's product}}$$

$$= \frac{35}{35+15} = \frac{35}{50} = 0.7$$

Suppose that the random number randomly chosen by the computer is 0.589. As this random number is *smaller* than 0.7, your network is the winner. Therefore, each member of your network receives a reward of 80 francs.

Your payoff for this trial will be:
Your payoff = endowment − your investment + reward = 24 − 20 + 80 = 84 francs.
Your partner's payoff = endowment − your partner's investment + reward
= 24 − 15 + 80 = 89 francs.

Example 2. Suppose that the capital endowed to each subject at the beginning of a trial is 24 francs, and the reward for winning the competition is 80 francs to each member of the successful network. Also suppose that you invest 20 francs and your partner invests 15 francs in the developing your network's product. Thus, as in

Example 1, the total investment in your network's product is 35 francs. Assume that the competing network invests a total of 15 francs in its product. Therefore, exactly as before,

$$\textit{probability of your network winning} = \frac{35}{35+15} = \frac{35}{50} = 0.7.$$

Suppose that the random number randomly chosen by the computer is 0.856. As this random number is *larger* than 0.7, your network loses the competition. Further, the investments are completely recoverable, and so the non-recoverability factor is 0. In other words, members of the losing network can all their investments.

Your payoff in this trial will be:
Your payoff = endowment – (non-recoverability factor × your investment) + reward = 24 – (0*20) + 0 = 24 francs.
Your partner's payoff = endowment – (non-recoverability factor × your partner's investment) + reward = 24 – (0*15) + 0 = 24 francs.

This concludes the description of the decision task. Paper and pencil are placed beside the computer terminal in case you wish to record the investments made in your network and in the competing network.

At the end of the experiment, your accumulated payoff will be converted to US dollars at the rate of 1000 francs = 2 dollars. You will be asked to sign a receipt for the money, and complete a brief questionnaire before leaving the lab. We are required to retain some biographical information about you, as we are paying you for participating in this experiment. However, during the course of this experiment you will remain anonymous. If you have any questions, please raise your hand and the supervisor will assist you.

After all the participants have understood the instructions, we will start the experiment. In order to help you become familiar with the decision task, you will participate in five practice trials and then the actual 150 trials.

Chapter 5

THE COGNITIVE ILLUSION CONTROVERSY: A METHODOLOGICAL DEBATE IN DISGUISE THAT MATTERS TO ECONOMISTS

Ralph Hertwig
University of Basel, Basel

Andreas Ortmann
Charles University and Academy of Sciences of the Czech Republic, Prague

Abstract

In the early 1970s, Daniel Kahneman and Amos Tversky launched a research program that showed that heuristic short-cuts can result in probability judgments that deviate from statistical principles. Because these *cognitive illusions* have important implications for economic behavior, the heuristics-and-biases program has attracted the attention of economists as well as numerous social scientists. Even as the heuristics-and-biases program gained acceptance outside psychology, it drew criticism within the field. In this chapter, we mine the debate among psychologists about the *reality of cognitive illusions* for methodological lessons of relevance to experimental economists. Our concern here is neither the controversy about cognitive illusions nor its implications for rationality. Instead, it is what we see as the important methodological insights that have emerged from the controversy, which can inform the choices that all behavioral experimenters wittingly or unwittingly make when they sample and represent stimuli for use in their experiments.

How do we make decisions? According to subjective expected utility (SEU) theory, which still holds sway throughout much of the social sciences, "decision makers behave *as if* utilities were assigned to outcomes, probabilities were attached states of nature, and decisions were made by taking expected utilities" (Mas-Colell, Whinston, & Green, 1995, p. 205, their emphasis). Although this is an elegant and often useful way to model decision outcomes, it imposes heroic knowledge and rationality requirements, and it clearly does not reflect the way people make decisions most of the time.[1]

Herbert Simon (1956) was the most outspoken critic of the assumption that SEU theory can be applied in any literal way to human choices. In his view, "the SEU model is a beautiful object deserving a prominent place in Plato's heaven of ideas" (Simon, 1990a, p. 194); real humans, however, "have neither the facts nor the consistent

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 113–130.
© 2005 *Springer. Printed in the Netherlands.*

structure of values nor the reasoning power at their disposal that would be required . . . to apply SEU principles" (p. 197). Simon did not limit himself to criticizing the "Olympian model" of SEU theory (Simon, 1990a, p. 198); he also proposed an alternative way to think about decision making, which he called *bounded rationality*.

Simon's vision of bounded rationality has two interlocking components: the limitations of the human mind and the informational structures of the environment in which the mind operates. Simon captured the interplay between these two components thus: "Human rational behavior . . . is shaped by a scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990b, p. 7). What Simon in effect argued was that rational behavior can only be understood in terms of both scissor blades: the mind and the environment. The cognitive blade requires that models of human judgment and decision-making rest on realistic assumptions about the mind's capacities rather than on idealized competencies. Due to the mind's limitations, people "*must use approximate methods to handle most tasks*" (Simon, 1990b, p. 6, his emphasis). These methods include recognition processes that often obviate the need for information search and, when information search is necessary, simple rules for guiding and terminating search and for making a decision based on the information obtained. The environmental blade is the statistical structure of the task environment. The extent to which the approximate methods of the cognitive blade are adapted to this statistical structure determines how well they perform.

The idea that environmental and cognitive structures work in tandem is not Simon's alone. Even before Simon coined the term bounded rationality, the psychologist Egon Brunswik (1955) proposed that the processes underlying human perception and cognition are adapted to the uncertain environments in which they evolved and now function. From this premise, he challenged the standard approach to psychological experimentation on ecological grounds (for a review of neo-Brunswikian research, see Dhami, Hertwig, & Hoffrage, in press). In the standard approach, which Brunswik called *systematic design*, experimenters vary one or a few independent variables in isolation and observe resulting changes in the dependent variable(s) while holding other variables constant or allowing them to vary randomly.

Systematic design strongly emphasizes internal validity, that is, the demonstration of causal relationships between variables. Brunswik believed that this approach thereby renders impossible the primary aim of psychological research, that is, to discover probabilistic laws that describe an organism's adaptation to the structure of its environment. In pursuit of this aim, experimenters must preserve this structure in the stimuli that they present to participants. If they tamper with this structure, Brunswik argued, they destroy the phenomenon under investigation or at least alter psychological processes such that the experimental findings are no longer representative of people's functioning outside the laboratory.

Brunswik also observed that psychologists followed a double standard in their practice of sampling in experimental research (Brunswik, 1944). Why, he asked, are procedures for sampling participants scrutinized while findings based on stimuli in the laboratory are blithely generalized to stimuli outside the laboratory? He argued that experimental stimuli should be representative of the population of stimuli to which the experimenter intends to generalize the findings in the same way that experimental participants should be representative of the population of people whose behavior the experimenter wishes to study. As an alternative to systematic design, Brunswik proposed *representative design*, which can take any of various forms. The one he seemed to favor is achieved by randomly sampling stimuli from the defined population of stimuli or conditions, or reference class, about which the experimenter aims to make inferences.

Simon's and Brunswik's ecological views of cognition share a methodological corollary: To understand how – and how well – cognitive algorithms work, behavioral researchers need to study them under conditions that are representative of the conditions under which they usually operate. In this chapter, we show how this ecological approach to experimentation has shed new light on findings from the *heuristics-and-biases* research program in psychology and argue that the resulting insights into cognition have important implications for experimental methods in economics.

1. COGNITIVE ILLUSIONS

In the early 1970s, Daniel Kahneman and Amos Tversky launched a research program that would strike a powerful blow to SEU theory as a descriptive model of human judgment and choice. The heuristics-and-biases program stresses that people have only limited "reasoning power" at their disposal, implicitly equating bounded rationality with irrationality: "Systematic, predictable differences between normative models of behavior and actual behavior occur because of what Herbert Simon . . . called 'bounded rationality'" (Thaler, 1980, p. 40). On this view, people's cognitive limitations necessitate reliance on cognitive heuristics to make judgments and choices. Although these heuristics are "highly economical and usually effective, . . . they lead to systematical and predictable errors" (Kahneman, Slovic, & Tversky, 1982, p. 20) that are variously referred to as biases, fallacies, or *cognitive illusions.*

In challenging the Olympian model of the human mind on which SEU theory rests, the heuristics-and-biases critique (e.g., Kahneman et al., 1982; Gilovich, Griffin, & Kahneman, 2002) has focused on the premise that the decision maker is able to form probabilistic beliefs or expectations about the state of nature and the effects of her action, and to process available information according to statistical principles (Schoemaker, 1982). In contrast to this premise, the heuristics-and-biases program has shown that people's statistical reasoning appears systematically biased and error-prone, and such biases were attributed to flawed cognitive software.

In recent years, the heuristics-and-biases program has attracted the attention of numerous social scientists, including economists (e.g., Barber & Odean, 2001; Camerer, 1995; Hirshleifer, 2001; Odean, 1999) and legal scholars (e.g., Sunstein, 2000). In fact, much of today's work in behavioral economics and behavioral finance draws inspiration and concepts from the heuristics-and-biases program (e.g., Shiller, 2000; Thaler, 1993). This attention is warranted because systematic biases may have important implications for economic behavior. In his analysis of "irrational exuberance" in the stock market during the late 1990s, for example, Shiller (2000) explicitly invoked Kahneman and Tversky's experimental results.

Even as the heuristics-and-biases program gained acceptance outside psychology, it drew criticism within psychology. Some critics suggested that the heuristics-and-biases research strategy has a built-in bias to find cognitive illusions (e.g., Krueger & Funder, 2004). Others claimed that some cognitive illusions were themselves illusory (e.g., Erev, Wallsten, & Budescu, 1994; Koehler, 1996). Perhaps the most influential objections were voiced by Gigerenzer (e.g., 1991, 1996), who argued that the heuristics onto which cognitive illusions were attributed were not precise process models; that the heuristics-and-biases program relied on a narrow definition of rationality; and that cognitive illusions can be reduced or made to disappear by representing statistical information differently than it typically had been in heuristics-and-biases experiments. A vigorous debate ensued (see Gigerenzer, 1996; Kahneman & Tversky, 1996).

Our concern here is neither the controversy about cognitive illusions nor its implications for rationality. Instead, it is what we see as the important methodological insights that have emerged from the controversy, which can inform the choices that all behavioral experimenters wittingly or unwittingly make when they sample and represent stimuli for their experiments. We have argued elsewhere that psychologists can learn from the experimental practices of economists (e.g., Hertwig & Ortmann, 2001; Ortmann & Hertwig, 2002). In this chapter, we mine the debate in psychology about the reality of cognitive illusions for methodological lessons of relevance to experimental economists. We begin by examining how stimuli are selected from the environment for inclusion in behavioral experiments.

2. SAMPLING STIMULI

Many kinds of real-world economic failures have been attributed to the *overconfidence bias*. Camerer (1995, p. 594), for example, suggested that the well-documented high failure rate of small businesses may be due to overconfidence, while Barber and Odean (2001; Odean, 1999) argued that overconfidence based on misinterpretation of random sequences of successes leads some investors, typically men, to trade too much. According to Shiller (2000), "[s]ome basic tendency toward overconfidence appears to be a robust human character trait" (p. 142). These conclusions are based on the results of psychological experiments in which confidence is studied using general-knowledge questions like the following:

> Which city has more inhabitants?
> (a) Canberra (b) Adelaide
> How confident are you that your answer is correct?
> 50%, 60%, 70%, 80%, 90%, 100%

Typically, when people say they are 100% confident of their answer, the relative frequency of correct answers is only about 80%. When they are 90% confident, the proportion correct is about 75%, and so on. The size of the bias is measured as the difference between participants' mean confidence and the mean percentage of correct answers. Like many other cognitive illusions, overconfidence bias is thought to be tenacious: "Can anything be done? Not much" (Edwards & von Winterfeldt, 1986, p. 656).

But is there really so little that can be done to undo the overconfidence bias? One implication of Brunswik and Simon's idea that cognitive strategies are adapted to the statistical structure of the task environment is that if the strategies are tested in environments that are unrepresentative of that environment, they will probably perform poorly. Adopting a Brunswikian perspective, Gigerenzer, Hoffrage, and Kleinbölting (1991) argued that this is why people appear overconfident in the laboratory. In other words, the way in which experimenters sample the questions posed to participants in overconfidence studies helps create the bias.

For illustration, let us assume that a person can retrieve only one piece of knowledge, or cue, pertaining to Australian cities, namely, whether or not a city is the national capital. How good would her inferences be if she inferred the relative population size of two Australian cities based solely on the capital cue? Consider the reference class of the 20 largest cities in Australia. Here the capital cue has an ecological validity of .74.[2] If a person's intuitive estimate of the validity of a cue approximates its ecological validity in the reference class[3] and if she uses the cue's validity as a proxy for her confidence, then her confidence judgments will be well calibrated to her knowledge. This prediction holds as long as the experimenter samples questions such that the cue's validity in the experimental item set reflects its validity in the reference class.

Gigerenzer et al. (1991) conjectured that the overconfidence effect observed in psychology studies stemmed from the fact that the researchers did not sample general-knowledge questions randomly but rather selected items in which cue-based inferences were likely to lead to incorrect choices. Suppose, for example, that an experimenter gives participants only five of the 190 possible paired comparisons of the 20 largest Australian cities: Canberra-Sydney, Canberra-Melbourne, Canberra-Brisbane, Canberra-Perth, and Canberra-Adelaide. In all these comparisons, a person who relies solely on the capital cue, (thus selecting Canberra) will go astray. In fact, if she assigns a confidence of 75% (the approximate ecological validity of the cue) to each pair, she will appear woefully overconfident, although the predictive accuracy of the capital cue is generally high. If the experimenter instead draws the pairs randomly from all possible paired comparisons of the 20 largest Australian cities, the person will no longer appear overconfident.[4] As they predicted, Gigerenzer et al. (1991, Study 1) found that when questions were randomly sampled from a defined reference class (e.g., all paired comparisons of the 83 German cities that have more than 100,000 residents) – that is, in a representative design – participants answered an average of 71.7% of the questions correctly and reported a mean confidence of 70.8%. When participants were presented with a selected set of items, as was typically the case in earlier studies, overconfidence reappeared: Participants answered an average of 52.9% of the questions correctly, and their mean confidence was 66.7%.

Recently, Juslin, Winman, and Olsson (2000) reviewed 130 overconfidence data sets to quantify the effects of representative and selected item sampling. Figure 1 depicts the overconfidence and underconfidence scores (regressed on mean confidence) observed in those studies. The overconfidence effect was, on average, large

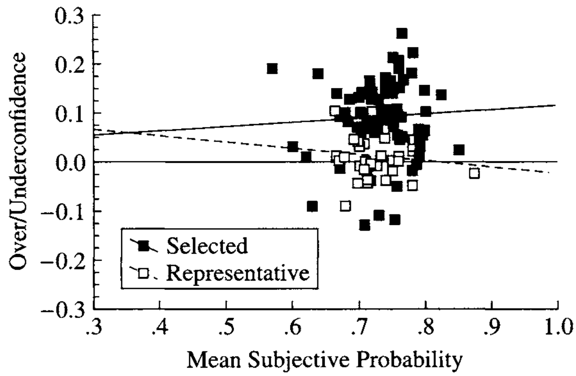


Figure 1. Regression lines relating over/underconfidence scores to mean subjective probability for systematically selected (black squares) and representative samples (open squares) (Reprint of Figure 2B from Juslin et al., *2000).*

when participants were given selected samples of questions and close to zero when they were given representative samples of questions. These results hold even when one controls for item difficulty, a variable to which the disappearance of overconfidence in Gigerenzer et al.'s (1991) studies has sometimes been attributed (see Griffin & Tversky, 1992; see also Brenner, Koehler, Liberman & Tversky, 1996).

The impact of item sampling on judgment and decision-making is not restricted to overconfidence. For instance, it has also been shown to affect the hindsight bias, that is, the tendency to falsely believe after the fact that one would have correctly predicted the outcome of an event. Hindsight bias is thought not only to undermine economic decision making (Bukszar & Connolly, 1988) but also to exert tremendous influence on judgments in the legal system (e.g., Sunstein, 2000; for an alternative view of the hindsight bias, see Hoffrage, Hertwig, & Gigerenzer, 2000). Like overconfidence, hindsight has been typically studied in psychology by having participants respond to general-knowledge questions.

To study the impact on hindsight of representative versus selected item sampling, Winman (1997) presented participants with selected or representative sets of general-knowledge questions such as "Which of these two countries has a higher mean life expectancy: Egypt or Bulgaria?" Before they were given an opportunity to respond, participants in the experimental group were told the correct answer (in this case, Bulgaria) and asked to identify the option they would have chosen had they not been told. Participants in the control group were not given the correct answer before they responded. If hindsight biased the responses to a given question, then the experimental group would be more likely to select the correct answer than would the control group. While this was the case, Winman also found that the size of the hindsight bias in the experimental group differed markedly as a function of item sampling: In the selected set, 42% of items elicited the hindsight bias, whereas in the representative set only 29% did so.

Using representative design, researchers have shown that cognitive illusions can be a byproduct of the slices of the world that earlier experimenters happen to take. The lesson is that methods of stimulus sampling can shape participants' performance and, by extension, inferences about human rationality. Experimenters who use selectively chosen or artificially constructed tasks in the laboratory risk altering the very phenomena that they aim to investigate. The issue is not that selected samples are inherently more difficult to handle but that cognitive strategies are adapted to the informational structure of the environment in which they have been learned (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999; Payne, Bettman, & Johnson, 1993).

3. DOES STIMULUS SAMPLING MATTER IN EXPERIMENTAL ECONOMICS?

The question of whether and how to sample from the environment has not been of much concern for experimental economists until recently, notwithstanding early calls for "parallelism" (e.g., Plott, 1987). Laboratory environments were typically created to test decision- or game-theoretic predictions derived from (possibly competing) formal models, with a focus on the equilibrium properties of those models. Given this research strategy, little attention was paid to how representative these environments were of their real-world counterparts. Indeed, why should it have been a concern? After all, the theories being tested were formulated to capture the essential characteristics of the world outside the laboratory.

Neglect of representative design in experimental economics was amplified by the practice of using abstract tasks. The rationale behind this methodological choice seems to have been that it would reduce the danger of eliciting participants' responses to field counterparts of the task rather than the task itself. There is now ample evidence that stripping away content and context prevents participants from applying the strategies that they use in their usual habitats. Relying mostly on evidence from psychology, Ortmann and Gigerenzer (1997) argued that experimental economists' convention of stripping the laboratory environment of content and context may be counterproductive and ought to be studied experimentally.

An early demonstration of the importance of representative design in economics was provided by economists Dyer and Kagel (1996) in an experimental investigation of the bidding behavior of executives from the commercial construction industry in one-shot common value auctions. Simple survivorship arguments suggest that such sophisticated bidders should be able to avoid the winner's curse in laboratory-based common value auctions designed to capture the essential characteristics of commercial bidding behavior. Dyer and Kagel (1996) found, however, that a significant number of the executives in their study fell victim to the winner's curse in the laboratory. The authors identified a number of differences between theoretical treatments in the literature – embodied in the experimental design – and practices in the industry that made the experimental design unrepresentative. For example, in the commercial construction industry, it seems to be possible for bidders to void the award of a contract that they realize would cost them dearly by claiming arithmetic errors. The executives' bidding behavior was maladapted to the laboratory situation because that situation failed to capture essential aspects of their natural ecology.[5]

In our view, the issue of representative design lies also at the heart of discussions about the existence of altruism, defined here – in line with recent usage – as a form of unconditional kindness (e.g., Fehr & Gächter, 2004). The debate has revolved around seemingly simple games such as symmetric and simultaneous prisoners' dilemmas (Colman, 1995); public good provision problems (Ledyard, 1995); asymmetric and sequential games such as dictator, ultimatum, and trust games (e.g., Camerer, 2003; Cox, 2004); and closely related gift exchange or principal-agent games. What these games have in common is that tests based on them seem to provide overwhelming evidence that participants are often altruistic, at least by the lights of deductive game theory as it is expounded in textbooks such as Kreps (1990) and Mas-Colell et al. (1995). Indeed, the ultimatum game "is beginning to upstage the PDG prisoner dilemma game in the freak show of human irrationality" (Colman, 2003, p. 147).

Or is it? Recall that the results that precipitated such conclusions are puzzling only if one takes as a benchmark deductive game theory's predictions for one-shot games or for finitely repeated games solvable through backward induction (Mas-Colell et al., 1995, Proposition 9.B.4). As various authors have pointed out (e.g., Hoffman, McCabe, & Smith, 1996), prisoners' dilemma, public good provision, dictator, ultimatum, trust, and gift exchange or principal-agent games are typically encountered indefinitely often in the game of life. As observed by Smith (1759/1982) and Binmore (1994, 1997), the game of life is therefore played using cognitive and behavioral strategies with consequences that probably differ markedly from the dire predictions of standard deductive game theory for one-shot and finitely repeated games. In Brunswik's terms, the standard implementations of prisoners' dilemma, public good provision, dictator, ultimatum, trust, and gift exchange or principal-agent games in experimental economics are unlikely to capture the conditions under which people usually encounter and make such choices. To the extent that participants perceive these games in the laboratory as some form of social dilemma, they are likely to retrieve experiences and strategies that, unbeknownst to the experimenter, change the nature of the game.

4. REPRESENTING STIMULI

After stimuli have been sampled, experimenters face another methodological question raised by the controversy about cognitive illusions, namely, how to represent the stimuli to participants. Just as the algorithms of a pocket calculator are tuned to Arabic rather than Roman numerals, cognitive processes are tuned to some information representations and not others (see Marr, 1982). A calculator cannot perform arithmetic operations on Roman numeral inputs, but this fact should not be taken to imply that it lacks an algorithm for multiplication. Similarly, the functioning of cognitive algorithms cannot be evaluated without considering the type of inputs for which the algorithms are designed. In their efforts to convey some aspect of reality to experimental participants, behavioral researchers use all kinds of representations, including words, pictures, and graphs. The choice of representation has far-reaching effects on the computations that a task demands and on the ease with which cognitive algorithms can carry out these operations.

The importance of task representation for cognitive performance has been extensively demonstrated in research on how people update probabilities to reflect new information. Given the importance to the SEU framework of the assumption that this updating process is Bayesian, it is not surprising that researchers in the heuristics-and-biases program have investigated the assumption's psychological plausibility. The results appear devastating for the premise that people are rational Bayesians. Time and again, experimenters found that people failed to make Bayesian inferences, even in simple situations where both the predictor and the criterion are binary. Kahneman and Tversky (1972) left no room for doubt: "Man is apparently not a conservative Bayesian: he is not Bayesian at all" (p. 450).

To get a feel for this research, consider the following study by Eddy (1982) of statistical inferences based on results of mammography tests. In the experiment, physicians received information that can be summarized as follows (the numbers are rounded):

> For a woman at age 40 who participates in routine screening, the probability of breast cancer is 0.01 [base rate, $p(\mathrm{H})$]. If a woman has breast cancer, the probability is 0.9 that she will have a positive mammogram [sensitivity, $p(\mathrm{D}|\mathrm{H})$]. If a woman does not have breast cancer, the probability is 0.1 that she will still have a positive mammogram [false-positive rate, $p(\mathrm{D}|\mathrm{not}-\mathrm{H})$]. Now imagine a randomly drawn woman from this age group with a positive mammogram. What is the probability that she actually has breast cancer?

The posterior probability $p(\mathrm{H}|\mathrm{D})$ that a woman who tests positive actually has breast cancer can be calculated using Bayes' rule, in which H stands for the hypothesis (e.g., breast cancer) and D for the datum (e.g., a positive mammogram):

$$p(\mathrm{H}|\mathrm{D}) = \frac{p(\mathrm{H})p(\mathrm{D}|\mathrm{H})}{p(\mathrm{H})p(\mathrm{D}|\mathrm{H}) + p(\mathrm{not}-\mathrm{H})p(\mathrm{D}|\mathrm{not}-\mathrm{H})}. \tag{1}$$

Inserting the statistical information from the mammography problem into Equation 1 yields:

$$\frac{(.01)(.90)}{(.01)(.90) + (.99)(.10)} \approx .08.$$

In other words, about 9 out of 10 women who receive a positive mammography result do not have breast cancer. Most of the physicians in Eddy's (1982) study overestimated the posterior probability: 95 of 100 physicians gave an average estimate of about .75. Many of them arrived at this estimate because they apparently mistook the sensitivity of the test $[p(\mathrm{D}|\mathrm{H})]$ for the posterior probability $p(\mathrm{H}|\mathrm{D})$ or because they subtracted the false positive rate from 100%. Any strategy that, like these two, ignores the base rate of breast cancer can lead to the *base-rate fallacy.*

Although the reality of the base-rate fallacy has been disputed on various grounds (e.g., Koehler, 1996), let us focus on the critique that is most closely related to the ecological approach to experimentation that is the focus of this chapter. Most studies that observed the base-rate fallacy presented information in the form of probabilities or percentages. Mathematically, probabilities, percentages, and frequencies are equivalent representations of statistical information. Psychologically, however, they are not equivalent. Physicist Richard Feynman (1967) described the consequences of information representation for deriving different mathematical formulations of the same physical law thus: "Psychologically they are different because they are completely unequivalent when you are trying to guess new laws" (p. 53). This insight is central to the argument that problems that represent statistical information in terms of *natural frequencies* rather than probabilities, percentages, or relative frequencies are more likely to elicit correct Bayesian inferences from both laypeople and experts (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). Natural frequencies are absolute frequencies of events that have not been normalized with respect to the base rates of the hypothesis or of its absence. In natural frequencies, the mammography problem would read:

> Of 1,000 women at age 40 who participate in routine screening, 10 women have breast cancer. Nine of those 10 women with breast cancer will test positive and 99 of the 990 women without breast cancer will also test positive. How many of those who test positive actually have breast cancer?

To see how natural frequencies are related to bounded rationality, recall Simon's (1990b) view that human rational behavior arises from the interplay between the structure of task environments and organisms' computational capabilities. In the case of statistical reasoning, this means that one cannot understand people's inferences without taking external representations of statistical information, as well as cognitive algorithms for manipulating that information, into account. For most of their existence, humans and animals have made statistical inferences on the basis of information encoded sequentially through their direct experience. Natural frequencies are the result of this process. The concept of mathematical probability, in contrast, emerged only in the mid-seventeenth century (Daston, 1988). Percentages seem to have become common representations only in the aftermath of the French revolution, mainly for purposes of calculating taxes and interest; only very recently have percentages become a way to represent risk and uncertainty more generally. Based on these observations, Gigerenzer and Hoffrage (1995) argued that minds have evolved to deal with natural frequencies rather than with probabilities.[6]

Independent of evolutionary considerations, Bayesian computations are simpler to perform when the relevant information is presented in natural frequencies than in probabilities, percentages, or relative frequencies because natural frequencies do not require figuring in base rates. Compare, for instance, the computations that an algorithm for computing the posterior probability that a woman has breast cancer given a positive mammogram when the information is represented in probabilities (shown in Equation 1) with those necessary when the same information is presented in natural frequencies:

$$p(\mathrm{H}|\mathrm{D}) = \frac{pos\ \&\ cancer}{pos\ \&\ cancer + pos\ \&\ \neg cancer} = \frac{9}{9+99} \approx .08. \tag{2}$$

Equation 2 is Bayes' rule for natural frequencies, where *pos&cancer* is the number of women with breast cancer and a positive test and *pos&¬cancer* is the number of women without breast cancer but with a positive test. In the natural frequency representation, fewer arithmetic operations are necessary, and those required can be performed on natural numbers rather than fractions.

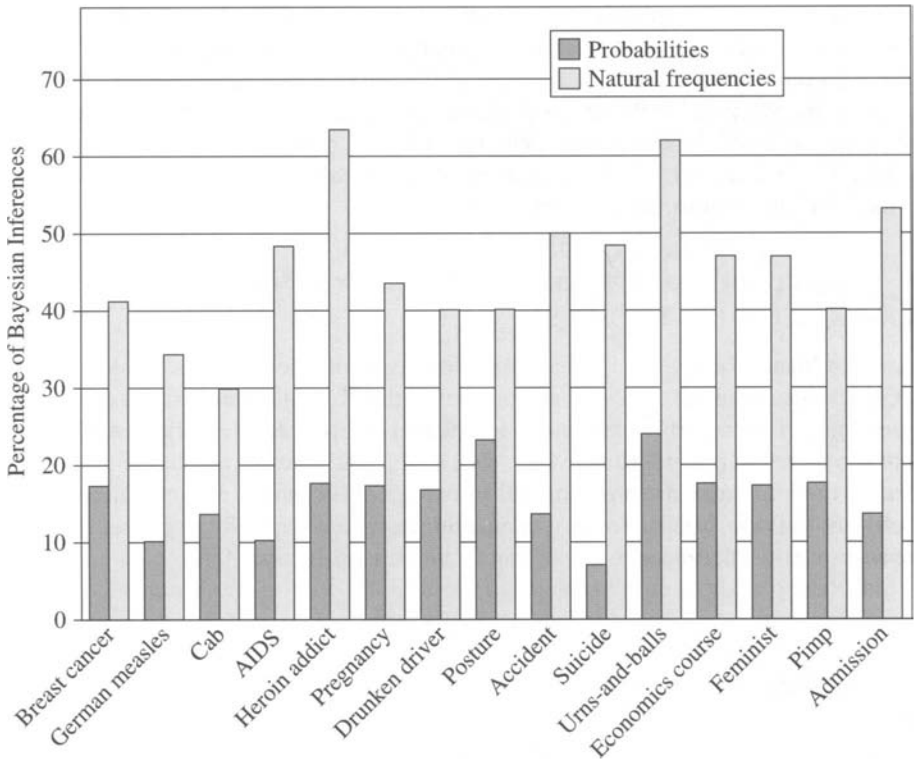


Figure 2. Across 15 Bayesian reasoning problems, statistical information was either presented in probabilities or in natural frequencies. In each problem, probabilistic reasoning improved when statistical information was communicated in natural frequencies (adapted from Gigerenzer and Hoffrage, 1995). To qualify as a Bayesian inference, the participant had to respond with the exact Bayesian estimate, and *the written protocol had to confirm that the response was derived from actual Bayesian reasoning.*

Probabilistic reasoning improves when statistical information is presented in terms of natural frequencies rather than probabilities. Take, for instance, Gigerenzer and Hoffrage's (1995) study of university students' ability to solve a set of 15 Bayesian reasoning problems that included many of the problems in which the base-rate fallacy had been observed (e.g., the mammography problem). Participants received the statistical information in each problem in terms of probabilities or natural frequencies. As Figure 2 shows, in each of the 15 problems, natural frequencies substantially increased the proportion of Bayesian inferences. On average, people reasoned the Bayesian way in only about 1 out of 6 cases given probabilities, whereas in 1 out of 2 cases they did so given natural frequencies. Other studies show that natural frequencies foster Bayesian reasoning among experts who make medical and forensic inferences (e.g., Hoffrage et al., 2000). Moreover, Sedlmeier and Gigerenzer (2001) designed a tutorial computer program that teaches people to translate probability information into natural frequencies (representation training) or, alternatively, to insert probabilities into Bayes' rule (rule training). Rule training resulted in the typical forgetting curve, whereas representation training resulted in robust probabilistic thinking even three months after the training.

Regardless of one's take on the evolutionary argument about natural frequencies,[7] it seems to be widely accepted that the extent to which people obey statistical principles or fall prey to biases such as base-rate fallacy depends on the way in which statistical information is presented.[8]

5. DOES STIMULUS REPRESENTATION MATTER IN EXPERIMENTAL ECONOMICS?

An important example of how information representation matters in economics experiments is the Allais paradox. Together with Ellsberg's paradox, it is the most prominent of the (early) violations of expected utility theory reported in the economics literature (Kreps, 1990; Mas-Colell et al., 1995). According to the independence axiom, aspects that are common to two gambles should not influence choice behavior (Savage, 1954). For any three alternatives X, Y, and Z taken from a set of options S, the independence axiom can be written (Fishburn, 1979):

$$\text{If } pX + (1 - p)Z \tag{3}$$

The following choice problems produce violations of the axiom:

A:	100 million	for sure
B:	500 million	$p = .10$
	100 million	$p = .89$
	0	$p = .01$

By eliminating a .89 probability to win 100 million from both gambles A and B, Allais obtained the following alternatives:

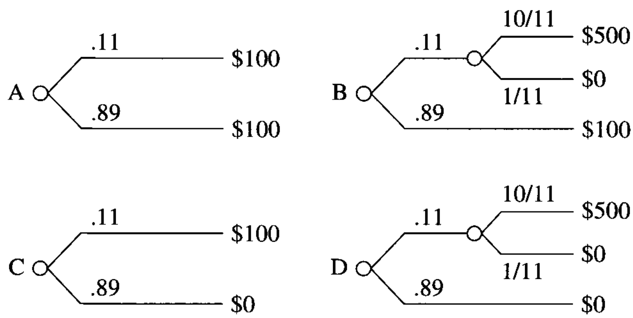


Figure 3. A graphical representation of the gambles involved in the Allais paradox that reduces the proportion of inconsistent choices (adapted from Kreps, 1990).

C:	100 million	$p = .11$
	0	$p = .89$
D:	500 million	$p = .10$
	0	$p = .90$

The majority of people choose A over B and D over C (e.g., MacCrimmon, 1968), which constitutes a violation of the axiom.

However, there is evidence that different task representations can lead to considerable reductions in the percentage of inconsistent choices. For example, when the gambles are presented to participants in the graphical form shown in Figure 3 (adapted from Kreps, 1990), then inconsistent behavior decreases sharply (see also Conlisk, 1989, for another example of the impact of task representation on the percentage of inconsistent choices). As with probability representations of Bayesian inference, the problem with the standard representation of the gambles, which coalesces probabilities and makes the payoffs more difficult to compare, is their complexity.

Uncertainty and risk are arguably the dominant theme of modern economics (e.g., Mas-Colell et al., 1995; Kreps, 1990). Probabilities are therefore an essential ingredient of solution concepts such as sequential equilibrium that are used to analyze problems of moral hazard, adverse selection, screening, and signaling that can be conceptualized as games of strategic interaction under incomplete information. Signaling games, the most prominent in this class, go back to Spence's highly influential analysis of informational transfers in hiring and related processes. The basic problem is that workers with higher abilities may not be able to signal this fact credibly to employers. Spence (1974) suggested that, to signal their type, such workers might invest in education. If it is easier for workers with higher abilities to invest in education (as is customarily assumed), then they might be able to distinguish themselves from their less able competitors.

Such signaling games typically have multiple Nash equilibria, the number of which theorists have tried to reduce by imposing various restrictions on out-of-equilibrium beliefs. This is where probabilities come in. Such refinements require the person who uses the model to use Bayes's rule so as to make the strategy profile and the belief system mutually consistent. As every graduate student in economics can attest, this is typically a computational task of a tall order. Not surprisingly, the experimental evidence from tests of signaling games and refinements indicates that some of the subtler refinements (e.g., beyond sequential equilibrium) overtax participants (e.g., Banks, Camerer, & Porter, 1994). It is important to repeat that these models make heroic knowledge and rationality assumptions as well as assumptions about commonly known identical beliefs. Where do they come from? And what is their ecological validity?

The few experimental tests of signaling models that exist have matched participants repeatedly and observed how they learned. Note that participants in such games, whether or not they know the distribution of types of workers, have to perform belief updating that is likely to be affected by information representation. The results are a mixed bag that shows, among other things, that meaningful context both facilitates learning within a game and across related games (Cooper & Kagel, 2003). In our view, the question of how to represent information is key to the design of such learning experiments.

6. CONCLUSION

Its implications for human rationality aside, the cognitive illusion controversy in psychology has spawned a body of research with important implications for experimental economics. This research demonstrates that theoretical questions such as how well people are calibrated to their own knowledge and whether people update probabilities in a Bayesian way cannot be disentangled from the methodological questions of how to sample and represent experimental stimuli from the environment. To the extent that cognitive strategies and environmental structures go hand in hand, the world that is realized or represented in the laboratory codetermines how well the strategies perform and, ultimately, experimenters' conclusions.

Germane here is Vernon Smith's (2002) recent discussion of the Duhem-Quine problem in the context of experimentation in economics. The crux of the problem is that any experiment represents a test of two things: the hypotheses derived from the theory of interest and the auxiliary hypotheses necessary to implement the experiment. In psychological and economic experiments, the latter include hypotheses about measurement instruments, participant payments, and instructions. Because of the auxiliary hypotheses, any failure of the experiment to confirm the theoretical hypotheses can be explained in one of three ways: The theory is wrong; one or more of the auxiliary hypotheses are wrong; or both the theory and the auxiliary hypotheses are wrong. Thus, in Lakatos's words (quoted in Smith, 2002, p. 98): "No theory is or can be killed by an observation. Theories can always be rescued by auxiliary hypotheses."

Although experimental outcomes are thus inherently ambiguous, Smith sees no reason for despair. On the contrary, he argues, the Duhem-Quine problem is a driving force behind methodological innovation and scientific progress. Ambiguous results spark not only controversy but also the execution of new experiments designed to narrow the range of tenable interpretations. The results of these experiments, in turn, illuminate the extent to which the behavior of interest is sensitive to methodological variation. They also suggest new research questions, thus initiating a new cycle of experiments. In Smith's (2002) words, "The bottom line is that good-enough solutions emerge to the baffling infinity of possibilities, as new measuring systems emerge, experimental tools are updated, and understanding is sharpened" (p. 104).

We share Smith's (2002) optimistic pragmatism, although, having observed the tug of war over cognitive illusions for a decade, we are not convinced that more experiments always bring more clarity. Still, the cognitive illusion controversy has yielded profound knowledge about how human reasoning, judgment, and choice are affected by stimulus representation and stimulus sampling. In experimental economics, the auxiliary hypotheses needed to perform an experiment are in themselves substantive theories of, for instance, the interaction between cognitive processes and environmental structures. It is here where psychology has something to contribute to experimental economics.

ACKNOWLEDGMENT

Ralph Hertwig, University of Basel, Basel, Switzerland; Andreas Ortmann, Charles University and Academy of Sciences of the Czech Republic, Prague. We would like to thank Dirk Engelmann and Pavlo Blavatskyy for many constructive comments and Valerie Chase for valuable editorial input.

Correspondence should be addressed to Ralph Hertwig, Institute for Psychology, University of Basel, Missionsstrasse 60/62, 4055 Basel, Switzerland. Electronic mail may be sent to ralph.hertwig@unibas.ch.

NOTES

[1] Interpreting the principal components of SEU theory in "as-if" terms, as is often proposed, skirts the question of what cognitive processes lead people to their decisions.

[2] Gigerenzer et al. (1991) defined the ecological validity of a cue as the proportion of correct inferences that a person using only that cue would make in the subset of paired comparisons where the cue discriminates between alternatives (e.g., where one city is a capital and the other is not).

[3] An extensive literature (e.g., Zacks & Hasher, 2002) indeed suggests that people are well calibrated to environmental frequencies.

[4] Creating an exhaustive set of paired comparisons of the 20 largest Australian cities results in 190 comparisons. In 171 of the 190 pairs, the capital cue does not discriminate (because neither of the cities is a capital). In such cases, let us assume that the person guesses and estimates her confidence to be 50%. In 19 cases, the capital cue discriminates. Let us assume that the person estimates her confidence to be the cue's ecological validity, which is 75%. Averaged across all cases, her mean confidence should therefore be 53%, as should be the percentage of comparisons to which she provides the correct answer.

[5] Since then, a small but increasing number of economics studies has addressed the issue of representative design. An encouraging development in this vein is field experiments that use nontraditional subject pools, real-life decision situations, and real-life goods and services (Harrison & List, in press).

[6] This argument is consistent with developmental studies indicating the primacy of reasoning about discrete numbers and counts over fractions and with studies of adult humans and animals showing that

they can monitor frequency information in their natural environment in fairly accurate and automatic ways (see Gigerenzer & Hoffrage, 1995).

[7] For discussion of these issues, see, for instance, Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni (1999), and Hoffrage, Gigerenzer, Kraus, and Martignon (2002).

[8] The possible reasons for why representation matters, however, are controversially discussed (e.g., Tversky & Kahneman, 1983; Hertwig & Gigerenzer, 1999; Mellers, Hertwig, & Kahneman, 2001).

REFERENCES

Banks, J., Camerer, C. F., & Porter, D. (1994). "An experimental analysis of Nash refinements in signaling games." *Games and Economic Behavior*, 6, 1–31.

Barber, B. M., & Odean T. (2001). "Boys will we boys: Gender, overconfidence, and common stock investment." *Quarterly Journal of Economics*, 116, 261–292.

Binmore, K. (1994). *Game theory and the social contract: Playing fair* (Vol. 1). Cambridge, MA: MIT Press.

Binmore, K. (1997). *Game theory and the social contract: Just playing* (Vol. 2). Cambridge, MA: MIT Press.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). "Overconfidence in probability and frequency judgments: A critical examination." *Organizational Behavior and Human Decision Processes*, 65, 212–219.

Brunswik, E. (1944). "Distal focussing of perception: Size constancy in a representative sample of situations." *Psychological Monographs*, 56, 1–49.

Brunswik, E. (1955). "Representative design and probabilistic theory in a functional psychology." *Psychological Review*, 62, 193–217.

Bukszar, E., & Connolly, T. (1988). "Hindsight bias and strategic choice: Some problems in learning from experience." *Academy of Management Journal*, 31, 628–641.

Camerer, C. F. (2003). *Behavioral game theory. Experiments in strategic interaction.* Princeton, NJ: Princeton University Press.

Camerer, C. F. (1995). "Individual decision making." In J. H. Kagel & A. E. Roth (eds.). *Handbook of experimental economics* (pp. 587–703). Princeton, NJ: Princeton University Press.

Colman, A. M. (2003). "Cooperation, psychological game theory, and the limitations of rationality in social interaction." *Behavioral and Brain Sciences*, 26, 139–198.

Colman, A. M. (1995). *Game theory and its applications in the social and biological sciences* (2nd ed.). Amsterdam: Butterworth-Heinemann.

Conlisk, J. (1989). "Three variants on the Allais paradox." *American Economic Review*, 79, 392–407.

Cooper, D., & Kagel, J. (2003). "The impact of meaningful context on strategic play in signaling games." *Journal of Economic Behavior and Organization*, 50, 331–337.

Cosmides, L., & Tooby, J. (1996). "Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty." *Cognition*, 58, 1–73.

Cox, J. C. (2004). "How to identify trust and reciprocity." *Games and Economic Behavior*, 46, 260–281.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (in press). "The role of representative design in an ecological approach to cognition." *Psychological Bulletin.*

Daston, L. (1988). *Classical probability in the Enlightenment.* Princeton, NJ: Princeton University Press.

Dyer, D., & Kagel, J. H. (1996). "Bidding in common value auctions: How the commercial construction industry corrects for the winner's curse." *Management Science*, *42*, 1463–1475.

Eddy, D. M. (1982). "Probabilistic reasoning in clinical medicine: Problems and opportunities." In D. Kahneman, P. Slovic, & A. Tversky (eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, UK: Cambridge University Press.

Edwards, W., & von Winterfeldt, D. (1986). "On cognitive illusions and their implications." In H. R. Arkes & K. R. Hammond (eds.), *Judgment and decision making* (pp. 642–679). Cambridge, UK: Cambridge University Press.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). "Simultaneous over- and underconfidence: The role of error in judgment processes." *Psychological Review*, 101, 519–527.

Fehr, E., & Gächter, S. (2004). "Fairness and retaliation: The economics of reciprocity." In C. F. Camerer, G. Loewenstein, & M. Rabin (eds.), *Advances in behavioral economics* (pp. 510–532). Princeton, NJ: Princeton University Press.

Feyman, R. (1967). *The character of physical law*. Cambridge, MA: MIT Press.

Fishburn, P. C. (1979). "On the nature of expected utility." In M. Allais & O. Hagen (eds.), *Expected utility hypotheses and the Allais paradox* (pp. 243–257). Dordrecht, Netherlands: Reidel.

Gigerenzer, G. (1991). "How to make cognitive illusions disappear: Beyond heuristics and biases." In W. Stroebe & M. Hewstone (eds.), *European Review of Social Psychology* (Vol. 2, pp. 83–115). Chichester, UK: Wiley.

Gigerenzer, G. (1996). "On narrow norms and vague heuristics: A reply to Kahneman and Tversky." *Psychological Review*, 103, 592–596.

Gigerenzer, G., & Hoffrage, U. (1995). "How to improve Bayesian reasoning without instruction: Frequency formats." *Psychological Review*, 102, 684–704.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). "Probabilistic mental models: A Brunswikian theory of confidence." *Psychological Review*, 98, 506–528.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. Oxford: Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, England: Cambridge University Press.

Griffin, D., & Tversky, A. (1992). "The weighing of evidence and the determinants of confidence." *Cognitive Psychology*, 24, 411–435.

Harrison, G. W., & List, J. (in press). *Field experiments*.

Hertwig, R., & Gigerenzer, G. (1999). "The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors." *Journal of Behavioral Decision Making*, 12, 275–305.

Hertwig, R., & Ortmann, A. (2001). "Experimental practices in economics: A challenge for psychologists?" *Behavioral and Brain Sciences*, 24, 383–451.

Hirshleifer, D. (2001). "Investor psychology and asset pricing." *The Journal of Finance*, 61, 1533–1597.

Hoffman, E., McCabe, K., & Smith, V. L. (1996). "On expectations and the monetary stakes in ultimatum games." *International Journal of Game Theory*, 25, 289–301.

Hoffrage, U., Gigerenzer, G., Kraus, S., & Martignon, L. (2002). "Representation facilitates reasoning: What natural frequencies are and what they are not." *Cognition*, 84, 343–352.

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). "Hindsight bias: A by-product of knowledge updating?" *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 566–581.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). "Communicating statistical information." *Science*, 290, 2261–2262.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). "Naive probability: A mental model theory of extensional reasoning." *Psychological Review*, 106, 62–88.

Juslin, P., Winman, A., & Olsson, H. (2000). "Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect." *Psychological Review*, 107, 384–396.

Kahneman, D., Slovic, P., & Tversky, A. (eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). "On the reality of cognitive illusions: A reply to Gigerenzer's critique." *Psychological Review*, 103, 582–591.

Kahneman, D., & Tversky, A. (1972). "Subjective probability: A judgment of representativeness." *Cognitive Psychology*, 3, 430–454.

Koehler, J. J. (1996). "The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges." *Behavioral and Brain Sciences*, 19, 1–53.

Kreps, D. M. (1990). A course in microeconomic theory. Princeton, NJ: Princeton University Press.

Krueger, J. I., & Funder, I. (2004). "Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition." *Behavioral and Brain Sciences*.

Ledyard, J. (1995). "Public good experiments." In J. H. Kagel & A. E. Roth (eds.), *Handbook of experimental economics* (pp. 111–194). Princeton, NJ: Princeton University Press.

Marr, D. (1982). *Vision*. New York: Freeman and Company.

Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory*. New York: Oxford University Press.

MacCrimmon, K. R. (1968). "Descriptive and normative implications of the decision-theory postulate." In K. H. Borch & J. Mossin (eds.), *Risk and Uncertainty* (pp. 3–23). New York: St. Martin's Press.

Mellers, B., Hertwig, R., & Kahneman, D. (2001). "Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration." *Psychological Science*, 12, 269–275.

Odean, T. (1999). "Do investors trade too much?" *American Economic Review*, 89, 1279–1298.

Ortmann, A., & Gigerenzer, G. (1997). "Reasoning in economics and psychology: Why social context matters." *Journal of Institutional and Theoretical Economics*, 153, 700–710.

Ortmann, A., & Hertwig, R. (2002). "The costs of deception: Evidence from psychology." *Experimental Economics*, 5, 111–131.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

Plott, C. R. (1987). "Dimensions of parallelism. Some policy applications of experimental economics." In A. E. Roth (ed.), *Laboratory experimentation in economics: Six points of view* (pp. 193–219). New York: Cambridge University Press.

Savage, L. J. (1954). *The foundations of statistics* (2nd ed.). New York: Dover.

Schoemaker, P. J. H. (1982). "The expected utility model: Its variants, purposes, evidence and limitations." *Journal of Economic Literature*, 20, 529–563.

Sedlmeier, P., & Gigerenzer, G. (2001). "Teaching Bayesian reasoning in less than two hours." *Journal of Experimental Psychology: General*, 130, 380–400.

Shiller, R. J. (2000). *Irrational exuberance*. Princeton, NJ: Princeton University Press.

Simon, H. A. (1956). "Rational choice and the structure of environments." *Psychological Review*, 63, 129–138.

Simon, H. A. (1990a). "Alternative visions of rationality." In P. K. Moser (ed.), *Rationality in action: Contemporary approaches* (pp. 189–204). New York: Cambridge University Press.

Simon, H. A. (1990b). "Invariants of human behavior." *Annual Review of Psychology*, 41, 1–19.

Smith, A. (1759/1982). *The theory of moral sentiments*. Indianapolis, IN: Liberty Classics.

Smith, V. L. (2002). "Method in experiment: Rhetoric and reality." *Experimental Economics*, 5, 91–110.

Spence, M. (1974). *Market signaling: Informational transfer in hiring and related processes*. Cambridge, MA: Harvard University Press.

Sunstein, C. R. (2000). *Behavioral law and economics*. Cambridge, UK: Cambridge University Press.

Thaler, R. (1980). "Toward a positive theory of consumer choice." *Journal of Economic Behavior and Organization*, 1, 39–60.

Thaler, R. H. (1993). *Advances in behavioral finance*. New York: Russell Sage Foundation.

Tversky, A., & Kahneman, D. (1983). "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment." *Psychological Review*, 90, 293–315.

Winman, A. (1997). "The importance of item selection in 'knew-it-all-long' studies of general. knowledge." *Scandinavian Journal of Psychology*, 38, 63–72.

Zacks, R. T., & Hasher, L. (2002). "Frequency processing: A twenty-fie year perspective." In Sedlmeier, P. & Betsch, T. (eds.), *Etc. Frequency processing and cognition* (pp. 21–36). Oxford: Oxford University Press.

Chapter 6

EXPLORING ELLSBERG'S PARADOX IN VAGUE-VAGUE CASES

Karen M. Kramer
Edward Hines Jr. Veterans Administration Medical Center, Hines, IL

David V. Budescu
University of Illinois, Urbana-Champaign, IL

Abstract

We explore a generalization of Ellsberg's paradox to the Vague-Vague (V-V) case, where neither of the probabilities (urns) is specified precisely, but one urn is always more precise than the other. We present results of an experiment explicitly designed to study this situation. The paradox was as prevalent in the V-V cases, as in the standard Precise-Vague (P-V) cases. The paradox occurred more often when differences between ranges of vagueness were large. Vagueness avoidance increased with midpoint for P-V cases, and decreased for V-V cases. Models that capture the relationships between vagueness avoidance and observable gamble characteristics (e.g., differences of ranges) were fitted.

Key words: Ellsberg's paradox, ambiguity avoidance, vagueness avoidance, vague probabilities, imprecise probabilities, probability ranges, logit models

Over eighty years ago, Knight (1921) and Keynes (1921) independently distinguished between the problems of choice under uncertainty and ambiguity. Forty years later, Ellsberg (1961) demonstrated the relevance of this distinction with the following simple problem: A Decision-Maker (DM) has to bet on one of two urns containing balls of two colors, say Red and Blue. The composition (proportions of two colors) of one urn is known, but the composition of the other urn is completely unknown. Imagine that one of the colors (Red or Blue) is arbitrarily made more desirable, simply by associating it with a positive prize of size $x. If DMs are asked to choose one urn when each color is more desirable, many are more likely to select the urn with known content *for both colors* and "avoid ambiguity[1]". This pattern of choices violates Subjective Expected Utility Theory (SEUT), and this tendency is widely known as the "(two-color) Ellsberg's paradox".

The most common and appealing explanation of Ellsberg's paradox (e.g., Camerer and Weber, 1992) is that it is due to "ambiguity (or, in our terms, vagueness)

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 131–154.
© 2005 *Springer. Printed in the Netherlands.*

aversion". The logic of this explanation is straightforward and compelling – *If within each pair, most DMs choose the more precise urn, the modal pattern of joint choices (across the two replications when Red or Blue are the target colors) would, necessarily, lead to the paradox.* Various psychological explanations were offered for the subjects' preference for the more precise urn. Subjects may simply choose the urn about which they have more knowledge and information (Edwards, cited in Roberts, 1963, footnote 4; Baron and Frisch, 1994; Keren and Gerritsen, 1999). The different levels of information may induce various levels of competence (Heath and Tversky, 1991). Other, more complex, explanations rely on perception of "hostile nature" (Yates and Zukowski, 1976; Keren and Gerritsen, 1999), anticipation of evaluation by others (Ellsberg, 1963; Fellner, 1961; Gärdenfors, 1979; Knight, 1921; MacCrimmon, 1968; Roberts, 1963; Toda and Shuford, 1965; Slovic and Tversky, 1974), self-evaluation (Ellsberg, 1963; Roberts, 1963; Toda and Shuford, 1965), perception of competition (Kühberger and Perner, 2003), and others (see reviews by Camerer and Weber, 1992 and Curley, Yates, and Abrams, 1986). Curley et al. (1986) tested empirically some of these theories and suggested that "evaluation by others" is the most promising for future research of the phenomenon's psychological rationale. Regardless of the underlying psychological reason(s), Ellsberg's paradox has become almost synonymous with vagueness avoidance. In fact, most empirical research has focused on single choices between pairs of gambles varying in their precision, and only very few studies (e.g., MacCrimmon and Larsson, 1979) have actually replicated the full paradoxical pattern across two choices.

Many researchers have tried to model the behavior underlying this paradox (see Camerer and Weber, 1992 for a comprehensive review, and Becker and Brownson, 1964; Curley and Yates, 1985, 1989; Einhorn and Hogarth, 1986; for typical studies). Most of this research has used Precise-Vague (P-V) cases, where the probabilities of the two colors in one urn are known precisely, but the probabilities in the other urn are vague (specified imprecisely). This work has identified some of the factors and conditions that contribute to the intensity of the preference for precision. For example, Einhorn and Hogarth (1986) used probability predictions, insurance pricing, and warranty pricing tasks, to show vagueness avoidance at moderate to high probabilities of gains, and vagueness seeking for low probabilities of gains. Kahn and Sarin (1988) and Hogarth and Einhorn (1990) confirmed these results.

An interesting trend in the literature has been the extension of the paradox to new, more general, situations. It is possible to show that the paradoxical pattern of choices is obtained when the vagueness in the second urn is only partial, i.e., when the DM knows that $\Pr(\text{Red}) \geq x$, $\Pr(\text{Blue}) \geq y$, s.t., $0 \leq x, y \leq 1$, but $(x + y) < 1$. This implies that $x \leq \Pr(\text{Red}) \leq (1 - y)$, i.e., Pr(Red) is within a range of size $R = (1 - x - y)$ centered at $M = (1 + x - y)/2$. Similarly, $y \leq \Pr(\text{Blue}) \leq (1 - x)$, i.e. in a range of size $R = (1 - x - y)$ centered at $M = (1 + y - x)/2$. The current study follows this trend by extending the paradox to Vague-Vague (V-V) cases, where the composition of both urns is only partially specified. Typically, the range of possible probabilities in one urn is narrower than the range of the second urn, but both ranges share the same central value. Thus, Pr(Red | Urn I) $\geq x_1$, Pr(Blue|Urn I) $\geq y_1$, Pr(Red | Urn II) $\geq x_2$, and Pr(Blue | Urn II) $\geq y_2$, subject to the constraints: $0 \leq x_1, y_1, x_2, y_2 \leq 1$, $(x_1 + y_1) < 1$, $(x_2 + y_2) < 1$. Furthermore, $|x_1 - y_1| = |x_2 - y_2|$, but $R_1 = (1 - x_1 - y_1) \neq R_2 = (1 - x_2 - y_2)$. In other words, $x_1 \leq$ Pr(Red | Urn I) $\leq (1 - y_1)$ and $x_2 \leq$ Pr(Red | Urn II) $\leq (1 - y_2)$, and the common midpoint of both ranges is $M = (1 + x_1 - y_1) = (1 + x_2 - y_2)$.

The effects of vagueness in P-V cases are relatively well understood (see for example the list of stylized facts in Camerer and Weber's 1992 review), but the V-V case is more complicated. Becker and Brownson (1964) found inconsistencies when they tried to relate vagueness avoidance to differences in the ranges of vague probabilities, and Curley and Yates' studies (1985, 1989) were inconclusive with regard to the presence and intensity of vagueness avoidance in V-V cases. Curley and Yates (1985) examined the choices subjects made in the P-V and V-V case as a function of the width(s) of the range(s) and the common midpoint of the range of probabilities. They showed that people were more likely to be vagueness averse as the midpoint increased in P-V cases, but not in V-V cases. Neither vagueness seeking nor avoidance was the predominant behavior for midpoints < .40. The range difference between the two urns was not sufficient for explaining the degree of vagueness avoidance, and no effect of the width of the range was found in preference ratings over the pairs of lotteries.

Undoubtedly, the range difference (wider range – narrower range) is the most salient feature of pairs of gambles with a common midpoint, and one would expect this factor to influence the degree of observed vagueness avoidance. Range difference captures the *relative precision* of the two urns, and DMs who are vagueness averse are expected to choose the more precise urn more often. In fact, it is sensible to predict a positive monotonic relationship between the relative precision of a pair of urns and the intensity of vagueness avoidance displayed. It is surprising that Curley and Yates could not confirm this expectation. We will consider this prediction in more detail in the current study.

However, the relative precision of a given pair can not fully explain the DM's preferences in the V-V case. Consider, for example, the following three urns: Urn A: $0.45 \leq p \leq 0.55$; Urn B: $0.30 \leq p \leq 0.70$; Urn C: $0.15 \leq p \leq 0.85$, where p is the probability of the desirable event (Red or Blue ball). All urns have a common midpoint (0.5) but vary in their (im)precision. Urn A has a range of 0.10, Urn B has a range of 0.40, and Urn C spans a range of 0.70. Imagine that a DM has to choose between A and B, and between B and C. In both pairs the range difference (relative precision) is the same (0.30), but vagueness avoidance is expected to be stronger for the A, B pair, because most people would prefer the higher certainty associated with A. If, on the other hand, there is a fair amount of vagueness in both urns, people may feel that vagueness is unavoidable, and may focus their attention on other features. For example, they may notice that, in the best possible case, Urn C offers a very high probability (0.85) of the desirable event. This shift of attention may reduce the tendency to avoid vagueness and may lead to indifference or vagueness seeking.

This example highlights the importance of the more precise urn in the pair. The range width of probabilities in this urn represents the greatest possible (an upper bound on) precision, which is what most DMs tend to seek (Becker and Brownson, 1964). We refer to this value as the pair's *minimal imprecision*. We predict that, everything else being equal, vagueness avoidance should increase as the minimal imprecision decreases. Conversely, as minimal imprecision increases (i.e., as the more precise urn becomes more vague), we should observe more instances of indifference between the two urns, and an increased tendency of vagueness preference.

The P-V pairs represent a special case in which the minimum imprecision is always 0. Thus, only considerations of relative precision are relevant for these choices. Otherwise, the level of vagueness avoidance depends on both minimal imprecision and relative precision. But the two factors are negatively correlated. Thus, one is unlikely to encounter large levels of relative precision in cases with large minimal imprecision. For example, if the more precise urn in a pair has a high minimal imprecision, say 0.70, the relative precision cannot exceed 0.30. On the other hand, if the more precise urn in the pair has a low minimal imprecision, say 0.20, the relative precision can be as high as 0.80. In general, Max(Relative Precision) $\leq$ (1 $-$ Minimal imprecision), or Max(Minimal imprecision) $\leq$ (1 $-$ Relative Precision). One factor that constrains the minimal imprecision (and, indirectly, the relative precision) in a pair is the midpoint of the range. Note that for any urn, Max(Minimal imprecision) $\leq 2 \times [\text{Min}\{M, (1 - M)\}]$, where M is the midpoint of the range, subject to $0 > M > 1$.[2] Thus, the effects of the two types of (im)precision may interact with the midpoint of the pair.

Choices in the V-V case can be summarized by the following reasonable scenario: DMs identify and focus first on the more precise urn. If it is "sufficiently precise" and/or "substantially more precise" than the other member of the pair, DMs are most likely to choose it. If, however, the narrower range urn is "not sufficiently precise" nor "substantially more precise" than the other member of the pair, DMs may be indifferent between the urns, and in some cases they may be tempted to favor the less precise urn. Choices in the P-V reflect only considerations of relative precision. This qualitative description avoids the difficult questions of what exactly constitutes "sufficient precision", what is considered "substantially more precise", and what is the relative salience of these two factors. We will address these issues in more detail when we fit quantitative models to the tendency to avoid vagueness.

A good portion of the literature on choice under vagueness focuses on the ranges of the two urns, and a good deal of the experimental work (e.g., Curley and Yates, 1985; Yates and Zukowski, 1976) has studied the effects of the ranges, R_i, ($i = 1, 2$), and midpoints, M_i ($i = 1, 2$), on DM's choices. Consistent with this approach our models will also emphasize the midpoint, relative precision, and minimal imprecision *of the pair*, where the latter two factors are defined by the range of probabilities of the two urns.

1. CURRENT STUDY

The purpose of the present study is to study DM's choices in the presence of vagueness, and their tendency to succumb to Ellsberg's paradox in the domain of gains. We will be especially concerned with the V-V case, where both lotteries are imprecise and will contrast them with the choices in the "standard" P-V case, using a design similar to the one used by Curley and Yates (1985). We will, however use a much larger number of V-V pairs covering more ranges at three different midpoints. The subjects' choices in each pair will be classified as vagueness seeking, vagueness avoiding, or indifferent to vagueness, and the proportions of vagueness avoidance choices will be analyzed as a function of the pairs' minimal imprecision, relative precision and their common midpoint.

As indicated earlier, vagueness avoidance is expected to increase with relative precision and with reduction in minimal imprecision. There is empirical evidence that the intensity of vagueness avoidance increases with midpoint (Curley and Yates, 1985; Einhorn and Hogarth, 1986), and the midpoint may interact with the two precision measures of a pair. For example, we expect pairs with low midpoints will induce less vagueness avoidance than pairs with high midpoints. In addition, if the more precise urn's range is closer to the other urn's range, people are expected to feel more indifferent (and possibly be more vagueness seeking) between the urns. For low midpoints, this behavior may exist with greater values of relative precision and smaller values of minimal imprecision than for other midpoints.

In our experiment we present each pair of urns twice, and make a different event (i.e., marble color) the "target" (i.e., the more desirable one) on each presentation. This allows us to analyze the subjects' choices not only in terms of their attitude to (im)precision on each trial but also in terms of the emerging response patterns when matched pairs are considered simultaneously. These patterns are (a) the *classical Ellsberg's paradox* (choosing twice the more precise urn); (b) the *reversed paradox* (choosing twice the more vague urn); (c) *consistency* (choosing different urns on the two occasions); (d) *indifference* on both occasions; and weak *indifference* (being indifferent on one occasion and exhibiting a clear preference on the other).

Thus, the experiment verifies the presence of the paradoxical pattern in the V-V case, and compares its prevalence with the P-V case. The prevalence of the paradox will be analyzed as a function of the midpoint, range widths, and/or range differences. In general, we expect the factors that induce higher levels of vagueness avoidance to also increase the frequency of the paradoxical pattern, but an intriguing question that was never fully examined is whether the occurrence of the paradox can be predicted precisely from the subjects' attitudes towards precision. We expect Ellsberg's paradox to be the modal, but not the universal, pattern. In those cases when the paradox does not occur, we predict different patterns as a function of the common midpoint. We expect subjects to exhibit more indifference for pairs with a midpoint of 50, where it is easier and more natural to either imagine symmetric distributions of probabilities (Ellsberg, 1963; footnote 8), and/or a greater number of possible distributions (Ellsberg, 1961; Roberts, 1963), than with extreme midpoints. On the other hand, we expect subjects to be consistent with SEUT more often with extreme midpoints, where the imagined distributions are more likely to be asymmetric and to be skewed in opposite directions.

2. METHOD

Subjects: Subjects were 107 undergraduates registered in an introductory psychology class at the University of Illinois in Urbana-Champaign. They received an hour of credit for participation, and had a chance to win additional money at the end of the experiment.

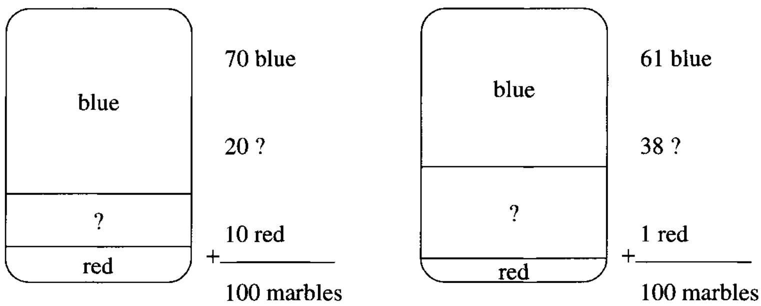
Stimuli: The subjects saw representations of 63 different pairs of urns. The colors of marbles in the two urns were red and blue. The pairs varied in terms of the (common) midpoint, and the ranges of values in each urn. Fifteen pairs had a midpoint of 20, fifteen pairs had a midpoint of 80, and thirty-three pairs had a midpoint of 50. Throughout the paper the midpoint is equivalent to the "expected" number of red marbles (and 100- the "expected" number of blue marbles) in each urn under a uniform distribution. Six different range widths were used with a midpoint of 20 or 80 (0, 2, 20, 30, 38, 40), and ten ranges were used with a midpoint of 50 (0, 2, 20, 30, 38, 40, 50, 80, 98, 100).

Two groups of subjects were recruited. In one group (80 subjects) the urn with the narrower range was always presented on the left; in the second group (27 subjects) the placement of the urn with a narrower range was randomly determined on every trial. Our analysis did not indicate any position effect, so the data from both groups were combined.

Procedure: Subjects were run individually on personal computers in a lab. In the first part of the experiment, each of the 63 pairs was presented twice. In one presentation the desirable outcome was associated with the acquisition of a red marble. In the other presentation, the desirable outcome was associated with the acquisition of a blue marble. The 126 pairs were presented, one at a time, in a different randomized order for each subject. For each pair the subjects had to decide whether to select Urn I, Urn II, or either urn (i.e., express indifference). Figure 1 shows an example of the display for a midpoint of 20 (which is equal to a blue midpoint of 80).

Before the experiment, subjects were told that two pairs would be randomly selected and played at the conclusion of the experiment, and that if they had selected "either urn" a coin toss would determine the urn choice. These instructions encouraged subjects to choose one urn, yet allowed them the opportunity to express indifference if truly desired.

In the second part of the experiment, the same 63 pairs were presented in random order and subjects were asked to indicate, on a scale from 1–7, how dissimilar the contents of the two urns were. These judgments were used to examine the subjects' subjective perceptions of the urns. The results of this (multidimensional scaling)



Which urn would you pick if you wanted to get a blue marble?

Urn I either urn Urn II

Figure 1. Example of a choice trial, red midpoint = 20. Actual colors were used with the words in the urn depictions.

analysis indicated a high similarity of subjectively scaled values to the actual stated values, so further discussion of these findings is unnecessary.

On average, subjects completed the experiment in approximately 30 minutes. At the conclusion of the experiment, a pair of urns was chosen, and the subjects' choices for each color were noted. To determine the subject's payoff, this pair of urns was prepared by placing 100 red and blue marbles in each urn. A random number generator, which used a uniform distribution over the relevant ranges of values[3], was used to determine the number of red marbles in the two urns. A marble was removed from the urn the subject (or the coin) selected. If the color of the selected marble matched the target color, the subject won $3. Otherwise, the subject did not receive any money. Twenty-one subjects received $0, 59 gained $3, and 27 gained $6 (average payoff = $3.17).

3. RESULTS

Ellsberg's paradox refers to an inconsistent pattern of revealed preference in two related choice problems. The first section of the analysis will focus on the intensity of the paradoxical pattern in these *joint choices*. It is common to attribute the paradoxical pattern to the subjects' tendency to avoid the more vague of the two gambles. Of course, this avoidance of vagueness can only be observed directly in a single choice, between gambles that vary only with respect to their imprecision. The second part of the analysis will focus on these choices and will model subjects' propensity to choose the more precise gamble within a pair.

3.1. Analysis of joint choice patterns

Distribution of responses: For any given pair of urns there are nine distinct possible responses that can be classified into five patterns: classic paradox (CP), reverse

Table 1. The possible patterns of joint selection for any given pair

	Blue		
Red	*VA*	*I*	*VS*
VA	Classic Paradox (CP)	Weak Indifference (WI)	Consistency #1 (C)
I	Weak Indifference (WI)	Indifference (I)	Weak Indifference (WI)
VS	Consistency #2 (C)	Weak Indifference (WI)	Reverse Paradox (RP)

Note: VA-vagueness avoidance, I-indifference, VS-vagueness seeking

paradox (RP), indifference (I), consistency (C), and weak indifference (WI). Indifference and consistency conform with SEUT. Weak indifference does not allow an unequivocal test of the paradox. All the patterns are illustrated in Table 1.

The distribution of responses was determined for each pair across all subjects and was compared to the expected distribution under the null hypothesis of random responses using χ^2 tests.[4] All the χ^2 values had right-hand p-values less than .05, and 61 (97%) had p-values less than .01. Thus, we reject the possibility that subjects' choices were random.

The distributions of choices over the nine patterns for P-V and V-V cases and for all midpoints are summarized in the various panels of Table 2. Panels 1–3 contain information for each midpoint separately and panel 4 is a subset of panel 2 that contains information for a midpoint of 50 but only for those ranges that were also

Table 2. Percentages of each pattern for the P-V and V-V cases, by midpoint

2.1. Red Midpoint = 20

$N = 535$ (P-V)	*Blue*							
$N = 1070$ (V-V)	VA		I		VS		Total	
Red	P-V	V-V	P-V	V-V	P-V	V-V	P-V	V-V
VA	33.60	28.50	3.70	6.20	7.10	18.30	44.40	53.00
I	9.20	6.20	8.60	8.20	1.90	4.60	19.70	19.00
VS	24.90	13.40	2.80	2.10	8.20	12.50	35.90	28.00
Total	67.70	48.10	15.10	16.50	17.20	35.40	100.00	100.00

Note: VA = vagueness avoidance, I = indifference, VS = vagueness seeking

Table 2. (cont'd)

2.2. Red Midpoint = 50 (includes all pairs)

$N = 963$ (P-V)	*Blue*							
$N = 2568$ (V-V)	VA		I		VS		Total	
Red	P-V	V-V	P-V	V-V	P-V	V-V	P-V	V-V
VA	40.90	38.00	5.40	5.80	6.10	7.90	52.40	51.70
I	7.10	5.20	16.50	18.50	3.10	3.50	26.70	27.20
VS	7.40	7.70	3.30	3.50	10.20	10.00	20.90	21.20
Total	55.40	50.90	25.20	27.80	19.40	21.40	100.00	100.00

Note: VA = vagueness avoidance, I = indifference, VS = vagueness seeking

2.3. Red Midpoint = 80

N = 535 (P-V)	*Blue*							
N = 1070 (V-V)	VA		I		VS		Total	
Red	P-V	V-V	P-V	V-V	P-V	V-V	P-V	V-V
VA	32.70	30.50	7.30	4.70	20.00	13.40	60.00	48.60
I	5.80	6.90	9.20	8.70	2.20	2.50	17.20	18.10
VS	11.70	18.80	2.10	3.20	9.00	11.40	22.80	33.40
Total	50.20	56.20	18.60	16.60	31.20	27.30	100.00	100.00

Note: VA = vagueness avoidance, I = indifference, VS = vagueness seeking

used for the midpoints 20 and 80. Finally, panel 5 is a summary across all midpoints based on the subset of common ranges (i.e., panels 1, 3 and 4).

The marginal distributions (the last row and column in the table, which are labeled Total) document the predominance of vagueness avoidance for each color and each midpoint, for P-V and V-V cases. They also revealed a greater tendency of vagueness seeking than indifference for the extreme midpoints (20 and 80), and a reversed trend (more indifference than vagueness seeking) for the midpoint of 50.

Table 2. (cont'd)

2.4. Red Midpoint = 50 (including only ranges used for all midpoints)

$N = 535$ (P-V) / $N = 1070$ (V-V)	*Blue*							
	VA		I		VS		Total	
Red	P-V	V-V	P-V	V-V	P-V	V-V	P-V	V-V
VA	38.70	33.40	6.00	6.80	6.90	7.00	51.60	47.20
I	6.70	5.50	17.80	20.90	3.00	3.70	27.50	30.10
VS	7.20	7.10	3.00	4.40	10.70	11.10	20.90	22.60
Total	52.60	46.00	26.80	32.10	20.60	21.80	100.00	100.00

Note: VA = vagueness avoidance, I = indifference, VS = vagueness seeking

2.5. All red midpoints, with only comparable pairs (Tables 2.1 + 2.3 + 2.4)

$N = 1605$ (P-V) / $N = 3210$ (V-V)	*Blue*							
	VA		I		VS		Total	
Red	P-V	V-V	P-V	V-V	P-V	V-V	P-V	V-V
VA	35.00	30.80	5.70	5.90	11.30	12.90	52.00	49.60
I	7.20	6.20	11.80	12.60	2.40	3.60	21.40	22.40
VS	14.70	13.10	2.60	3.20	9.30	11.70	26.60	28.00
Total	56.90	50.10	20.10	21.70	23.00	28.20	100.00	100.00

Note: VA = vagueness avoidance, I = indifference, VS = vagueness seeking

The distribution of the five general patterns for P-V and V-V cases are displayed in Figure 2. There is some slight variation across midpoints but, in general, the classic paradox was the most prevalent, and the reverse paradox was the least prevalent one. As predicted, indifference was almost twice as prevalent for a midpoint of 50 than for the other two midpoints. Conversely, consistency was twice as frequent for extreme midpoints than for the midpoint of 50. In general, the results for P-V and V-V pairs were highly similar.

Consider again Table 2 that summarizes all choices and patterns. The margins documented the predominance of vagueness avoidance, and the upper left cell

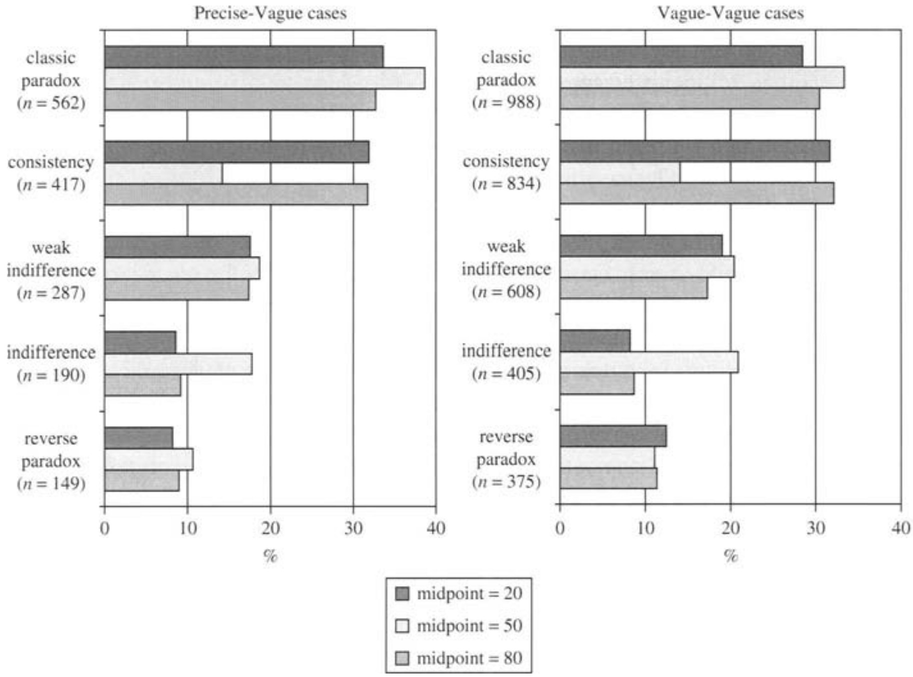


Figure 2. Distribution of the five general patterns for P-V and V-V cases, by midpoint.

(VA/VA, e.g., 33.60 and 28.50 in Table 2.1) in every sub-table indicated that the classic paradox was the modal pattern. A natural question is whether the frequency of the paradox can be predicted exclusively from the subjects' *global* tendency to choose the more precise lottery. In other words, is Pr(Classic Paradox) = Pr(VA|Red) × Pr(VA|Blue)? Surprisingly, the answer is negative! In fact, in all tables the paradox occurred more frequently than one would predict from independent vagueness avoidance choices (overall, 5.83% above expectation). Conversely, the indifferent pattern and the reverse paradox were under-predicted by the marginal distributions (by 7.67% and 3.60%, respectively). Clearly, the rate of the various patterns (e.g., CP) was not driven exclusively by a *constant* tendency to avoid/prefer vagueness. The intensity of this tendency varied as a function of various features of the gambles. The rest of this paper is devoted to modeling the effects of these features on the intensity of vagueness avoidance.

Log-linear models of the joint patterns: The frequency of each of the five patterns in Figure 2 was tabulated as a function of the urns' midpoint and their relative precision. Log-linear models were fit to each pattern, to determine the effect of the two factors on the observed frequency of the target pattern. The saturated model is:

$$\ln(f_{ij}) = \lambda + \lambda_{M(i)} + \lambda_{D(j)} + \lambda_{MD(ij)} \tag{1}$$

where M is the Midpoint effect,
D is the range Difference effect, and
MD is the interaction of these effects.

Reduced models are defined by constraining some of the parameters to equal 0. The fits of reduced versions of model (1) for the classic paradox are presented in Table 3, separately for the P-V and V-V pairs. For each case we show the frequencies being modeled, as well as the results of the model fits. For each model, we report the degrees of freedom (df), the likelihood ratio (G^2) and the ratio G^2/df. Usually, the model's goodness of fit is tested by comparing G^2 with its asymptotic sampling distribution (χ^2). In this situation, this would be inappropriate because the observations are not independent, as required for a valid application of this test. An alternative procedure is to use the ratio G^2/df as a descriptive measure of the fit of a model. In general, the closer the G^2/df ratio is to 1, the better the fit of the model (e.g., Goodman, 1971a, 1975; Haberman, 1978). In both cases, the reduced model including the range difference effect alone was the best, judged by the proximity of its G^2/df ratio to unity. It appears that the pair's relative precision is the most important predictor of the incidence of CP.

Table 3. Log-linear analysis of frequency of the Classic Paradox

3.1a. Frequency table of CP in the P-V Case

	Range Difference				
Midpoint	*2*	*20*	*30*	*38*	*40*
20	32	25	42	41	40
50	27	39	47	47	47
80	27	32	41	36	39

3.1b. Log-linear model results for the P-V case

model	*df*	G^2	G^2/df*
Complete Independ.	8	3.37	.42
Just Midpoint	12	18.00	1.50
Just Range Diff.	10	6.49	.65 *

($N = 562$)
Note: * if $G^2/df \approx 1$, model fits

Table 3. (cont'd)
3.2a. Frequency table of CP in the V-V case

	Range Difference							
Midpoint	*2*	*8*	*10*	*18*	*20*	*28*	*36*	*38*
20	20	27	57	62	35	42	27	35
50	14	21	58	83	40	47	46	48
80	17	18	50	75	30	49	41	46

3.2b. Log-linear model results for the V-V case

model	*df*	G^2	G^2/df*
Complete Independ.	14	12.33	.88
Just Midpoint	21	178.61	8.51
Just Range Diff.	16	16.47	1.03 *

($N = 988$)
Note: * if $G^2/df \approx 1$, model fits

Set-association models: A more detailed analysis distinguishes between pairs with various levels of minimal imprecision. Table 4.1 shows the frequency of the CP pattern as a function of the narrower and wider ranges of the urns involved (across all three midpoints). This analysis involves constrained (triangular) arrays of frequencies, and requires fitting special types of log-linear models to measure the effects of the relevant factors. The set-association model (e.g., Wickens, 1989), allows testing the significance of hypothesized "treatment effects" in such triangular arrays of frequencies. The most general form of the model is:

$$\ln(f_{ij}) = \lambda + \lambda_{N(i)} + \lambda_{W(j)} + \lambda_{T(k)} \tag{2}$$

where N is the Narrower range effect,
W is the Wider range effect, and
T is the "treatment effect."

Naturally, when $\lambda_{T(k)} = 0$, there is no treatment effect and we obtain the "quasi-independence model", that is similar to a regular independence model but applies to partial tables (Bishop, Fienberg, and Holland, 1975; Wickens, 1989; Rindskopf, 1990). A variety of treatment effects can be specified to reflect various hypotheses. We fitted two such "effects". The first was the "CP pattern" in which it was hypothesized that the frequency of the Classic Paradox pattern would be greater for pairs where the relative precision was larger and the minimal imprecision was smaller.[5] The second model simply distinguished between the P-V and V-V cases. All three models for the classic paradox are shown in Table 4, across all midpoints as well as

Table 4. Set-association models of Classic Paradox frequencies

4.1 Triangular table of frequencies over all midpoints

	Wide Range					
Narrow Range	*0*	*2*	*20*	*30*	*38*	*40*
0	–	86	96	130	124	126
2	–	–	111	138	114	129
20	–	–	–	83	109	105
30	–	–	–	–	66	82
38	–	–	–	–	–	51
40	–	–	–	–	–	–

4.2 Set-association model results, midpoint = 20.

model	*df*	G^2	G^2/df*
Quasi-independence	4	14.69	3.67
P-V vs. V-V	3	14.21	4.74
"CP" pattern	3	12.81	4.27

($N = 485$)
Note: * if $G^2/df \approx 1$ model fits

4.3 Set-association model results, midpoint = 50.

model	*df*	G^2	G^2/df*
Quasi-independence	4	41.87	10.47
P-V vs. V-V	3	39.99	13.33
"CP" pattern	3	22.49	7.50

($N = 564$)
Note: * if $G^2/df \approx 1$ model fits

Table 4. (cont'd)

4.4 Set-association model results, midpoint = 80.

model	*df*	G^2	G^2/df*
Quasi-independence	4	26.64	6.66
P-V vs. V-V	3	24.49	8.16
"CP" pattern	3	11.21	3.74

($N = 501$)
Note: * if $G^2/df \approx 1$ model fits

4.5 Set-association model results, all midpoints.

model	*df*	G^2	G^2/df*
Quasi-independence	4	70.88	17.72
P-V vs. V-V	3	66.65	22.22
"CP" pattern	3	38.77	12.92

($N = 1550$)
Note: * if $G^2/df \approx 1$ model fits

for each midpoint separately. Again, the closer the ratio G^2/df is to 1, the better the fit of the model. Note that the G^2/df ratios of the models with the "P-V vs. V-V" treatment were comparable to those of the quasi-independence model, which suggested that subjects did not treat P-V and V-V pairs differently, and the paradoxical pattern occurred with similar intensity in both cases. On the other hand, for midpoints greater than, or equal to, 50 and over all midpoints, the model including the "CP pattern" is clearly superior over the quasi-independence and the "P-V vs. V-V" models. Thus, Ellsberg's paradox was more likely to occur in pairs with large relative precision and small minimal imprecision when the midpoint was greater than 20. With the low midpoint, the occurrence of the paradox appears to be independent of these joint effects of relative precision and minimal imprecision.

3.2. *Analysis of choices within a single gamble*

Distribution of responses: We have shown in Table 2 that in most cases subjects tend to choose the more precise of the two gambles in a pair. The marginal means of Table 2.5 indicate that across all (4,815 × 2 =) 9,630 cases examined, the more precise option was chosen (2,426 + 2,520 =) 4,946 times (i.e., 51.36% of the time). Vagueness preference was observed (1,325 + 1,274 =) 2,599 times (in 26.99% of

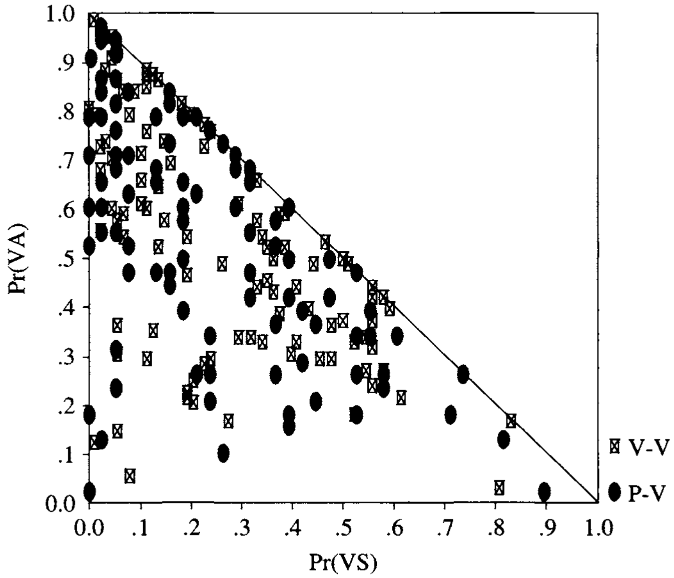


Figure 3. Proportions of VA and VS choices in P-V and V-V pairs, for 107 subjects.

the cases), and subjects expressed indifference towards (im)precision on (1,021 + 1,064 =) 2,085 occasions (21.65% of the cases). This general pattern held for extreme midpoints, for both colors and for the two types of pairs (P-V and V-V). The distribution over the three choices varied slightly over midpoints, colors, and types of pairs (in particular, for the midpoint of 50, indifference was more prevalent than vagueness preference). However, the distinct preference for precision was almost constant across all cases.

The predominance of vagueness avoidance holds for most individual subjects as well. Figure 3 displays the trinomial distribution of choices for all 107 subjects, for P-V and V-V cases. Each subject is represented by two points (P-V and V-V cases) in the plane whose coordinates are the probability of choosing the more vague gamble, Pr(VS), on the x-axis, and the probability of choosing the more precise gamble, Pr(VA), on the y-axis. The third probability (of being indifferent) is implied by these two, and it can be determined by simple subtraction: Pr(Ind) = 1 − Pr(VA) − Pr(VS), and inferred from each point's location relative to the origin, where Pr(Ind) = 1, and the negative diagonal (where Pr(Ind) = 0). The most important feature of this display for the current purposes is that 83 subjects (78%) for P-V, and 81 subjects (76%) for V-V are located in the upper corner (above the main diagonal along which Pr(VA) = Pr(VS)), indicating that they displayed vagueness avoidance much more frequently than vagueness seeking.

Modeling vagueness avoidance: In this section we seek to model the subjects' choices at the pair level as a function of the pair's type (P-V or V-V), midpoint, relative precision, minimal imprecision, and the interactions among these factors. We focus on those cases where the subjects expressed a clear preference between the two options, and discard cases where subjects expressed indifference. The dependent variable is the log-odds (also called the logit) of choosing the more precise urn in a pair, i.e., Log{Pr(VA)/Pr(VS)}, as measured across the two complementary color choices for each pair. The predictors used in the model are:

1. The pair's Relative Precision (RELPR) = Difference in widths between the two urns;
2. The pair's Minimal Imprecision (MINIM) = Width of the imprecise range of the more precise urn;
3. The pair's Midpoint (MID);
4. The pair's type (TYPE) = a binary variable that distinguishes between the V-V and the P-V cases; and
5. All pair-wise interactions between these four (centered) factors.

The models were fitted to 57 of the pairs examined. We excluded six pairs with minimal imprecision greater that 40, because such extreme values are incompatible with the extreme midpoints (20 and 80)[6]. The best model without interactions has an R^2 of 0.29 ($R^2_{\text{adj}} = 0.26$) and is achieved by the following equation (all coefficients are standardized):

$$\text{Logit(VA)} = 0.40{*}\text{RELPR} - 0.24{*}\text{MINIM},$$

As predicted, the tendency to avoid vagueness depends primarily on the relative precision ($r = 0.50$) and, to a lesser degree, on the minimal imprecision ($r = -0.40$). Although the midpoint and the type of the pair are not significant predictors ($r =$ 0.02 and 0.21, respectively), they contribute to the prediction of the target behavior through their interactions with other factors. A model with the four factors and two interactions involving the midpoint, achieves an impressive fit of R^2 of 0.71 ($R^2_{\text{adj}} = 0.68$):

$$\begin{aligned}\text{Logit(VA)} = {} & 0.40{*}\text{RELPR} - 0.22{*}\text{MINIM} + 0.05{*}\text{MID} - 0.03{*}\text{TYPE} \\ & - 0.54{*}(\text{MINIM}{*}\text{MID}) - 0.17{*}(\text{TYPE}{*}\text{MID}).\end{aligned}$$

To fully understand the effects of the two interactions, consider Table 5 that lists the mean probability of choosing the more precise option (and avoid vagueness) for all relevant combinations of the factors in question. The first column of the table shows that for the P-V pairs the tendency to avoid vagueness peaks at the highest midpoint (80). In the other columns (corresponding to the V-V pairs) the pattern is reversed with the weakest vagueness aversion measured at the high midpoint (80). The table also shows that the tendency to avoid vagueness across various levels of minimal imprecision depends on the midpoint: Vagueness avoidance decreases for high midpoints (50 and 80), but it increases for the low midpoint of 20, as minimal

Table 5. Interaction between the absolute imprecision (range width) of the pair of urns and its midpoint

	Minimum imprecision/ range width of pair					
	P-V	V-V				
Midpoint	*0*	*2*	*20*	*30*	*38*	*All*
20	.59 (5)	.63 (4)	.68 (3)	.71 (2)	.70 (1)	.64 (15)
50	.73 (9)	.73 (8)	.70 (7)	.67 (2)	.55 (1)	.71 (27)
80	.76 (5)	.69 (4)	.57 (3)	.49 (2)	.34 (1)	.53 (15)
All	.70 (19)	.70 (16)	.67 (13)	.62 (6)	.53 (3)	.68 (57)

Notes: – In each cell, the probability of choosing the more precise of the two urns is displayed. This probability is inferred from the mean Log{Prob(VA)/Prob(VS)}.
– Number in parentheses indicates the number of pairs.

imprecision increases. This pattern is inconsistent with the "perceived information" effect described by Keren and Gerritsen (1999).

The two interactions are not distinct because all P-V pairs have a minimal imprecision of 0. Thus, it is possible to fit a simpler version of the model by including only one interaction term, without sacrificing much in term of goodness of fit. Indeed, the model:

$$\text{Logit(VA)} = 0.40*\text{RELPR} - 0.24*\text{MINIM} + 0.06*\text{MID} - 0.64*(\text{MINIM}*\text{MID}),$$

fits the data almost equally well ($R^2 = 0.70$, $R^2_{\text{adj}} = 0.67$). This model does not include the binary factor corresponding to the sharp dichotomy (P-V vs. V-V), but rather a continuous variable that captures the level of minimal imprecision. This highlights the fact that the two situations are not qualitatively distinct. It is, however, instructive to note that in the P-V case, where the minimal imprecision is 0, the relative precision is, simply, the range of the vague urn and the model is reduced to simple additive form involving the common midpoint (center) and the range of the more vague urn, as suggested by Curley and Yates (1985).

4. DISCUSSION

This study shows that people prefer precisely specified gambles and succumb to Ellsberg's paradox in "dual vagueness" (V-V) situations. The tendency to avoid the more vague urn and the prevalence of the classic paradox is similar in the P-V and the V-V situations. Our results indicate that P-V and V-V cases are not qualitatively different, and it is more appropriate to think of them as defining a continuum of "degree of vagueness". In both cases, the prevalence of the paradoxical pattern of choices depends primarily on the ranges of the two gambles (i.e., the relative precision and minimal imprecision of the pair) and, to a lesser degree, on the pair's common midpoint. The model fitted for the choices within a single pair also shows that the subjects' tendency to choose the more precise urn does not reflect a sharp P-V vs. V-V dichotomy. Rather, it is determined by the degree of minimal imprecision. The P-V case is just one, admittedly critical and intriguing, point on this imprecision continuum.

Several empirical regularities apply to all cases (P-V and V-V). One is the robust effect of the common midpoint: There are more choices consistent with SEUT for extreme midpoints, and a higher rate of indifference for the central value of 50. This can be attributed to the symmetry that underlies all the decisions for the 50 midpoint. In this case most, if not all, hypothetical and imagined distributions over the range are symmetric and the midpoint is the most salient focal point of the range, regardless of the range width. This, of course, can increase the likelihood of indifference between the two urns. For the extreme midpoints, 20 or 80, the most salient feature is the asymmetry between the two colors, which favors consistent choices over indifference.

Becker and Brownson (1964) suggested that subjects are sensitive to the amount of information in each urn when making their decisions, and this resonates in some of the modern behavioral work (e.g., Heath and Tversky, 1991; Keren and Gerritsen, 1999). A sensible index of the differential level of information in the two urns is obtained by considering the difference in the range width (relative precision) between the two urns. Log-linear models confirmed the relevance of the relative precision as a predictor of the *rate of paradoxical pattern*, and the logit models results confirm the importance of relative precision for predicting the *rate of vagueness avoidance within single pairs*. These results indicate, unequivocally, that as relative precision increases, vagueness avoidance (and the tendency to succumb to the famous paradox) increases. Interestingly, this robust observation contradicts one of the conclusions drawn by Curley and Yates (1985) who determined that "ambiguity avoidance did not significantly increase with the interval range *R*."

Relative precision is the most important, but not the single, predictor of the regularities in the data. We have argued that its effects are complemented by, and contingent on, the minimal imprecision in a pair, as measured by the width of the narrower range. This expectation was also confirmed by two analyses. The fit of the set-association model results for predicting the *rate of paradoxical pattern*, and of the logit model for predicting the *rate of vagueness avoidance within a single pair*, was increased by the addition of predictors that capture the effect of the minimal imprecision and its interaction with the midpoint.

Although the P-V and V-V cases are similar, they are not identical. Indeed, we have uncovered several subtle, but systematic, differences between them. The first difference highlights the distinction between the two extreme midpoints. The marginal frequencies in Tables 2.1 and 2.3 show that for the P-V case there is less vagueness avoidance (and more vagueness seeking) for the low midpoint (20), than for the high midpoint (80). On the other hand, for V-V pairs, we found more vagueness avoidance (and less vagueness seeking) for the low midpoint than for the high midpoint.[7] This difference is reflected in the results for the two consistent patterns: Although the overall level of consistency is about equal for the two types, as the midpoint increases there is a greater tendency to choose the more precise gamble in a P-V pair, whereas in the V-V case there is an opposite trend that favors less vagueness avoidance (see similar results in Curley and Yates, 1985; Einhorn and Hogarth, 1986; and Gärdenfors and Sahlin 1982, 1983).

What psychological processes can account for the particular pattern of observed differences between the P-V and V-V cases? In the P-V case the precise urn provides a clear reference point and subjects have to consider primarily the parameters of the vague urn. Its upper limit offers an attractive probability (higher than that of the precise), but this is accompanied with the risk of a lower probability (the lower limit). The subjects' behavior in these cases seems to indicate that when the precise probability is "sufficiently high" (i.e., high midpoint) they resist the temptation of the upper limit and prefer the security of the precise urn (hence, the high level of vagueness avoidance). But for low midpoints the security offered by the precise option is not sufficient, and there is a greater tendency to opt for the vague urn, presumably because of its attractive upper limit (see Stasson et al. 1993, for a similar approach).

The V-V cases do not guarantee a security level since the more precise urn is also vague. In most cases one would expect DMs to focus on the lower limits to ascertain the guaranteed security level in each urn. The higher security level would always be found in the more precise urn, hence for low midpoints DMs are likely to choose the more secure (i.e., the more precise) urn. However, the concern with security decreases for higher midpoints. Thus, vagueness avoidance decreases as the midpoint increases in the urns.

An alternative explanation for behavior in the V-V choices is that when comparing two vague urns with a common midpoint, subjects focus on the information available about the frequency of the two colors. In particular it is easy to imagine that the unknown marbles in the urn are distributed according to the same rule as the known marbles. Consider two hypothetical urns (consisting of 100 marbles) with the same (high) midpoint of 70 Red marbles. If the DM knows that in Urn A there are 50 Red marbles and 10 Blue marbles (so, the number of Reds is between 50 and 90), he/she may estimate the ratio of Red and Blue among the other (unknown) 40 marbles to also be 5:1. The DM's best guess would be that (100*5/6 =) 83 of the marbles in Urn A are Red and (100 – 83 =) 17 are Blue. Imagine that in Urn B there are 60 Red marbles and 20 Blue (so the number of Reds is between 60 and 80). The DM may infer that the ratio of the two colors is the same for the 20 unknown marbles, and his/her best guess would be that (100*3/4 =) 75 of the marbles in Urn B are Red, and the remaining (100 – 75 =) 25 are Blue. In this case, the DM would be more likely to choose the more vague Urn A, because he/she would expect it to have more marbles that are Red. If however the DM had to choose between the two urns when Blue marbles are desirable (low midpoint = 30), he/she would be more likely to pick the more precise Urn B. This is, indeed, the observed pattern in the data.

5. AN ALTERNATIVE CLASS OF MODELS

We conclude by pointing out that the DM's evaluations of vague options can also be modeled in terms of the (lower and upper) bounds of the ranges that are, typically, presented numerically and/or graphically to the subjects. Specifically, let l_i and u_i be the lower and upper bounds of range i ($i = 1, 2$), respectively, and assume that when faced with a range of probabilities, the DM "resolves its vagueness" by considering a weighted average of the two end points: $v_i = wl_i + (1 - w)u_i$, where $0 \leq w \leq 1$ indicates the relative salience of the lower bound.[8] Then the choice between the two vague lotteries can be thought of as a choice between two regular lotteries with probabilities v_1 and v_2, respectively. From a modeling point of view, focusing on the two bounds suggests a different parameterization of the problem, but the new parameters are simple linear transformations of the midpoints and ranges: $l_i = M_i - R_i/2$ and $u_i = M_i + R_i/2$. Note that if $w > 0.5$, the DM would, necessarily, exhibit vagueness avoidance, and if $w < 0.5$ he/she will appear to favor imprecision. And, if $w = 0.5$ the DM is insensitive to the range's (im)precision. Thus, we can think of w as a "coefficient of vagueness avoidance".

The two forms can be used interchangeably and most models based on the ranges can be mapped into models involving lower and upper bounds. For example, consider the probabilistic model that assumes that the tendency to choose the more precise urn depends on the difference between the two ranges:

$$\log[\Pr(\text{VA})/\Pr(\text{VS})] = (v_1 - v_2) = w(l_1 - l_2) + (1 - w)(u_1 - u_2). \qquad (3)$$

It is easy to see that $(l_1 - l_2) = -(u_1 - u_2) = \text{RELPR}/2$ (i.e., half of the relative precision). Thus, fitting model (3) amounts to fitting a model invoking only relative precision. The coefficient of vagueness avoidance, w, can be inferred from the coefficient associated with the pair's relative precision.

Although the two classes of models are statistically interchangeable, one form can be chosen over the other on the basis of its psychological plausibility, i.e., the congruence between its formulation and the assumed psychological processes underlying the subjects' behavior. We believe that the "end points" form of the model captures the psychological process involved in tasks where the subjects are required to evaluate one prospect at a time (see Budescu, Kuhn, Kramer, and Johnson, 2002; for studies of the CEs of vague lotteries). On the other hand, we think that when the DMs are asked to perform pair-wise choices between vague lotteries, as in the present study, they do not necessarily resolve the vagueness of each lottery before choosing. Rather they are more likely to rely on direct comparisons of key features of the two alternatives, such as the relative and absolute (im)precision, as indicated in our models.

This distinction is based on the lucid analysis offered by Fischer and Hawkins (1993), who distinguished between *qualitative* and *quantitative* response tasks. Quantitative tasks (pricing, rating, ranking, and matching) are, typically, compensatory and rely on quantitative strategies involving trade-offs between the various attributes that define the options. Qualitative tasks (choice, strength of preference judgments) are non-compensatory and rely on a multi-stage mix of qualitative and quantitative strategies applied in a dimension-wise fashion. The non-compensatory rules are self-terminating and do not necessarily exhaust all the attributes of the options being compared. Fischer and Hawkins (1993) have argued that in a *direct qualitative choice* where neither option strongly dominates the other, people choose the option that is superior on the more important (prominent) dimension (see also, Slovic, 1975). The more *quantitative rating task* is expected to induce a mental strategy of trade-offs between attribute values and, therefore, the more prominent attribute is not weighted as heavily. These principles apply here as well and suggest an intriguing possibility that attitudes to vagueness may vary across tasks, inducing a "reversal" of attitudes to imprecision. This hypothesis should be tested systematically in future studies.

ACKNOWLEDGMENT

This research was supported, in part, by a National Science Foundation grant (SBR-9632448). Karen Kramer's work was supported, in part, by a NIMH National Research Service Award (MH 14257) to the University of Illinois at Urbana-Champaign. The research was conducted while the first author was a predoctoral trainee in the Quantitative Methods Program of the Department of Psychology, University of Illinois at Urbana-Champaign.

NOTES

1 We will use the terms "vagueness" and "imprecision" interchangeably instead of the usual (but in our opinion, inaccurate) "ambiguity" (e.g., Budescu, Weinberg and Wallsten, 1988; Budescu, Kuhn, Kramer, and Johnson, 2002).

2 This implies that the effects of minimal imprecision can be best studied by focusing on $M = 0.5$.

3 No reference was made to a uniform distribution during the study when subjects were making their choices, so their preferences were not affected by an assumption of equal chances. This distribution was chosen because of its convenience and intuitive appeal to determine the payoffs to the subjects.

4 If subjects choose Urn I, Urn II and indifference randomly (i.e., with equal probability) and independently across the various pairs, we should observe the following distribution: (11% CP, 11% RP, 11% I, 22% C, and 44% WI).

5 We distinguished between two classes of pairs. One class consisted of all pairs where the narrower range was under 5 and the range difference was greater than 15. We expected that in all 8 pairs with these characteristics the frequency of the CP pattern would be higher than in the other (7) pairs where the ranges were closer to each other in size.

6 We also fitted all the models to the full data set including the 63 pairs. All the qualitative trends were replicated and the quantitative details varied only slightly, so we do not reproduce these results here.

7 A blue midpoint of 20 is equivalent to a red midpoint of 80, and a blue midpoint of 80 is equivalent to a red midpoint of 20, when examining the marginals. Table 2 is organized by the red midpoint.

8 This form is closely related to the one proposed by Ellsberg in his 1961 paper.

REFERENCES

Baron, J., and Frisch, D. (1994). "Ambiguous Probabilities and the Paradoxes of Expected Utility", in Wright, G. and Ayton, P. (Eds.), *Subjective Probability*, Chichester: John Wiley & Sons Ltd.

Becker, S. W., and Brownson, F. O. (1964). "What Price Ambiguity? Or the Role of Ambiguity in Decision-Making." *Journal of Political Economy*, 72, 62–73.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis*, Cambridge, MA: MIT Press.

Budescu, D. V., Kuhn, K. M., Kramer, K. M., & Johnson, T. (2002). "Modeling certainty equivalents for imprecise gambles." *Organizational Behavior and Human Decision Processes*, 88, 748–768. (Erratum in the same volume, page 1214).

Camerer, C., and Weber, M. (1992). "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity." *Journal of Risk and Uncertainty*, 5, 325–70.

Curley, S. P., and Yates, J. F. (1985). "The Center and Range of the Probability Interval as Factors Affecting Ambiguity Preferences." *Organizational Behavior and Human Decision Processes*, 36, 273–87.

Curley, S. P. and Yates, J. F. (1989). "An Empirical Evaluation of Descriptive Models of Ambiguity Reactions in Choice Situations." *Journal of Mathematical Psychology*, 33, 397–427.

Curley, S. P., Yates, J. F., and Abrams, R. A. (1986). "Psychological Sources of Ambiguity Avoidance." *Organizational Behavior and Human Decision Processes*, 38, 230–56.

Einhorn, H. J., and Hogarth, R. M. (1986). "Decision Making under Ambiguity." *Journal of Business*, 59, S225–S250.

Ellsberg, D. (1961). "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics*, 75, 643–69.

Ellsberg, D. (1963). "Risk, Ambiguity, and the Savage Axioms: Reply." *Quarterly Journal of Economics*, 77, 336–42.

Fellner, W. (1961). "Distortion of Subjective Probabilities as a Reaction to Uncertainty." *Quarterly Journal of Economics*, 75, 670–89.

Fischer, G. W., & Hawkins, S. A. (1993). "Strategy compatibility, scale compatibility, and the prominence effect." *Journal of Experimental Psychology: Human Perception and Performance*, 19, 580–597.

Gärdenfors, P. (1979). "Forecasts, Decisions, and Uncertain Probabilities." *Erkenntis*, 14, 159–81.

Gärdenfors, P., and Sahlin, N. E. (1982), "Unreliable Probabilities, Risk Taking, and Decision Making." *Synthese*, 53, 361–86.

Gärdenfors, P., and Sahlin, N. E. (1983). "Decision Making with Unreliable Probabilities." *British Journal of Mathematical and Statistical Psychology*, 36, 240–51.

Goodman, L. A. (1971a). "The Analysis of Multidimensional Contingency Tables: Stepwise Procedures and Direct Estimation Methods for Building Models for Multiple Classifications." *Technometrics*, 13, 33–61.

Goodman, L. A. (1975). "On the Relationship Between Two Statistics Pertaining to Tests of Three-Factor Interaction in Contingency Tables." *Journal of the American Statistical Association*, 70, 624–25.

Haberman, S. J. (1978). *Analysis of Qualitative Data*, New York: Academic Press.

Heath, C., and Tversky, A. (1991). "Preference and Belief: Ambiguity and Competence in Choice under Uncertainty." *Journal of Risk and Uncertainty*, 4, 5–28.

Hogarth, R. M., and Einhorn, H. J. (1990). "Venture Theory: A Model of Decision Weights." *Management Science*, 36, 780–803.

Kahn, B. E., and Sarin, R. K. (1988). "Modeling Ambiguity in Decisions under Uncertainty." *Journal of Consumer Research*, 15, 265–72.

Keren, G., and Gerritsen L. E. M. (1999). "On the Robustness and Possible Accounts of Ambiguity Aversion." *Acta Psychologica*, 103, 149–172.

Keynes, J. M. (1921). *A Treatise on Probability*, London: Macmillian.

Knight, F. H. (1921). *Risk, Uncertainty, and Profit*, Boston: Houghton Mifflin.

Kühberger, A., and Perner, J. (2003). "The Role of Competition and Knowledge in the Ellsberg Task." *Journal of Behavioral Decision Making*, 16, 181–191.

MacCrimmon, K. R. (1968). "Descriptive and Normative Implications of the Decision Theory Postulates," in Borch, K., and Mossin, J. (Eds.), *Risk and Uncertainty*, London: MacMillan.

MacCrimmon, K. R., and Larsson, S. (1979). "Utility Theory: Axioms versus 'Paradoxes,'" in Allais, M., and Hagen, O. (Eds.), *Expected Utility and the Allais Paradox*, Dordrecht, Holland: D. Reidel.

Rindskopf, D. (1990). "Nonstandard Log-Linear Models." *Psychological Bulletin*, 108, 150–62.

Roberts, H. V. (1963). "Risk, Ambiguity, and the Savage Axioms: Comment." *Quarterly Journal of Economics*, 77, 327–36.

Slovic, P. (1975). "Choice between equally valued alternatives." *Journal of Experimental Psychology: Human Perception and Performance*, 1, 280–287.

Slovic, P., and Tversky, A. (1974). "Who Accepts Savage's Axiom?" *Behavioral Science*, 19, 368–73.

Stasson, M. F., Hawkes, W. G., Smith, H. D., Lakey, W. M. (1993). "The Effects of Probability Ambiguity on Preferences for Uncertain Two-Outcome Prospects." *Bulletin of the Psychonomic Society*, 31, 624–626.

Toda, M., and Shuford, Jr., E. H. (1965). "Utility, Induced Utilities, and Small Worlds." *Behavioral Science*, 10, 238–54.

Wickens, T. D. (1989). *Multiway Contingency Table Analysis for the Social Sciences*, Hillsdale, NJ: Lawrence Erlbaum.

Yates, J. F. and Zukowski, L. G. (1976). "Characterization of Ambiguity in Decision Making." *Behavioral Science*, 21, 19–25.

Chapter 7

OVERWEIGHING RECENT OBSERVATIONS: EXPERIMENTAL RESULTS AND ECONOMIC IMPLICATIONS

Haim Levy
Hebrew University of Jerusalem

Moshe Levy
Hebrew University of Jerusalem

Abstract

We conduct an experimental study in which subjects choose between alternative risky investments. Just as in the "hot hands" belief in basketball, we find that even when subjects are explicitly told that the rates of return are drawn randomly and independently over time from a given distribution, they still assign a relatively large decision weight to the most recent observations – approximately double the weight assigned to the other observations. As in reality investors face returns as a time series, not as a lottery distribution (employed in most experimental studies), this finding may be more relevant to realistic investment situations, where a temporal sequence of returns is observed, than the probability weighing of single-shot lotteries as suggested by Prospect Theory and Rank Dependent Expected Utility. The findings of this paper suggests a simple explanation to several important economic phenomena, like momentum (the positive short run autocorrelation of stock returns), and the relationship between recent fund performance and the flow of money to the fund. The results also have important implications to asset allocation, pricing, and the risk-return relationship.

1. INTRODUCTION

Normative economic theory of decision-making under uncertainty asserts how people should behave. Experimental studies dealing with choices under conditions of uncertainty report how people actually do behave when they are faced with several hypothetical alternative prospects. In many cases there is a substantial discrepancy between the observed experimental investment behavior and the normative theoretical behavior. This discrepancy casts doubt on the validity of the theoretical economic models which rely on the normative behavior,[1] and may explain several economic "anomalies". This paper experimentally investigates and quantitatively

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 155–183.
© 2005 *Springer. Printed in the Netherlands.*

measures individuals' tendency to overweigh recent observations, and analyzes the economic implications of this behavioral phenomenon to capital markets.

The importance of overweighing recent information in capital markets is not new and has been noted by several researchers. Arrow [1982], in the context of a discussion of Kahneman and Tversky's work, highlights

> "... the excessive reaction to current information which seems to characterize all the securities and futures markets." (p. 5)

De Bondt and Thaler [1985] assert that:

> "... investors seem to attach disproportionate importance to short-run development". (p. 794)

The present paper is an attempt to experimentally quantify this phenomenon, and to estimate some of its economic effects.

The result asserting that subjects tend to interpret a series of i.i.d. observations in a biased fashion is not new. The "Law of Small Numbers" (see Tversky and Kahneman [1971]) shows that subjects exaggerate the degree to which the probabilities implied by a small number of observations resemble the probability distribution in the population. The overweighing of recent observations can be considered as a special case of the "representativeness heuristic" suggested by Tversky and Kahneman [1974], by which people think they see patterns even in truly random sequences. For example, the pioneering work of Gilovich, Vallone, and Tversky [1985] shows that basketball fans believe that players have "hot hands", meaning that after making a shot a player becomes more likely to make the next shot. This belief is very widely held despite of the fact that it is statistically unjustified (see also Albright [1993] and Albert and Bennett [2001]). Similarly, Kroll, Levy and Rapoport [1988] study an experimental financial market and show that subjects look for trends in returns even when they are explicitly told that returns are drawn randomly from a given distribution.

In a series of papers Rapoport and Budescu [1992, 1997] and Budescu and Rapoport [1994] document the phenomenon of "local representativeness", by which subjects expect even short strings within a long sequence of binary i.i.d. signals to contain proportions of the two outcomes which are similar to those in the population. Rabin [2002] presents a model with the following results: when the proportions of the two possible outcomes in a binary i.i.d. process are known, a draw of one outcome increases the belief that in the next draw the other outcome will be realized. However, when the proportions of the two outcomes are unknown, subjects infer these proportions from very short sequences of outcomes. For example, if subjects believe that an average fund manager is successful once every two years, then they believe that an observation of two successful years in a row indicates that the manager has good investment talent. As we shall see below, the experimental results we obtain conform with this assertion by Rabin.

Another related issue is that of subjective probability distortion, or the use of decision weights (see Preston and Baratta [1948], Edwards [1953], [1962], Kahneman and Tversky [1979], Tversky and Kahneman [1992], and Prelec [1998]). In most of the above studies related to decision weights, the subjects choose between two options $(x, p(x))$ and $(y, p(y))$, but the payoffs, x and y, are not given as time series. Thus, we have single-shot decisions. The subjects have to choose between two lotteries, or one lottery and a certain income. Such experiments may have limited relevance for actual investing as, in practice, investors in the market observe rates of return as time series, e.g., several years of corporate earnings, several years of mutual fund returns, etc. Therefore, the time dimension may be very important to investors, and thus should be incorporated into the analysis. In the present study, which is relevant for phenomena taken from the capital market, we present the subjects with a choice between two alternatives with given historical time series of returns, (x_t) and (y_t), where t stands for time (year, month, etc.). Subjects are told that the time series are generated **randomly from fixed distributions**, thus they should rationally attach the same weight to each observation. We test whether they indeed do so, or whether they attach more weight to the recent observations. Thus, we are dealing with the subjective distortion of probabilities as a function of the temporal sequence, not as a function of the probability itself as in the more standard frameworks of decision weights (e.g., Prospect Theory, CPT, or Quiggin's [1982] Rank Dependent Expected Utility (RDEU)), which ignore the temporal sequence.

This paper has three main goals:

(i) To experimentally test whether the most recent observations are overweighed even though the subjects are told that rates of return are i.i.d.
(ii) To estimate quantitatively the magnitude of the decision weights that the subjects attach to the most recent observations.
(iii) To analyze the economic implications of this phenomenon in terms of momentum (the positive autocorrelation of stock returns), the relationship between mutual fund performance and the flow of money to the fund, and in terms of asset pricing.

The structure of the paper is as follows: Section I describes the experiments and provides the results. In Section II we suggest a method of quantitatively estimating the overweighing of the most recent observation. Section III discusses the economic implications of the results. Section IV concludes the paper.

2. THE EXPERIMENTS AND RESULTS

In order to investigate the importance attached to recent observations we take two approaches. In the first approach we compare the choices of subjects among a set of alternative risky investments under two setups: once when the subjects are given the means and standard deviations of the normal return distributions, and once when instead they are given a time series of the returns on the alternative investments, such that the means and standard deviations are exactly as before. This approach is employed in Experiment I. In the second approach (Experiment II) we provide only the time series of the returns on the alternative investments. All subjects are given the exact same returns, but different subjects get a different time ordering of the returns. In this experiment we test directly whether the order of the returns affect the subjects' choices, i.e., whether they assign a higher decision weight to the most recent observation.

Altogether we have 287 subjects who made 415 choices (128 subjects made two choices each). The subjects are business school students and practitioners in financial markets (financial analysts and mutual funds' managers).

All of the subjects successfully completed at least one statistics course and were familiar with the normal distribution and the concept of independence over time and, in particular, with the random walk. In all the tasks where rates of return are available, the subjects were told that the rates of return were drawn randomly and independently (i.i.d.) from fixed normal distributions. Moreover, in all tasks, the subjects were explicitly told that the next realized rate of return (which is relevant for their investment) is drawn randomly and independently from the corresponding normal distribution. These facts were emphasized in the instructions to the subjects.

2.1. Experiment I

In this experiment we have 128 subjects, 64 of them third-year undergraduate business students and 64 of them mutual fund managers and financial analysts whom we call "practitioners".[2] All of the subjects had the questionnaire for a relatively long period of time (at least a week), hence, they could make any needed calculation and make the choices without any time pressure.

The experiment, as many other experiments, did not involve any real financial reward or financial penalty to the subjects, which may constitute a drawback. However, Battalio, Kagal and Jiranyakul [1990] have shown that experiments with and without real money differ in the magnitude of the results but not in their essence. Harless and Camerer [1994] have shown that when real money is involved, the variance of the results decreases. Thus, it seems that the absence of money does not drastically change the results.[3] Yet, because no real money was involved one always suspects that the subjects may fill out the questionnaire randomly without paying close attention to the various choices. Fortunately, this was not the case, as shown below.

In this experiment the subjects are requested to complete two tasks. In Task I they are presented with five mutual funds and are told that the return distribution for each of the funds is normal, with given parameters, as presented in Table 1. The subjects are asked the following question: "*Assuming that you wish to invest in only one mutual fund for one year, which fund will you select?*".

In Task II the subjects are again asked to choose one of five mutual funds, and again they are told that the return distributions are normal and that returns are independent over time. However, in this task the subjects are given the last 5 annual return observations of each fund instead of the fund's mean and standard deviation (see Table 2). The returns in Task II are constructed such that the means and standard deviations of each fund are exactly identical to those in Task I.

Table 1. Means and Variances of Returns in Experiment I Task I

	Fund				
Fund	*A*	*B*	*C*	*D*	*E*
Mean	12.40%	10.40%	12.60%	10.60%	14.00%
Standard Deviation	18.87%	15.82%	13.15%	8.33%	17.17%

Table 2. Annual Rates of Return in Experiment I Task II

	Fund				
Year	*A*	*B*	*C*	*D*	*E*
1	14%	14%	14%	12%	14%
2	45%	18%	35%	20%	45%
3	−10%	−10%	2%	14%	−2%
4	15%	34%	15%	12%	15%
5	−2%	−4%	−3%	−5%	−2%

2.1.1. Results

Table 3 reports the choices in Tasks I and II corresponding to the 5 mutual funds. As there are no significant differences in the choices of the students and the practitioners, we report here only the aggregate results. The main results are as follows:

1) The choices are not random: we test whether the subjects filled out the questionnaire randomly to quickly "get it over with", by employing the Chi-square goodness-of-fit test. To illustrate, in Task I, the subjects had to choose one out of five mutual funds. If the subjects select the fund randomly, we expect on average $128/5 \cong 26$ subjects choosing each fund. Using the observed choices, and the expected number of choices of each fund, we employ the Chi-square goodness-of-fit test with four degrees of freedom. We obtain in Task I a sample statistic of $\chi_4^2 = 129.3$, when the 1% critical value is 13.3. In Task II the sample statistic is $\chi_4^2 = 100.4$. Thus, both the sample statistics are substantially larger than the corresponding critical value, hence regarding each of the two tasks, the hypothesis that the subjects made a random choice is strongly rejected. Thus, it seems that despite the fact that there was no financial reward/penalty, most of the

Table 3. Results of Experiment I

Fund	*Task I*	*Task II*
A	3	1
B	0	2
C	**45**	**29**
D	**66**	**34**
E	**14**	**62**
Total	128	128

subjects in our experiment made a choice according to their preference and not randomly.

2) When the return distributions are normal, the mean-variance rule is well known to be optimal under risk aversion (see Tobin [1958]). Moreover, it is also optimal under the Markowitz [1952b] reverse S-shape value function, and under the CPT S-shape value function (see Levy and Levy [2003]). Thus, it is natural to examine the mean-variance efficiency of the subjects' choices. Figure 1 presents the five funds in the mean-standard deviation space. It can easily be seen that funds {D, C, E} are mean-variance efficient and funds {B, A} are inefficient (see Figure 1). The inefficient funds, A and B, together were selected by only 3 out of 128 subjects in both Task I and in Task II.

   Thus, we have the encouraging results showing that 98% of the choices are mean-variance efficient. Thus, "framing" the choices in terms of μ-σ or in terms of annual rates of return does not affect the percentage of the efficient choices, which remains very high.

3) In Task I, the choices were mainly of C and D and not E. Looking at Table 1, we see that Fund E has a little higher expected return than Fund C but much larger standard deviation. It is possible that this risk-return tradeoff induces most of the subjects to select Funds C and D and not Fund E.[4]

4) The importance of the time sequence: Because rates of return are i.i.d., theoretically framing the choices in two ways should not affect the choices. This is not the case, because choices have been dramatically changed *within* the efficient set. While in Task I, choices C and D were very popular, in Task II there is a substantial shift from Funds C and D to Fund E, which became the most popular choice with almost half of the subjects selecting it (compared to less than 11% selecting E in Task I). Focusing on the shifts in choices in Task I and II within the efficient set, we conducted a χ^2 test to examine whether the shifts are significant. We obtain a sample statistic of 44.0 while the 1% critical value is only

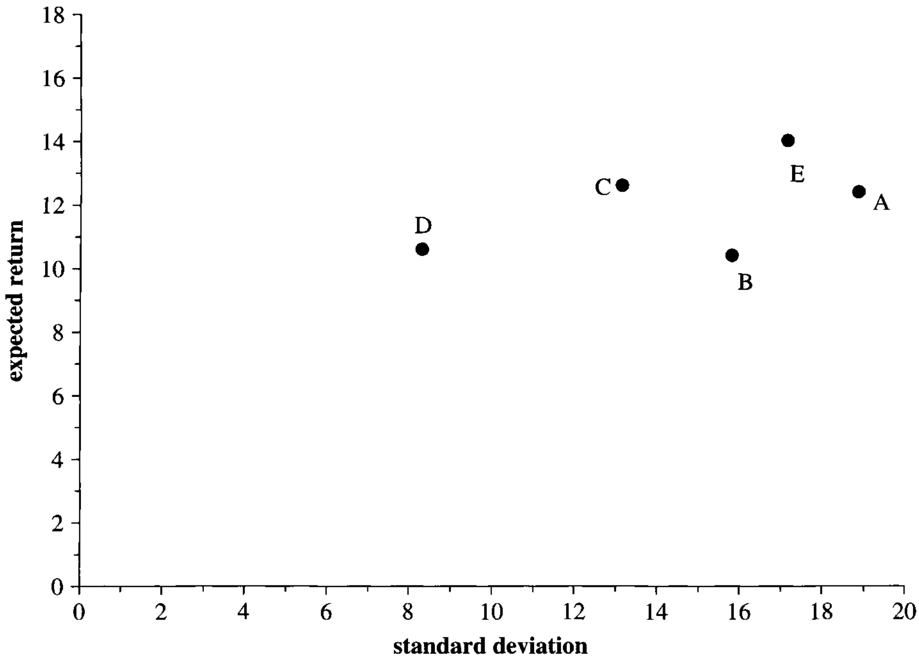


Figure 1. The Funds in Experiment I.

9.2, hence the change in choices is highly significant. There is a wide range of possible explanations as to why subjects switched from C and D to E. However, a close look at the rates of return in Table 2 reveals two important characteristics: in four out of five years, E shows a higher rate of return than D, and more importantly, in the last two years the returns on E are better than the returns on D. Though this information is irrelevant under the i.i.d. property, it seems that the subjects made use of this information. This experimental finding, i.e., switching to the fund with the highest short-term performance (e.g., the performance in the last two years) conforms with the results of Kroll Levy and Rapoport [1988], with Rabin [2002], and with Arrow's [1982] assertion of an "excessive reaction to current information". Thus, despite of the randomness and independence over time of rates of return, investors switch between funds based on short-term performance.

The comparison of the rates of return on E and C is a little more involved: in two years they have the same rates of return, in two years E is better and in one year C is better (see Table 2). However, in the last year, which probably was more important to the subjects, E is better, even though by only 1%. Thus, the "seemingly" superiority of E over D is stronger than the superiority of E over C, which may explain why a larger shift occurred from D to E than from C to E (see Table 3).

Regardless of whether all rates of return affect choices, or only the last one or two observations affect the switch in the choices, one thing is clear: the subjects either misperceive randomness and overweigh recent outcomes, do not believe the i.i.d. information or do not believe the normality.

To sum up, the subjects create patterns, and draw conclusions from the irrelevant order of the historical rates of return. This is because, theoretically, under the i.i.d. information, and the data of Tasks I and II, no switch in choices should occur.

Finally, it is possible that in Task II the subjects do not assign relatively large decision weights to the last 1–2 observations, but rather employ some other complicated decision rules, e.g., "select the fund with the highest possible gain and the smallest possible loss" (like Fund E), or select the mutual fund based on mean, variance and, say, skewness, though skewness is irrelevant under normal distribution. To address this issue, in Experiment II we refine the analysis regarding the role that recent rates of return play in decision making. This experiment is very simple, and more directly attempts to figure out the role of the most recent observation on the decision making process.

2.2. *Experiment II*

The subjects participating in this experiment are 159 undergraduate business school students. The subjects have to choose between only two investment alternatives. As in Task II of the first experiment, the last five returns of each of these alternatives are presented to the subjects, and the subjects are told that the returns are drawn randomly and independently over time from normal distributions. We divide the subject population into two groups, and each subpopulation is given a different version of the questionnaire. One subpopulation is presented with two investment alternatives exactly identical to Funds D and E of Task II in Experiment I (see Questionnaire 1 in Table 4). The other subpopulation is presented with the same

Table 4. Experiment II

	Questionnaire 1		*Questionnaire 2*	
Year	*D*	*E*	*D*	*E*
1	12%	14%	–5%	–2%
2	20%	45%	12%	15%
3	14%	–2%	14%	45%
4	12%	15%	12%	–2%
5	–5%	–2%	20%	14%

Table 5. Results of Experiment II (in percent)

Questionnaire 1 (n = 66)		*Questionnaire 2 (n = 93)*	
D	29%	D	45%
E	71%	E	55%
Total	100%	Total	100%

set of returns for each fund, but the time ordering of the returns are different (see Questionnaire 2 in Table 4). Specifically, Questionnaire 2 is designed such that if more weight is assigned to recent returns Fund D becomes more attractive. Note that if an equal weight of 0.2 is attached to each observation the results in the two questionnaires should be roughly the same. However, if one assigns a relatively large decision weight to the last one or two years, then E is improved relative to D in Questionnaire 1, while D is improved relative to E in Questionnaire 2.

2.2.1. Results

The results of Experiment II are reported in Table 5. Only 29% of the choices were D in Questionnaire 1 versus 45% in Questionnaire 2. A χ^2 test with one degree of freedom reveals that the differences are significant with $\alpha = 5\%$, with a sample statistic of chi-square of 4.35, while the critical value is $\chi^2_{(1,5\%)} = 3.84$. Thus, there is a significant change, albeit not a very strong one, in choices in favor of the fund with the relatively good performance in the last two years. This is so despite the fact that the returns are exactly identical in the two questionnaires. Thus, Experiment II clearly reveals that the last two observations have an important role in determining choices.

We advocate in this paper that probability is distorted in a particular way, emphasizing the last one or two observations. This is in contradiction to the CPT and RDEU probability distortion. For example, by the CPT distortion, probabilities should be distorted in the same way in both questionnaires 1 and 2, overweighing the extreme probabilities of –2% and 45% in Fund E and –5% and 20% in Fund D, regardless of the sequence of appearance of these observations. Therefore, according to CPT the choices should not change across the two questionnaires. This is not the case in our experiment, indicating that the CPT weighing function may be inappropriate for time series returns, as observed in the capital market.

Finally, as not all subjects choose E in Questionnaire 1, and not all subjects choose D in Questionnaire 2, it is obvious that the decision weight assigned to the last 2 observations is less than 100%, and some of the investors may perceive randomness correctly. In many cases some complicated decision rules are probably employed. Yet, it is enough that some investors overweigh recent observations to create several important economic phenomena. In the next section we attempt to quantitatively estimate the overweighing of the most recent observation.

3. ESTIMATING THE DECISION WEIGHTS

In this section we estimate the decision weights corresponding to temporal sequence data, which is conceptually different than the decision weights in single-shot lottery-type situations, as suggested by Prospect Theory and other models. In order to analyze the shift in choices and the decision weights applied to the most recent observations one needs to make some assumptions regarding preferences. We start with general assumptions about the preference class (e.g., risk aversion), and then we refine the analysis by employing specific commonly acceptable utility/value functions.

Under the assumptions of normal rate of return distributions and risk aversion, the optimal investment rule which is consistent with von-Neumann and Morgenstern [1953] expected utility maximization is the Markowitz [1952a] mean-variance rule (see Tobin [1958] and Hanoch and Levy [1969]). In this case the mean-variance rule coincides with Second degree Stochastic Dominance (SSD). When rates of return are drawn randomly and independently from normal distributions then the best estimates of the mean and variance are the corresponding sample statistics, assuming each observation has an equal weight of $1/n$, n being the number of observations. Our findings imply that in expected utility calculation decision weights, $w(p(x))$, are employed rather than the objective probabilities, $p(x)$, where $w(p(x)) > p(x)$ for the last one or two observations. In this section, we attempt to estimate $w(p(x))$. We take two approaches. The first is the Stochastic Dominance approach which allows us to place an upper bound on $w(p(x))$. In the second approach we assume various typical utility functions and obtain estimates of the median $w(p(x))$ in the population.

Several studies highlight the importance of overweighing the most recently observed return (see Kroll, Levy and Rapoport [1988], Chevalier and Ellison [1997], and Rabin [2002]). The results of Experiment I support this view. An increase in the decision weight of the most recent return explains the shift in choices from Funds C and D in Task I to Fund E in Task II. In contrast, the penultimate observation is not overweighed much, because such overweighing would have implied a shift in the choices to Fund B in Task II, a shift which did not occur (in the 4th year, the rate of return on Fund B was 34%, much higher than the 15% of Fund E, see Table 2). Thus, from the rates of return data and from the specific shift in choices, we conclude that the overweighing of the most recent return is probably the main factor, albeit not the only factor, inducing the shifts in choices observed in our experiments. Therefore, in what follows we analyze the subjects' choices by making the assumption that for the 5th year $w_5(p) > p = 0.2$ and for all the other four years $w_i(p) = \dfrac{1 - w_5(p)}{4} < 0.2$, where $w_i(p)$ is the decision weights corresponding to year i ($i = 1, 2, 3$ and 4).[5] As we employ Stochastic Dominance rules in estimating $w_5(p)$, let us first define these rules.

3.1. Stochastic Dominance Approach

a. Definitions

Consider the funds in Experiment I. When decision weights are employed such that the most recent observation is overweighed Fund E becomes more attractive relative to the other funds. In employing the stochastic dominance approach we ask the following question: what should $w_5(p)$ be such that E will stochastically dominate the other funds? The answer to this question gives an upper bound on $w_5(p)$, because if all subjects assign a weight equal or greater than this critical value of $w_5(p)$ to the fifth observation, they should all prefer Fund E in Task II. We investigate the critical value of $w_5^*(p)$ by employing First and Second degree Stochastic Dominance rules. These decision rules are defined below.

i) *First degree Stochastic Dominance (FSD)*:
Distribution F dominates distribution G for all increasing utility functions if and only if $F(x) \le G(x)$ for all x, and there is a strict inequality for some value x_0. Namely,

$$F(x) \le G(x) \text{ for all } x \quad \Leftrightarrow \quad E_F U(x) \ge E_G U(x) \text{ for all } U, \text{ with } U' \ge 0 \quad (1)$$

ii) *Second degree Stochastic Dominance (SSD)*:
Define F and G as before, and U is a concave utility function ($U' \ge 0$, $U'' \le 0$). Then,

$$\int_{-\infty}^{x} [G(t) - F(t)]\mathrm{d}t \ge 0 \quad \text{for all } x \quad \Leftrightarrow \quad E_F U(x) \ge E_G H(x) \qquad (2)$$

for all U with $U' \ge 0$, $U'' \le 0$.

Thus, if risk aversion is assumed, SSD can be employed.[6,7] Though we focus in this study on SSD (i.e. risk aversion), experimental studies show that risk-seeking also exists in preferences (see Friedman and Savage [1948], Markowitz [1952b], and Kahneman and Tversky [1979]). In particular, Levy and Levy [2001] show that at least 50% of the subjects are not risk averse. Hence, if preferences other than risk-aversion are assumed, the corresponding Stochastic Dominance criteria should be employed. For example, the Prospect Stochastic Dominance (PSD)[8] rule corresponds to the class of all Prospect Theory S-shape value functions, and the Markowitz

Stochastic Dominance (MSD)[9] rule corresponds to the class of all reverse S-shape value functions as suggested by Markowitz [1952b]. Here we focus on risk-aversion and the SSD rule.[10]

b. Implementation of the Stochastic Dominance Rules

First, note that Fund E dominates Fund A by FSD with the objective probabilities $p_i = 0.2$ (see Table 2). Any overweighing of the fifth year probability, $w_5 > 0.2$, does not affect this FSD dominance.

Now let us turn to the more interesting case of Funds D and E, as given in Table 2 (and Questionnaire 1 in Table 4). Figure 2a provides the cumulative distributions of these funds when an equal probability of $p = 0.2$ is assigned to each observation, as should be done with a random sample composed of five independent observations. As we can see, the two cumulative distributions F_D and F_E intersect, so by equation (1) neither fund dominates the other by FSD.

Also, as can be seen from Figure 2a, $\int_{-\infty}^{-2\%} (F_E(x) - F_D(x))\, dx < 0$, hence D does not dominate E by SSD, (see equation (2)) and $\int_{-\infty}^{12\%} (F_D(x) - F_E(x))\, dx < 0$, hence E does not dominate D by SSD. Also, there is no dominance by the Mean-Variance rule. Thus, it is very reasonable that with the objective probability $p = 0.2$, some risk averters (SSD or Mean-Variance decision makers) will select Fund D and some would select Fund E. Let us now demonstrate how with $w_5(p) > 0.2$, Fund E may be considered better by some risk averters, and beyond some critical value $w_5^*(p)$ Fund E even dominates Fund D by SSD, i.e., should be preferred by *all* risk-averters (SSD dominance).

Assume that $w_5(p) > 0.2$. As the most recent observation is also the smallest return for both Funds D and E, this overweighing corresponds to an increase of the first positive area (see Figure 2b) and the negative area decreases (recall that increasing $w_5(p)$ induces a decrease in the other decision weights $w_i(p)$, $i = 1, 2, 3, 4$). Thus, there is some critical value $w_5^*(p)$ such that the negative area will be equal to the first positive area, hence E will dominate D by SSD. To find the critical $w_5^*(p)$ the following condition must be fulfilled (i.e., equating the first areas enclosed between the two cumulative distributions):

$$w_5^*(p)(-2 - (-5)) = \frac{1 - w_5^*(p)}{4}(12 - (-2))$$

or: $3w_5^*(p) = (1 - w_5^*(p))(14/4)$. Hence,

$$12w_5^*(p) = 14 - 14w_5^*(p)$$

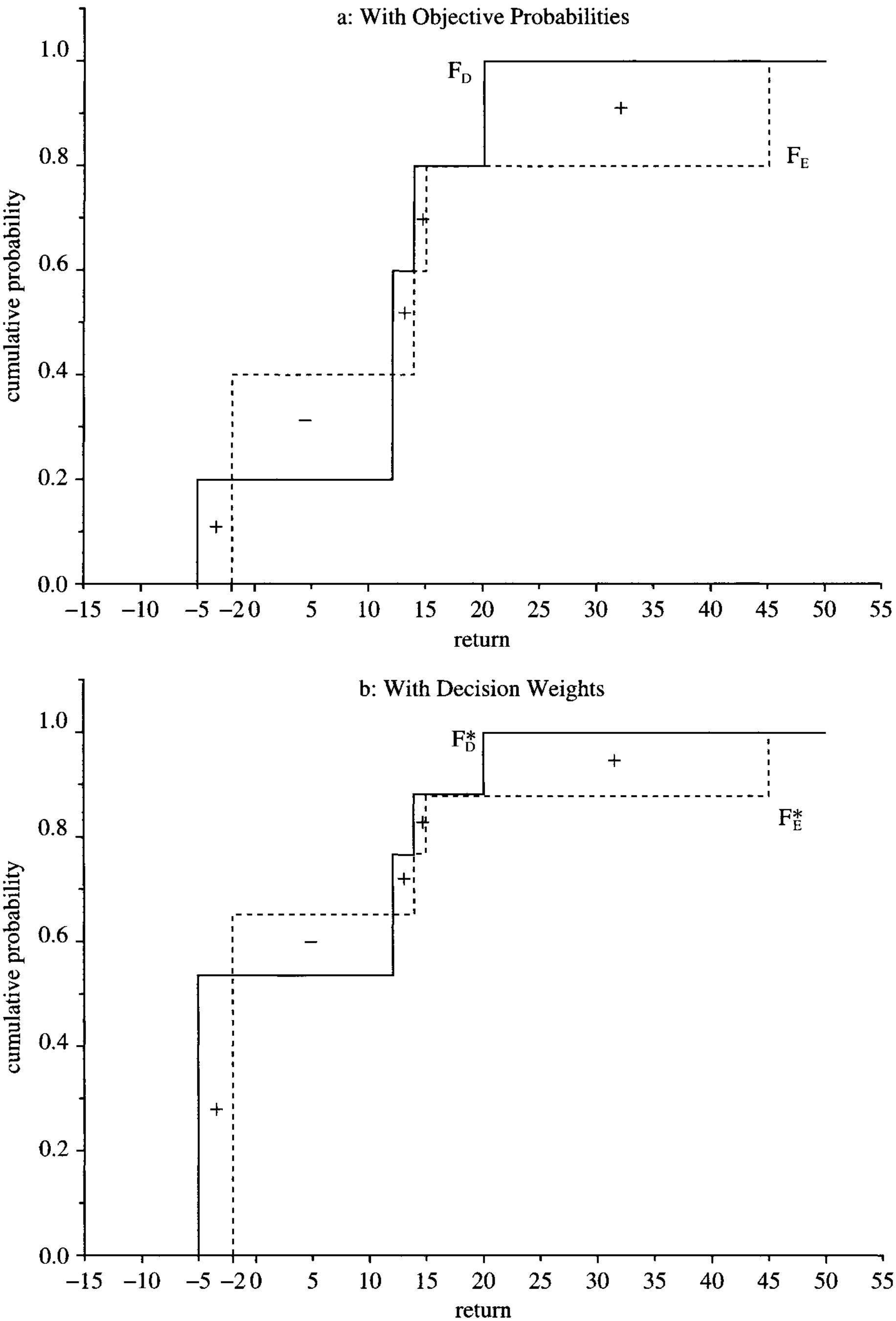


Figure 2. The Cumulative Distributions of D and E.

which finally yields,

$$w_5^*(p) = \frac{14}{26} = 0.54$$

and the other subjective probabilities are:

$$w_i(p) = \frac{12}{26 \cdot 4} = \frac{3}{26} \quad (\text{for } i = 1, 2, 3, 4)$$

Figure 2b draws the cumulative distributions of E and D with these decision weights, denoted by F_D^* and F_E^*. As can be seen from the figure, with the decision weights the negative area is equal to the first positive area (because $(-2-(-5)) \cdot \frac{14}{26} = (12-(-2))\frac{3}{26}$). Because all other areas enclosed between the two distributions are positive, we have: $\int_{-\infty}^{x} (F_D^*(t) - F_E^*(t))\, dt \geq 0$, for all x, (with at least one strict inequality for some x), when the superstar emphasizes that these are subjective cumulative distribution with decision weights rather than the objective cumulative distributions, F_D and F_E (compare Figure 2a and 2b). Thus, with $w(p) \geq w^*(p)$, Fund E (subjectively) dominates Fund D by SSD, and all *risk averters* are expected to choose E. Hence, with risk aversion $w_5^*(p) = 0.54$ is an upper bound on the fifth year decision weight. If all subjects were risk-averse and had $w(p) \geq w^*(p)$, they would all choose Fund E in Task II. As 62 out of the 128 subjects selected Fund E and 34 still selected Fund D, we conclude that either these 34 subjects are not risk averse, or that for these subjects $w_5 < 0.54$.

Using the same technique in the comparison of Funds C and E, we find that $w_5^*(p) = 0.5$, i.e., for $0.2 < w_5(p) < 0.5$, some subjects may switch from C to E, and for $w_5(p) \geq 0.5$ all risk averters are expected to shift from C to E. For the sake of brevity, we do not provide the detailed calculation of $w_5^*(p)$ corresponding to C and E.

c. Relaxing the Risk-Aversion Assumption

The Second degree Stochastic Dominance approach is non-parametric, hence it does not make assumptions about the specific utility function. This approach provides us with an upper bound on the decision weight in the sense that with risk aversion the experimental results reveal that it is not possible that *all* subjects have $w(p) \geq w^*(p)$. Alternatively, it is possible that not all subjects are risk averters. Thus, in what follows we do not confine ourselves to concave preferences. In particular, we discuss Prospect Theory's S-shape preferences and Markowitz's reverse S-shape preferences (about these two preference types, see footnote 10).

c1. PSD and MSD

So far we employ SSD in the comparison of E and D. The experimental results can also be explained with non-concave preferences. Employing MSD and PSD reveals the following results: Fund E dominates Fund D by MSD (for the MSD rule see footnote 9). This dominance holds for the objective probabilities, $p_i = 0.2$, as well as for any overweighing of the most recent observation, $w_5 > 0.2$. On the other hand, neither E nor D dominate one another by PSD (see footnote 8 for the PSD rule), and this is true both for the objective probabilities and for any overweighing $w_5 > 0.2$. Therefore, the results of Table 3 regarding Funds D and E conform either with risk-aversion and an increase in w_5, or alternatively, with no overweighing and with about 2/3 of the choices (62 out of 96) conforming with MSD, i.e., with a reverse S-shape value function.

3.2. Direct Estimation of $w_5(p)$

Assuming a specific utility function enables a direct estimation of $w_5(p)$. Surprisingly, the estimates obtained under different utility functions are very similar, which makes the results quite robust. Below we describe the estimation of $w_5(p)$ under the assumption of a logarithmic utility function, a linear utility function, the Prospect Theory S-shape value function suggested by Kahneman and Tversky [1992], and the reverse S-shape value function suggested by Markowitz [1952b].

In applying the direct estimation approach it is beneficial to employ Questionnaire 2 of Experiment II, because here the subjects' choices were split almost evenly between the two funds (see Table 5). This allows us to obtain an estimate of the median $w_5(p)$, as detailed below.

Logarithmic Utility Function

Consider Funds D and E of Questionnaire 2 in Experiment II (see Table 4). What is the value of $w_5(p)$ which makes an individual with logarithmic preferences indifferent between these two funds? The answer is given by the solution to:

$$\begin{aligned}&w_1 \log(W(1-0.05)) + w_2 \log(W(1+0.12)) + w_3 \log(W(1+0.14))\\&\quad + w_4 \log(W(1+0.12)) + w_5 \log(W(1+0.20))\\&= w_1 \log(W(1-0.02)) + w_2 \log(W(1+0.15)) + w_3 \log(W(1+0.45))\\&\quad + w_4 \log(W(1-0.02)) + w_5 \log(W(1+0.14))\end{aligned}$$

where W is the initial wealth, and w_i is the decision weight of observation i. Recalling that in our framework $w_i = \dfrac{1-w_5}{4}$ for $i = 1, 2, 3, 4$, and noticing that W cancels out, we have:

$$\left(\frac{1-w_5}{4}\right)[\log(0.95) + \log(1.12) + \log(1.14) + \log(1.12)] + w_5 \log(1.20) =$$

$$\left(\frac{1-w_5}{4}\right)[\log(0.98) + \log(1.15) + \log(1.45) + \log(0.98)] + w_5 \log(1.14),$$

which yields:

$$w_5 = \frac{[\log(0.98) + \log(1.15) + \log(1.45) + \log(0.98)] - [\log(0.95) + \log(1.12) + \log(1.14) + \log(1.12)]}{[\log(0.98) + \log(1.15) + \log(1.45) + \log(0.98)] - [\log(0.95) + \log(1.12) + \log(1.14) + \log(1.12)] + 4(\log(1.2) - \log(1.14))}$$

or:

$$w_5 = 0.44.$$

Suppose that different individuals with this specific type of preferences overweigh the fifth observation differently. Any individual with log utility who assigns a weight higher than 0.44 to the fifth observation prefers Fund D over E, and any individual who assigns a weight lower than 0.44 to the fifth observation prefers Fund E. Assuming a logarithmic utility function, the fact that approximately half of the subjects chose Fund D and half chose Fund E (see Table 5) implies that the median w_5 in the population is approximately 0.44.[11]

Linear Utility

A similar analysis under the assumption of linear utility (risk neutrality) leads to:

$$\left(\frac{1 - w_5}{4}\right)[0.95 + 1.12 + 1.14 + 1.12] + w_5 1.20$$

$$= \left(\frac{1 - w_5}{4}\right)[0.98 + 1.15 + 1.45 + 0.98] + w_5 1.14$$

Rearranging we obtain:

$$(1 - w_5)\left(\frac{4.56 - 4.33}{4}\right) = (1.20 - 1.14)w_5,$$

which yields $w_5 = 0.49$.

Prospect Theory Value Function

Tversky and Kahneman [1992] suggest that preferences are described by the following value function:

$$V(x) = \begin{cases} x^{\alpha} & \text{if } x > 0 \\ -\lambda(-x)^{\beta} & \text{if } x < 0 \end{cases}$$

where x is the change in wealth, and α, β, and λ are constants which Tversky and Kahneman experimentally estimate as: $\alpha = 0.88$, $\beta = 0.88$, and $\lambda = 2.25$. With this value function, an indifference between Funds D and E implies:

$$\left(\frac{1-w_5}{4}\right)[-2.25(0.05)^{0.88} + (0.12)^{0.88} + (0.14)^{0.88} + (0.12)^{0.88}] + w_5(0.20)^{0.88} =$$

$$\left(\frac{1-w_5}{4}\right)[-2.25(0.02)^{0.88} + (0.15)^{0.88} + (0.45)^{0.88} - 2.25(0.02)^{0.88}] + w_5(0.14)^{0.88}.$$

Rearranging we obtain:

$$(1 - w_5)(0.1349 - 0.0814) = (0.2426 - 0.1773)w_5$$

or:

$$w_5 = 0.49.$$

Markowitz Value Function
Assuming the Markowitz reverse S-shape value function also yields similar results: taking

$$V(x) = \begin{cases} x^{\alpha} & \text{if } x > 0 \\ -\lambda(-x)^{\beta} & \text{if } x < 0 \end{cases}$$

with $\alpha = \beta = 1.1$ and $\lambda = 2.25$ yields $w_5 = 0.54$.

Thus, under different utility functions we obtain similar estimates for w_5, in the range 0.44–0.54. These values are also close to the upper bound $w_5 = 0.54$ obtained in the Stochastic Dominance approach (and, again, recall that the upper bound states that it is not possible for *all* risk-averse subjects to exceed the bound). The overweighing we find is quite substantial – the decision weight is more than twice the objective probability! In the next section we discuss the economic implications of this phenomenon.

4. ECONOMIC IMPLICATIONS

The experimentally observed overweighing of recent observations can provide a simple explanation for several phenomena observed in the capital market, and can induce substantial mispricing of financial assets. In this section we discuss some of these implications.

a. Mutual Fund Performance and the Flow of Money to Funds

Several researchers have demonstrated that the performance of mutual funds does not exhibit statistically significant trends. Sharpe [1966] finds that a high performance of a mutual fund in one period does not increase the probability of a high performance in the next period. Beckers [1997] also finds that luck is the major factor explaining mutual fund performance. Samuelson [1989] who advocates market efficiency asserts:

> "Those lucky money managers who happen in any period to beat the comprehensive averages in total return seem primarily to have been merely lucky. Being on the honor roll in 1974 does not make you appreciably more likely to be on the 1975 honor roll." (Samuelson [1989] p. 4).

Yet, despite of these findings Chevalier and Ellison [1997] have shown that a relatively high performance of a fund in the recent period (recent 1–2 years), increases the inflow of cash to the fund (see also Agarwal, Daniel and Naik [2003], Luo [2003], and Sapp and Tiwari [2003]). Thus, there is a positive relationship between flow and performance, even though the empirical studies reveal that mutual funds managers do not have "timing" and "selectivity" ability, and performance is rather random.

The overweighing of recent returns provides a very simple explanation for this phenomenon. It seems that investors attach a high weight to the most recent returns of funds, and tend to classify funds with relatively high recent returns as "good", although this is statistically baseless. Thus, investors' money "chases" the funds that have the best recent returns, even though this is economically unjustified and creates inefficient allocations and unnecessary transactions (and transaction costs).

b. Momentum and Price Reversals

The common belief for years, at least among academics, was that the market is efficient and, in particular, that past rates of return cannot be employed to earn abnormal profit (weak form efficiency, see Fama [1970], [1991]). Recent studies question this assertion. It has been found that for short lags (6–12 months) there is a positive return auto-correlation, and for longer term lags (3–5 years) there is negative return auto-correlation.[12]

The significant auto-correlation of rates of return implies some predictability of stock returns, and thereby challenges the notion of weak market efficiency. Jegadeesh [1990], Jegadeesh and Titman [1993], and Levy and Lim [1998], investigate whether exploiting the return auto-correlations can lead to significant abnormal returns. The results are mixed: Jegadeesh, and Jegadeesh and Titman find that significant abnormal returns are attainable, while Levy and Lim find that there are no abnormal returns after transaction costs.

Several researchers have suggested various different explanations for the pattern of return autocorrelations.[13] The overweighing of recent returns offers a simple explanation for the return autocorrelation pattern. To see this, consider the following oversimplified example. Suppose that there are two assets with i.i.d. random normal returns: $x \sim N(\mu_1, \sigma_1)$ and $y \sim N(\mu_2, \sigma_2)$. For simplicity, assume that $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$. Thus, without misperception of randomness, the future returns are i.i.d. Now, suppose that $x_t = -10\%$ and $y_t = +5\%$ where t is the most recent observation. If investors believe the i.i.d. property of rates of return, the next rate of return will be indeed random. However, this is not the case with an overweighing of recent observations and a misperception of randomness. The relatively large decision weight assigned to the most recent observation will induce some of the investors to sell x and to buy y, which by itself would create a negative rate of return for x and a positive rate of return on y, creating further excess supply for x and demand for y, and so forth. Thus, the overweighing of recent returns may create a positive feedback loop leading to the empirically observed short-term momentum.

This momentum, however, cannot continue forever. Because the firms' earnings and dividends are not affected by investors' perceptions, the dividend component in the rate of return becomes smaller as the price becomes higher. Namely, if the momentum continues for a long period of time, the price of y will be very high, the dividend component in the rate of return will be relatively small and the rate of return will tend to decrease. A price reversal may therefore be obtained, again, reinforced by the positive feedback induced by the overweighing of recent returns (and the exact opposite happens to asset x). Thus, the short-term momentum is followed by a longer-term reversal, explaining the empirically observed U-shaped-auto-correlation pattern.[14]

c. Asset Allocation, Pricing, and Beta

The overweighing of recent returns as experimentally estimated in this paper has a dramatic effect on asset allocation, asset pricing, and the risk-return relationship. To illustrate this claim we perform the following analysis. We randomly select 10 stocks from the CRSP database, and record the last five annual returns on these stocks. Then, based on the objective probabilities (with a an equal weight of 0.2 for each observation) we calculate the objective means, standard deviations, and covariances, and calculate the mean-variance optimal portfolio of these assets based on the objective probabilities.[15] Next, we repeat this analysis, but this time with decision weights as we find in our experiment. We assign a decision weight of $w_5 = 0.45$ to the most recent observation (somewhere in the middle of the range that we estimate for w_5 in section II), and $w_1 = w_2 = w_3 = w_4 = 0.55/4$. We recalculate the means, standard deviations, and covariances, and the optimal mean-variance portfolio based on these decision weights. To measure the effect of the decision weights on asset allocation and pricing we compare the portfolio weight of each of the assets in these two optimal portfolios and report the relative difference:

$$\Delta_i \equiv \frac{x_i^{dw} - x_i^{ob}}{x_i^{ob}} \tag{3}$$

where x_i^{ob} is the proportion of asset i in the objective optimal portfolio, and x_i^{dw} is the proportion of asset i in the optimal portfolio based on decision weights.[16]

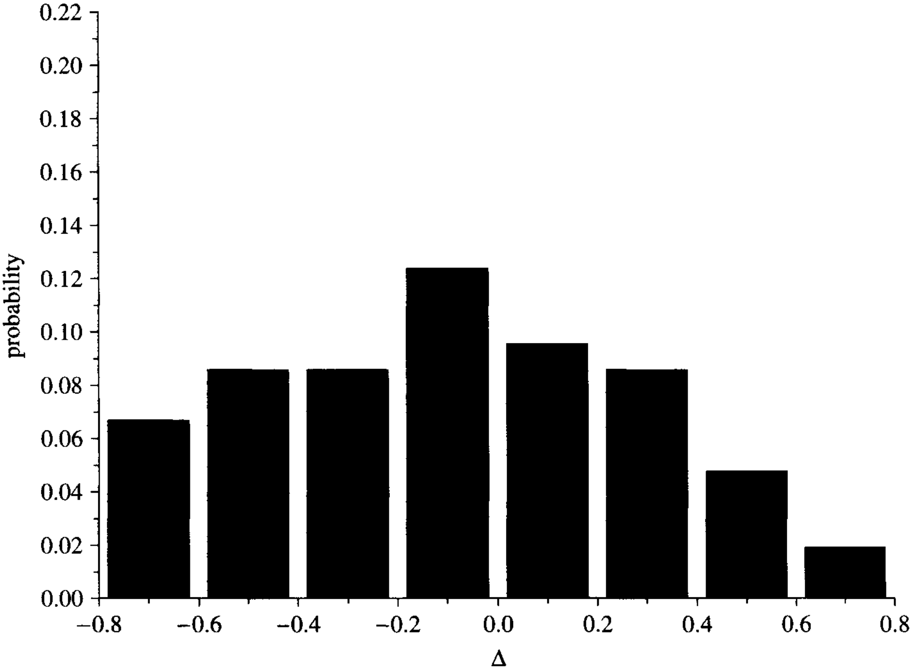


Figure 3. Deviations from Objective Portfolio Proportions.

The deviation Δ *is defined as follows:* $\Delta_i \equiv \frac{x_i^{dw} - x_i^{ob}}{x_i^{ob}}$, *where* x_i^{ob} *is the proportion of asset i in the optimal portfolio derived with the objective probabilities (1/n), and* x_i^{dw} *is the proportion of asset i in the optimal portfolio derived with the experimentally found decision weights.*

We repeat this procedure for 100 independent samplings of 10 stocks to obtain 1000 observations of Δ. Figure 3 shows the distribution of Δ. The mean absolute deviation is 63% ($E|\Delta| = 0.63$). This average absolute deviation of 63% in portfolio weights implies an average absolute deviation of 63% in asset pricing! Thus, it is evident that the overweighing of the most recent return induces substantial deviations in asset allocation and pricing. These deviations correspond to large economic inefficiencies and utility loss. Let us turn now to the effect of overweighing recent returns on the risk-return relationship.

The Sharpe [1964]-Lintner [1965] CAPM risk-return relationship is given by:

$$\mu_i = r + (\mu_m - r)\beta_i \tag{4}$$

where μ_i is the expected return on asset i, r is the riskless interest rate, μ_m is the expected return on the market portfolio, and β_i is the risk of asset i. The empirical test of the CAPM is given by:

$$\bar{R}_i = \gamma_0 + \gamma_1 \hat{\beta}_i + \varepsilon_i \tag{5}$$

where $\bar{R}_i$ is the empirically measured average return, and $\hat{\beta}_i$ is empirically measured beta, both of which are calculated with ex-post data. If the CAPM precisely holds and there are no sample errors we expect to find a perfect fit with $R^2 = 100\%$ in the regression given by eq. (5). Moreover, Roll [1977] shows that in an empirical study where beta is calculated against *any ex-post* mean-variance efficient portfolio (with short-sells), eq. (5) yields a perfect fit. Unfortunately, most empirical tests reveal a relatively low R^2, which implies either a rejection of the CAPM, or that the *ex-post* market portfolio employed is not mean-variance efficient. We show below that the overweighing of recent observations may induce the empirically observed deviation from the CAPM. If investors overweigh recent observations when making their investment decisions, but the econometrician who tests the CAPM assigns an equal weight of $1/n$ to each observation (where the number of observations is n), a dramatic reduction in the R^2 of regression (5) is obtained. Moreover, the reduction in R^2 is obtained even if the market portfolio employed is mean-variance efficient. Let us elaborate.

Consider the case where there are n *ex-post* return observations on each asset. Let us denote the average return on asset i calculated with a weight of $1/n$ for each observation by $\bar{R}_i^{ob}$, where the superscript ob stands for objective probability. Similarly, let us denote the optimal mean-variance portfolio calculated with the $1/n$ weights by R_m^{ob}, with average return $\bar{R}_m^{ob}$, and let us denote the betas calculated against this portfolio by β_i^{ob}. If recent observations are overweighed, these parameters and the optimal mean-variance portfolio will be different. We denote the corresponding values by $\bar{R}_i^{dw}$, R_m^{dw}, $\bar{R}_m^{dw}$ and β_i^{dw}, where *dw* stands for the decision weights implied by overweighing the recent observations. We would like to emphasize that if overweighing occurs the investment proportions are affected, and R_m^{dw} would be the observed, as well as optimal market portfolio.

The researcher who tests the CAPM typically employs the $1/n$ weights, and will therefore employ $\bar{R}_i^{ob}$ in testing the CAPM. She will also use the equal weights to calculate beta, but this beta will be different from β_i^{ob}, because it is calculated against the market portfolio R_m^{dw}; we denote this beta by β^*. Thus, the empirical research community tests the CAPM with the following regression:

$$\bar{R}_i^{ob} = \gamma_0 + \gamma_1 \beta_i^* + \varepsilon_i, \tag{6}$$

and will not obtain a perfect fit. Hence, even though β_i^* is calculated against R_m^{dw}, which is mean-variance efficient, unlike the case of Roll [1977], we will obtain in testing (6) $R^2 < 1$, and the CAPM may be erroneously rejected.[17]

Thus, the experimentally observed overweighing of recent observations may have a dramatic impact on the risk-return relationship. In order to examine the magnitude of this effect we randomly select 10 stocks from the CRSP dataset and calculate the means, standard deviations, and covariances based on the last five annual returns of these stocks with the objective weights of $p_i = 0.2$. Then, we calculate the optimal mean-variance portfolio R_m^{ob} (again assuming $r_f = 3\%$), and we calculate the betas of

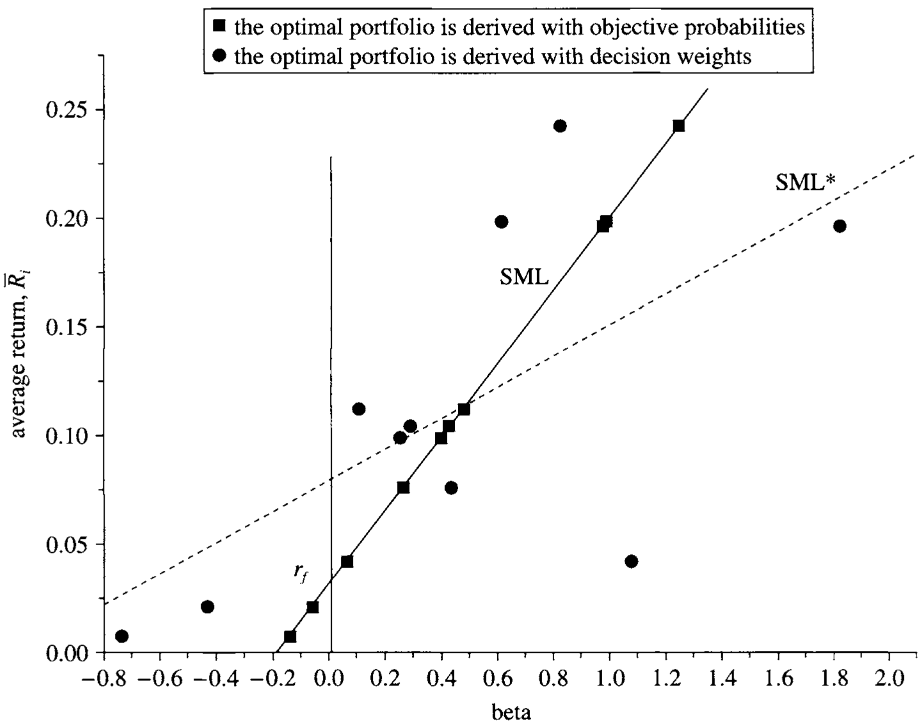


Figure 4. SML with Objective Probabilities and with Decision Weights.

the 10 stocks (relative to this optimal portfolio), β_i^{ob}. Figure 4 depicts the risk-return relationship between beta and the expected return, and, of course, as shown by Roll [1977], the analysis with the objective probabilities leads to the well-known linear Security Market Line (SML) relationship (squares in Figure 4).

Next, using the above argument, we analyze the risk return relationship with the experimentally observed overweighing of the most recent return observation, namely with $w_5 = 0.45$. Given these decision weights we find a new optimal mean-variance portfolio, R_m^{dw}, and calculate the betas relative to this portfolio. Though we use R_m^{dw} the means and betas (β*) are calculated with the objective probabilities, as an econometrician would have empirically measured them, and the effect on the risk-return relationship comes only from the change of the optimal portfolio due to the decision weights. The risk return relationship with overweighing is given by the circles in Figure 4. As can be seen in the figure, when decision weights are employed the expected return is no longer a linear function of beta, and R^2 is reduced from 1 to only 0.43. It is also interesting to note that the regression line in this case, denoted by SML* (see dotted line in Figure 4), has a smaller slope and a higher intercept than the SML. This result is typical of the empirical tests of the CAPM, and warrants further investigation.

While the above exercise is for illustration only, it shows the dramatic effects of overweighing recent observations on the risk-return relationship. This overweighing may provide a new explanation for the low R^2 obtained in many empirical studies of the CAPM risk-return relationship. The common explanation in the finance literature for the results of empirical tests of the CAPM is that the market portfolio is simply inefficient. Here we show that even with a mean-variance efficient market portfolio the different probabilities employed by the investors and the econometrician can explain the empirically observed low R^2.

5. CONCLUDING REMARKS

We experimentally test the overweighing of recent return observations in an investment experiment with 287 business school students and financial practitioners. We find that it is mainly the most recent observation which is overweighed, and that this overweighing is very strong – we estimate the decision weight attached to the most recent observation as approximately twice the objective probability.

In this framework probabilities are subjectively distorted based on the temporal sequence of the observations, rather than the distortion which takes place in single-shot lottery type decisions (as in PT, CPT, and RDEU models). This framework is applicable to circumstances where individuals are given observations as time series, as in financial markets, rather than a "given" set of outcomes and probabilities, as in many decision-making experimental setups. The case of the temporal probability distortion seems more relevant to actual economic decisions, because in practice investors observe time series data regarding corporate earnings, mutual fund returns, etc., and decisions are made based on these time series.

Clearly, the results presented here call for further investigation. It would be very informative to replicate the experiments with various different subject populations and experimental setups, perhaps with high stakes for the subjects (see footnote 3). Our analysis focuses on the overweighing of the most recent observations. Further studies may investigate other potential effects in the interpretation of time series data, such as the identification of trends (real or apparent), etc.

Our results are consistent with the "hot hand" belief of basketball fans, even though the hot hand phenomenon is statistically baseless (see Gilovich, Vallone and Tversky [1985], and Camerer [1989]). The results are also in line with the findings of Kroll, Levy and Rapoport [1988], who show that subjects make significant changes in their choices based on the two most recently observed rates of return, even though the subjects knew that the returns are drawn independently over time (see also Rabin [2002]).

One may argue that if the return distributions are non-stationary it is rational to attach a higher weight to more recent observations. Indeed, this logic may be the origin of the phenomenon we observe. However, we find that individuals attach more weight to recent observations even in circumstances where this is completely unjustified – like in our experiment where the subjects are explicitly told that the distributions are stationary and returns are i.i.d. Another example is manifested by the empirical observation of a flow of money to mutual funds with good recent performance, even though future performance has been shown to be unrelated to past performance.

There seem to be many important economic implications of this phenomenon. For example, money pointlessly "chases" the funds with the best recent performance, leading to inefficient allocations and to unnecessary transaction costs. Another phenomenon which can be straightforwardly explained by the overweighing of recent returns is the short-term momentum and longer-term reversal of stock returns. The overweighing of recent returns is also shown to lead to significant deviations from "objective" pricing. A simple analysis we perform based on the experimentally estimated overweighing leads to an average absolute deviation of 63% from the objective pricing. Thus, it seems that the heuristic of assigning more weight to more recent observations may lead in many realistic circumstances to bad economic decisions and to large economic inefficiencies.

Finally, the cornerstone risk-return relationship in the finance literature is the Sharpe-Lintner CAPM. Unfortunately, most empirical tests reveal a relatively low R^2 between $\bar{R}_i$ and β_i, with only partial support for the CAPM. Roll [1977] has shown that even with *ex-post* data $R^2 = 1$, as long as beta is calculated against a mean-variance efficient portfolio. We show that the empirically observed overweighing of recent observations implies a drop of R^2 from 100% to only 43%, even when betas are calculated against a mean-variance efficient portfolio. This result is induced by the fact that the mean-variance efficient portfolio (and hence the market portfolio) is determined by the subjective decision weights, while the average returns and betas are calculated assigning an equal probability of $1/n$ to each observation – as it is done in almost all empirical studies which test the CAPM.

ACKNOWLEDGMENT

The authors are grateful for the comments of the participants of the Second Asian Conference on Experimental Business Research and for the helpful suggestions of an anonymous referee. Haim Levy acknowledges the financial support of the Krueger Center of Finance. Moshe Levy acknowledges the financial support of the Zagagi Fund.

NOTES

1 Friedman [1953], introduces the "positive economics" concept, asserting that a model is evaluated by its predictive power and not by the validity of its assumptions on agents' rationality. Better models of individual decision making may not improve market level prediction. Whether they improve predictability is an empirical question that experimental economics tries to answer, see for example Plott [1986] and Camerer [1995]. In this study we show that investors' overweighing of recent observations can explain observed market price momentum and the flow of money to mutual funds.

[2] The experiments reported here are a part of a wider experimental project which also investigates subjects' ability to effectively employ "homemade leverage" in their investments (see Levy, Levy, and Alisof [2003]).

[3] Paying a financial reward to the subjects with no financial penalty (which is legally difficult to implement) induces a bias in the results, because the subjects tend to take extreme risks. When you receive a prize for performance and no financial penalty, you probably take a lot of risk, unlike what you do in actual investments. Indeed, when prizes are involved but no potential penalty, the subjects take extremely high leverage (see Kroll, Levy and Rapoport [1988]). When in a similar experiment, the subjects could loose out-of-pocket money, they became net lenders, i.e., shy away from taking risks, see Levy [1997].

[4] Another possible explanation is that even though in this experiment the subjects select only among the mutual funds, it is possible that their knowledge of the real-world riskless interest rate (which was 6%–8% in Israel at the time of the experiment) induced a rejection of Fund E.

[5] This has been done to make sure that the decision weights like probabilities sum up to 1. This normalization is not necessary to explain the shifts in choices from C to D to E. It is sufficient to assume that the 5[th] year is assigned a higher decision weight than the other years.

[6] If distributions are normal, the mean-variance rule is identical to SSD, namely

$$\left.\begin{aligned} E_F(x) &\geq E_G(x) \\ \sigma_F(x) &\leq \sigma_G(x) \end{aligned}\right\} \Leftrightarrow E_F U(x) \geq E_G U(x) \text{ for all concave } U \Leftrightarrow F \text{ dominates } G \text{ by SSD.}$$

Thus, if the subjects adopt the information given to them regarding the randomness and the normality of distributions of rates of return, they can employ either SSD or M-V rules. However, if the normality is violated due to the employment of decision weights $w(p)$, one can continue to employ (subjective) SSD but the M-V rule loses ground (see Tobin [1958] and Hanoch and Levy [1969]).

[7] In the above two rules we require that there is at least one strict inequality (in both sides of equations ((1) and (2)) to avoid trivial cases. (for the stochastic dominance rules see (Fishburn [1964], Hadar and Russell [1969], Hanoch and Levy, [1969], Rothschild and Stiglitz [1970] and Levy [1992] [1998]).

[8] The PSD rule is as follows: F dominates G for all S-shape utility/value functions, ($U'' \leq 0$ for $x > 0$ and $U'' \geq 0$ for $x < 0$), if and only if

$$\int_y^0 [G(t) - F(t)]\, dt \geq 0 \quad \text{for all } y \leq 0 \qquad \text{and:} \qquad \int_0^x [G(t) - F(t)]\, dt \geq 0 \quad \text{for all } x \geq 0$$

(Once again, we require a strict inequality for some pair (y_0, x_0) and for some U_0). A proof of PSD and more detail can be found in Levy [1998].

[9] The MSD rule is as follows: F dominates G for all reverse S-shaped value functions, ($U'' \geq 0$ for $x > 0$ and $U'' \leq 0$ for $x < 0$), if and only if

$$\int_{-\infty}^y [G(t) - F(t)]\, dt \geq 0 \quad \text{for all } y \leq 0 \qquad \text{and} \qquad \int_x^{-\infty} [G(t) - F(t)]\, dt \geq 0 \quad \text{for all } x \geq 0$$

(With at least one strict inequality). For proof see Levy and Levy [2002a].

[10] We analyze here various possible preference types, without getting into the question of which of these preference types better describes actual behavior. We have showed elsewhere (Levy and Levy [2002a] that Cumulative Prospect Theory's S-shape value function is experimentally rejected, and that there is support for Markowitz's reverse S-shape preference. These findings were criticized on the grounds that the study does not take into account probability distortion as advocated by CPT. However, we argue that in the case where all outcomes are equally likely (like the case employed in Levy and Levy

[2002a], with $p = 1/4$ for each of the four possible outcomes) probability distortion is not likely to play an important role. (In contrast, CPT probability distortion implies the following rather extreme subjective weights, $1/4 \rightarrow 0.29$, $1/4 \rightarrow 0.15$, $1/4 \rightarrow 0.15$, and $1/4 \rightarrow 0.29$, corresponding to the order of the outcomes, even when all outcomes are of similar magnitude). Moreover, in a different study we show that the S-shape value function is rejected even when CPT probability distortion is taken into account (see Levy and Levy [2002b]).

[11] Note that a similar analysis of Questionnaire 1 is meaningless in this case, because E is preferred over D for log utility with the objective probabilities, and overweighing the recent return in Questionnaire 1 only makes E even more attractive relative to D. This is true in the cases of linear and Prospect Theory preferences as well.

[12] For the empirical findings of auto-correlation see De Bondt and Thaler [1985], [1987], Chopra, Lakonishok and Ritter [1992], Jegadeesh and Titman [1993], Daniel [1996], Fama and French [1988], Poterba and Summers [1988].

[13] See De Bondt and Thaler [1985], Shefrin and Statman [1985], Poterba and Summers [1998], Lo and MacKinlay [1990], De Long, Shleifer, Summers and Waldmann [1990] Wang [1993], Barberis, Shleifer, and Vishny [1998], Hong and Stein [1999], and Daniel, Hirshleifer, and Subrahmanyan [1999].

[14] If returns are drawn randomly, but the investors have some beliefs which are based on the last few observations (bounded rationality), some non-random returns may be created. Based on this behavior, Levy, Levy and Solomon [2000] develop a decision-making model, where only past data is employed by investors as a predictor of future distributions of rates of return. This model induces price behavior with booms and crashes in the stock market. In addition, stock price momentum and stock price reversals are obtained. In this paper we test this behavior experimentally, in an ideal setting where the subjects are told that rates of return are i.i.d. In addition, we assume increasing vector of decision weights rather than focusing on a subset of rates of return and ignoring all other previous rates of return.

[15] Assuming an annual risk-free rate of 3%.

[16] In this optimization we assume no shortselling, and we report Δ only for those assets with $x_i^{ob} \neq 0$.

[17] Obviously, if one uses the decision weights employed by the investors she will obtain a perfect fit with $R^2 = 1$.

REFERENCES

Agarwal, V., N. D. Daniel, and N. Y. Naik (2003). "Flows, Performance, and Managerial Incentives in the Hedge Fund Industry." EFA 2003 Paper No. 501.

Albert, J., and J. Bennett (2001). *Curve ball: Baseball, Statistics, and the Role of Chance in the Game.* New York: Copernicus.

Albright, S. C. (1993). "A Statistical Analysis of Hitting Streaks in Baseball." *Journal of the American Statistical Association*, 88, 1175–1183.

Arrow, K. J. (1982). "Risk Perception in Psychology and Economics." *Economic Inquiry*, 20, 1–9.

Barberis, N., A. Shleifer, and R. Vishny (1998). "A Model of Investor Sentiment." *Journal of Financial Economics*, 49, 307–43.

Battalio, R. C., H. Kagel, and K. Jiranyakul (1990). "Testing between Alternative Models of Choice under Uncertainty: Some Initial Results." *Journal of Risk and Uncertainty*, 3, 25–50.

Beckers, S. (1997). "Manager Skill and Investment Performance: How Strong is the Link?" *The Journal of Portfolio Management*, 23, 9–23.

Budescu, D., and A. Rapoport (1994). "Subjective Randomization in One- and Two-Person Games." *Journal of Behavioral Decision Making*, 7, 261–278.

Camerer, C. F. (1989). "Does the Basketball Market Believe in the 'Hot Hand?' *American Economic Review*, 79, 1257–61.

Camerer, C. F. (1992). "The Rationality of Prices and Volume in Experimental Markets." *Organizational Behavior and Human Decision Processes*, 51, 237–72.

Camerer, C. F. (1995). "Individual Decision Making," in *Handbook of Experimental Economics*. J. Kagel and A. E. Roth, eds., Princeton: Princeton U. Press.

Camerer, C. F., and M. Weber (1992). "Recent Developments in Modeling Preferences: Uncertainty and Ambiguity." *Journal of Risk and Uncertainty*, 5, 325–370.

Chevalier, J., and Ellison, G. (1997). "Risk Taking by Mutual Funds as Response to Incentives." *Journal of Political Economy*, 105, 1167–1200.

Chopra, N., J. Lakonishok, and J. R. Ritter (1992). "Measuring Abnormal Performance: Do Stocks Overreact?" *Journal of Financial Economics*, 31, 235–68.

Daniel, K. D. (1996). "The Power and Size of Asset Pricing Tests." University of Michigan Mimeo, 1996.

Daniel, K., D. Hirshleifer, and A. Subrahmanyan (1999). "Investor Psychology and Market- and Over-Reactions." *Journal of Finance*, 53, 1839–85.

De Bondt, W., and R. Thaler (1985). "Does the Stock Market Overreact?" *Journal of Finance*, 40, 793–805.

De Bondt, W., and R. Thaler (1987). "Further Evidence on Investor Overreaction and Stock Market Seasonality." *Journal of Finance*, 42, 557–81.

DeLong, B. J., A. Shleifer, L.H. Summers, and R.J. Waldman (1990). "Positive Feedback Investment Strategies and Destabilizing Rational Speculation." *Journal of Finance*, 45, 370–395.

Edwards, W. (1953). "Probability Preference in Gambling." *American Journal of Psychology*, 66, 349–364.

Edwards, W. (1962). "Subjective Probabilities Inferred from Decisions." *Psychology Review*, 69, 109–35.

Fama, E. F. (1970). "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, 25, 383–417.

Fama, E. F. (1991). "Efficient Capital Markets: II." *Journal of Finance*, 46, 1575–1617.

Fama, E. F., and K. French (1988). "Permanent and Temporary Components of Stock Prices." *Journal of Political Economy*, 96, 246–273.

Fishburn, P. C. (1964). *Decisions and Value Theory*, Wiley, New York.

Friedman, M. (1953). *Methodology of Positive Economics*, Chicago: University of Chicago Press.

Gilovich, T., B. Vallone, and A. Tversky (1985). "The Hot Hand in Basketball: On the Misconception of Random Sequences." *Cognitive Psychology*, 17, 295–314.

Hadar, J., and W. Russell (1969). "Rules for Ordering Uncertain Prospects." *American Economic Review*, 59, 25–34.

Hanoch, G., and H. Levy (1969). "The Efficiency Analysis of Choices Involving Risk." *Review of Economic Studies*, 36, 335–346.

Harless, D. W. and C. F. Camerer (1994). "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica*, 62, 1251–89.

Hong, H., and J. C. Stein (1999). "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets." *Journal of Finance*, 2143–2184.

Jegadeesh, N. (1990). "Evidence of Predictable Behavior of Security Returns." *Journal of Finance* 45, 881–898.

Jegadeesh, N., and S. Titman (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *Journal of Finance*, 48, 65–91.

Kahneman, D., and A. Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica*, 47, 263–291.

Kroll, Y., H. Levy, and A. Rapoport (1988). "Experimental Test of the Mean-Variance Model for Portfolio Selection." *Organizational Behavior and Human Decision Processes*, 42.

Levy, H. (1992). "Stochastic Dominance and Expected Utility: Survey and Analysis." *Management Science*, April, 555–593.

Levy, H. (1997). "Risk and Return: An Experimental Analysis." *International Economic Review*, February, 38, 119–149.

Levy, H. (1998). *Stochastic Dominance*, Kluwer academic Press.

Levy, H., and Levy, M. (2002b). "Experimental Test of the Prospect Theory Value Function." *Organizational Behavior and Human Decision Processes* 89, 1058–1081.

Levy, H., and M. Levy (2003). "Prospect Theory and Mean-Variance Analysis." *Review of Financial Studies*, forthcoming.

Levy, H., M. Levy, and N. Alisof (2003). "'Homemade Leverage': Theory versus Experimental Evidence." *Journal of Portfolio Management*, forthcoming.

Levy, H., and K. C. Lim (1998). "The Economic Significance of the Cross-Sectional Autoregressive Model: Further Analysis." *Review of Quantitative Finance and Accounting*, 11, 37–51.

Levy, M., and H. Levy (2001). "Testing for Risk-Aversion: A Stochastic Dominance Approach." *Economics Letters*, 71, 233–240.

Levy, M., and H. Levy (2002a). "Prospect Theory: Much Ado About Nothing?" *Management Science*, 48, 1334–1349.

Levy, M., H. Levy, and S. Solomon (2000). *Microscopic Simulation of Financial Markets: From Investor Behavior to Market Phenomena*, Academic Press.

Lintner, J. (1965). "Security Prices, Risk, and Maximal Gains from Diversification." *Journal of Finance* 20, 587–615.

Lo, A., and C. MacKinlay (1990). "When are Contrarian Profits due to Stock Market Overreaction?" *Review of Financial Studies*, 3, 175–205.

Luo, D. (2003). "Chasing the Market? Driving the Market? A Study of Mutual Fund Cash Flows", Yale ICF Working Paper No. 03-22.

Markowitz, H. (1952a). "Portfolio Selection." *Journal of Finance* 7, 77–91.

Markowitz, H. (1952b). "The Utility of Wealth." *Journal of Political Economy*, 60, 151–156.

Payne, J. W., J. Bettman, and E. Johnson (1992). "Behavioral Decision Research: A Constructive Processing Perspective." *Annual Review of Psychology*, 43, 87–131.

Plott, C. R. (1986). "Rational Choice in Experimental Markets." *Journal of Business*, 59, 301–327.

Poterba, J., and L. Summers (1988). "Mean Reversion in Stock Prices: Evidence and Implications." *Journal of Financial Economics*, 22, 27–59.

Prelec, D. (1998). "The Probability Weighting Function." *Econometrica*, 66, 497–527.

Preston, M. G., and P. Baratta (1948). "An Experimental Study of the Auction-Value of Uncertain Outcomes." *American Journal of Psychology*, 61, 183–193.

Quiggin, J. (1982). "A Theory of Anticipated Utility." *Journal of Economic Behavior Organization*, 323–43.

Rabin, M. (2002). "Inferences by Believers in the Law of Small Numbers." *Quarterly Journal of Economics*, 117, 817–869.

Rapoport, A. and Budescu, D. (1992). "Generation of Random Binary Series in Strictly Competitive Games." *Journal of Experimental Psychology: General*, 121, 352–364.

Rapoport, A. and Budescu, D. (1997). "Randomization in individual choice behavior." *Psychological Review*, 104, 603–617.

Roll, R., (1977). "A Critique of the Asset Pricing Theory's Test, Part I: On Past and Potential Testability of Theory." *Journal of Financial Economics* 4, 129–176.

Rothschild, M., and J. Stiglitz (1970). "Increasing Risk: I. A definition." *Journal of Economic Theory*, 2, 225–243.

Samuelson, P. A. (1989). "The Judgment of Economic Science on Rational Portfolio Management: Indexing, Timing and Long-Horizon Effects." *The Journal of Portfolio Management*, 1989, 4–12.

Sapp, T., and A. Tiwari (2003). "Does Stock Return Momentum Explain the 'Smart Money' Effect?" *Journal of Finance*, forthcoming.

Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *Journal of Finance* 19, 425–442.

Sharpe, W. F. (1966). "Mutual Fund Performance." *Journal of Business*, September, 39, 119–138.

Shefrin, H. M., and M. Statman (1985). "The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence." *Journal of Finance*, 40, 774–90.

Tobin, J. (1958). "Liquidity Preference as Behavior Toward Risk." *Review of Economic Studies*, 25, 457–59.

Tversky, A., and D. Kahneman (1971). "Belief in the Law of Small Numbers." *Psychological Bulletin*, 76, 105–110.

Tversky, A., and D. Kahneman (1974). "Judgment under Uncertainty: Heuristics and Biases." *Science*, 185, 1124–1130.

Tversky, A., and D. Kahneman (1992). "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty*, 297–323.

von Neuman J. and O. Morgenstern (1953). *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.

Wang, J. (1993). "A Model of Intertemporal Asset Prices Under Asymmetric Information." *Review of Economic Studies*, 60, 249–282.

This page intentionally blank

Chapter 8

COGNITION IN SPATIAL DISPERSION GAMES

Andreas Blume
University of Pittsburgh

Douglas V. DeJong
University of Iowa

Michael Maier
University of Iowa

Abstract

In *common-interest spatial-dispersion games* the agents' common goal is to choose distinct locations. We experimentally investigate the role of cognition in such games and compare it with the role of cognition in spatial matching games. In our setup cognition matters because agents may be differentially aware of the dispersion opportunities that are created by the history of the game. We ask whether cognitive constraints limit the agents' ability to achieve dispersion and, if there is dispersion, whether these constraints affect the mode by which agents achieve dispersion. Our main finding is that strategic interaction magnifies the role of cognitive constraints. Specifically, with cognitive constraints, pairs of agents fail to solve a dispersion problem that poses little or no problem for individual agents playing against themselves. When we remove the cognitive constraints in our design, pairs of agents solve the same problem just as well as individuals do. In addition, we find that when playing against themselves agents do not change the mode by which they solve the dispersion problem when our design removes the cognitive constraints.

1. INTRODUCTION

In *spatial dispersion games* the agents' common goal is to choose distinct locations. Such games have been used to study congestion problems, habitat selection, and networking issues, e.g., Alpern and Reyniers [2002] and Alpern and Gal [2003]. More generally, dispersion incentives in location games appear in models of product differentiation, e.g., Salop [1979], and variants of the voting models of Hotelling [1929] and Downs [1957], e.g., Palfrey [1984].[1]

We experimentally investigate the role of cognition in such games and compare it with the role of cognition in spatial matching games, where the common goal of

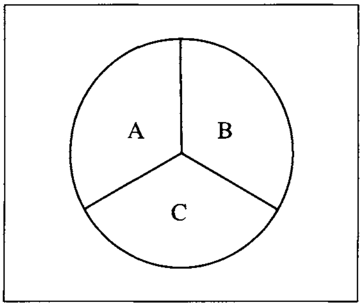
R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 185–202.
© 2005 Springer. Printed in the Netherlands.

the agents is to choose the same location. In our setup cognition matters because agents may be differentially aware of the dispersion opportunities that are created by the history of the game. Once agents achieve dispersion in a repeated spatial dispersion game and if they can remember past choices, they have the option to maintain dispersion by simply maintaining their previous choices. When agents do not have a simple record of their own past choices there may be other ways of sustaining dispersion. Cognitive issues arise when agents do not have a simple record of their own past choices, but there is a procedure for inferring own past choices. Some agents may be aware of this procedure while other agents may be unaware of it.

Unawareness of this sort requires more than simple lack of knowledge. In addition to not knowing the procedure the agent must not know that he does not know the procedure, i.e., he must lack *negative introspection*. Unawareness seems commonplace in everyday life, and yet has only recently attracted attention in the literature. One likely reason is that unawareness does not easily fit into conventional models of information economics. Violations of negative introspection are not compatible with the standard partitional state space model of knowledge, Aumann [1976], as pointed out by Geanakoplos [1992]. More recently, Dekel, Lipman and Rustichini [1998] have demonstrated that *any* standard state space model precludes unawareness. They suggest that one way to avoid this conundrum is to make a distinction between the agent's and the analyst's description of the state space, and to treat the state space as "representing the agent's view of possibilities." Recently, there have been a few proposals of models of knowledge that permit unawareness, e.g., Li [2003] and Schipper [2002]. Furthermore, there have been suggestions that properly incorporating unawareness into our models may shed light on issues related to contractual incompleteness and no-trade theorems.

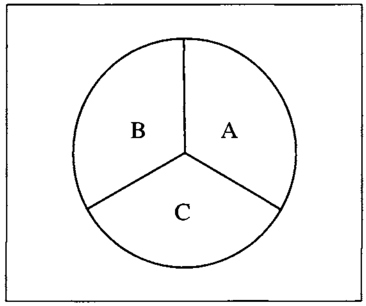
Our objective is more modest. We accept unawareness as a simple empirical phenomenon and ask what happens when agents differ in their awareness in a simple strategic setting, i.e., when there is interactive unawareness. Common-interest games are attractive for this purpose because they help us focus on the central issue of how unawareness affects players' strategic reasoning about others. We need not worry for example about how differential awareness interacts with signaling motives, bargaining motives, deception, threats, punishments, or other-regarding preferences. Location games with a spatial structure are appealing because agents may differ in how much of this structure and its possible uses they perceive.

For a formal model of interactive unawareness in our games we follow Bacharach [1993]. He calls for a model of games in which "one specifies the way players conceive the situation and how this varies." He provides details of such a model of *variable universe games* for the case where the players' aim is to choose a common action, i.e., for matching games. In Bacharach's model, a player's perception is essentially given by a partition of the set of actions. Blume and Gneezy [2002] extend Bacharach's approach to permit a more general structure on the sets of actions than partitions, or collections of partitions. It permits the spatial (circular) structure that is used in Blume and Gneezy [2002], Blume, DeJong and Maier



Player 1

Figure 1a.



Player 2

Figure 1b.

[2003] and that will be used in the present paper to address spatial dispersion games.[2]

A basic version of the dispersion game (that we expand upon and fully develop later in the paper) consists of two players who are randomly paired together for a one-shot game. The two players simultaneously and independently choose one of three identical unlabeled sectors of a disc, as illustrated in Figure 1. One player sees a disc whose labels have the directional order indicated in Figure 1a. The other player sees a disc with the directional order of the labels reversed, as in Figure 1b. The locations are randomized at the beginning of the one-shot game and neither player sees the labels A, B, and C themselves. In a spatial dispersion game, the payoffs are one if both players choose different sectors, A and B, B and C, or C and A, and zero if they choose the same sector, A, B, or C. For a simple spatial matching game the payoffs are just the reverse.

Blume and Gneezy [2002] have experimentally demonstrated that there are differences in awareness in spatial matching games. Blume and Gneezy consider one-shot spatial matching games in which players simultaneously choose a single sector from a disc with five sectors. All sectors are identical in size and shape, three are white, and two are black. They compare two scenarios, one in which a single individual plays against him- or herself, and one in which two distinct players play against each other. In either case, given the symmetry constraints imposed by the task, there is a unique optimal way to play the game. Success is only guaranteed if both choices correspond to the midpoint of the odd distance between the two black sectors. Cognitive differences can be shown to exist by having players play against themselves. When playing against themselves, players who are aware of the guaranteed success strategy will use it, while others will be attracted to the obvious alternative, to choose one of the black sectors. Blume and Gneezy find that a significant percentage of participants do not solve the game when playing against themselves.

In the matching games of Blume and Gneezy [2002], cognitive differences prevent players from coordinating on the unique optimal solution. Cognitive differences are likely to play a different role in dispersion games. Even though in both kinds of games agents have a common objective, the structure of equilibria is different. Unlike in matching games, in dispersion games typically none of the equilibria are strict: As long as there are more locations than agents, an agent can always switch to an unused location and still maintain dispersion. Also, while the matching games of Blume and Gneezy [2002] have a unique optimal solution, there are multiple ways in which dispersion can be achieved in our games. This makes the questions of whether any equilibrium is attained and, if so, which one will be selected important.

The present paper has agents interact repeatedly in spatial dispersion games. Repeated interaction in spatial matching games with a circular structure has been investigated by Blume, DeJong and Maier [2003]. There, players are randomly paired each period. The stage game played in each period consists of two rounds. In the first round of the stage game two players simultaneously and independently choose one of n identical sectors of a disc, where n is odd.[3] In the second round, after observing first round choices, but without being able to distinguish one's own from one's partner's choice, both players choose again. In both rounds, payoffs are one if both players choose identical sectors and zero otherwise. Note that the second round induces essentially the same choice problem as the task in Blume and Gneezy [2002] and therefore has a unique optimal solution.

In the repeated spatial matching games of Blume, DeJong and Maier, learning can occur at two levels. At one level, in each period, agents can learn by labeling actions in the first round and using these labels in the second round.[4] At the other level, agents can learn across periods about how to learn within a period. This type of learning, which we call *cognitive learning*, has to the best of our knowledge of the literature only been addressed in the Blume, DeJong and Maier [2003] paper. Initially, there may be agents who are unaware of the fact that the labels introduced by first-round choices can always be used to identify a unique distinct sector. Other agents may be aware of this possibility. In the course of the multi-period interaction, agents may *become aware* of this possibility, i.e., engage in *aha learning*, Bühler [1907, 1908], Köhler [1925] and Weber [2003].[5] The results from our matching games support coordination outcomes and we find evidence for cognitive learning. That is, in simple environments agents learn across periods to make better use within a period of labels created in that period. We observe transfer of cognitive learning from simple environments to more complicated environments.

As previously noted, the structures of the action space that agents may or may not be aware of have different uses in dispersion games than matching games. For example, the circular structure of the matching game of Blume, DeJong and Maier [2003] enables agents to identify a unique candidate for a common action. The same circular structure in a dispersion game generates a "coordination problem" characterized by multiple, non-strict equilibria. This difference in the possible use of structures suggests that the learning may also be different.

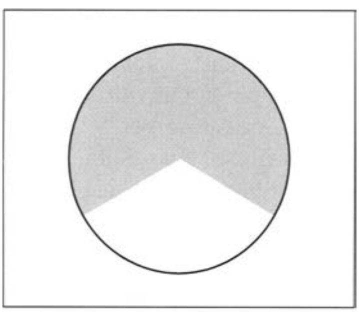
Our main finding in the present paper is that in spatial dispersion games, strategic interaction magnifies the role of cognitive constraints. Specifically, with cognitive constraints, pairs of agents fail to solve a dispersion problem that poses little or no problem for individual agents playing against themselves. When we remove the cognitive constraints in our design, pairs of agents solve the same problem just as well as individuals do. In addition, we find that when playing against themselves agents do not change the mode by which they solve the dispersion problem when our design removes the cognitive constraints.

2. GAME AND EXPERIMENTAL DESIGN

We study a repeated dispersion game in which two players are randomly paired together and stay paired for twenty-one periods. In the first period, the two players simultaneously and independently choose one of three identical unlabeled sectors of a disc, as illustrated in Figure 1.[6] One player sees a disc whose labels have the directional order indicated in Figure 1a. The other player sees a disc with the directional order of the labels reversed, as in Figure 1b. Neither player sees the labels A, B, and C themselves. The payoffs are one if both players choose different sectors, A and B, B and C, or C and A, and zero if they choose the same sector, A, B, or C. At the end of period one, the two players are informed about the sectors that were chosen.

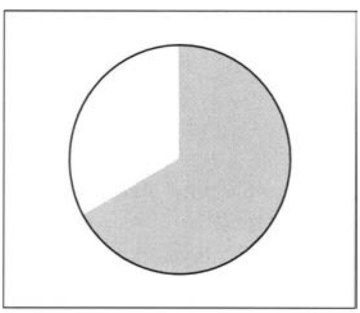
At the beginning of the second period, players observe the previous period's choices but without being able to distinguish one's own from one's partner's choice, for example see Figure 2 where the players achieved a dispersion outcome and where the discs with the first period choices have been randomly spun and presented to the players, Figure 2a and 2b respectively, at the beginning of period two. Both players then choose again. The payoffs are again one if both players choose different sectors and zero if they choose the same sector. At the end of period two, the two players are informed about the sectors that were chosen. Specifically they see the choices made in period 2, marked by red dots, on the background of the choices made in the previous period, marked by shaded sectors. Each of the subsequent periods follows the same sequence outlined for the second period.

We implement a two-by-two design. The first dimension is the information provided to players about their choices. The *relative-location information condition* is described above. In the theory for dispersion games, it is common practice to assume



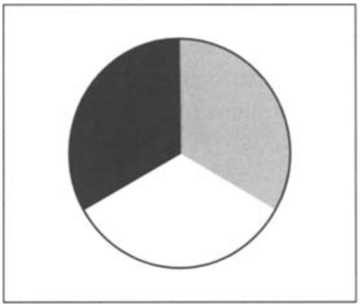
Player 1

Figure 2a.



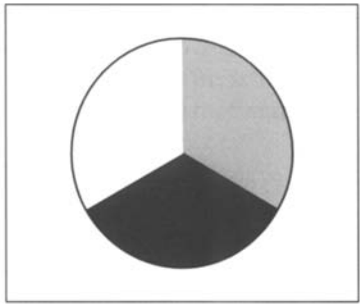
Player 2

Figure 2b.



Player 1

Figure 3a.



Player 2

Figure 3b.

that agents know their present location and the location of other agents when taking their future choice of action. This describes the *precise-location information condition* and is illustrated in Figure 3a and 3b, where the choice of one player is noted in dark shading and the other player's choice is lightly shaded.

The second dimension of the design is the pairing of the players. The first condition is *fixed pairing*, as describe above. The second condition is *self pairing* where a player is paired with him or herself for the duration of the repeated spatial dispersion game. The purpose of this dimension is to separate the cognition problem from the coordination problem. Thus, there are four treatments in our design; fixed-pairing with relative and precise location information, and self-pairing with relative and precise information.

The experiment was conducted using a series of six cohorts; two cohorts or replications each for the two information treatments with fixed-pairing and one replication each for the two information treatments with self-pairing. A cohort consisted of twelve participants. Such a design provides the same number of pair observations in each of the four treatments. All participants were recruited from undergraduate (sophomore and above) and graduate classes at the University of Iowa. None of the participants had previously taken part in or otherwise gained experience with this series of treatments. Upon arrival, participants were seated at separate computer terminals and given a copy of the instructions.[7] Before each replication, instructions were read aloud and participants individually filled out questionnaires confirming their knowledge and understanding of the instructions. We then went over the questionnaire orally and answered questions. Since these instructions were read aloud, we assume that the information contained in them was mutual knowledge.

Each cohort played a repeated spatial dispersion game for twenty-one periods from one of the four treatments in the design. Each period had the following structure. Prior to the beginning of the first period, participants were paired using a random-matching procedure or paired with themselves. In the first period, participants chose a sector from a symmetric disc with 3 identical sectors. At the beginning of the first period, the discs were randomly rotated, independently across participants or across the two computer screens used by a participant in the self-pairing treatments, to eliminate all possibilities for *a priori* coordination. Then, participants made their choices by using a mouse to click on their chosen sector. They were given an opportunity to either revise or confirm their choices. At the end of the period, when all participants had made and confirmed their choices, they were informed about which sectors were chosen in their match.

At the beginning of period two, each disc was randomly rotated and period-one choices were displayed in the new configurations. In the display for the relative information treatments no distinction was made between one's own choice and one's partner's choice, see Figure 2a and 2b. This procedure ensured that in the second period, participants only had information about the configuration of choices. In the precise information treatments, each player's choice was indicated for both players and for the self-pairing treatments the choices made on each computer screen were indicated for the player, see Figure 3a and 3b. In the second period, participants once more chose one of the three sectors from the same disc as before with the prior choices displayed as just described. At the end of the period, when all participants had made and confirmed their choices, they were informed about which sectors were chosen in their pair along with the relative (precise) locations of the previous period's choices. Each subsequent period through period twenty-one followed the same sequence detailed for period two.

Each replication lasted from one-half to one hour. Participants' earnings ranged from \$7.50 to \$15.75 plus a "show up" payment of \$5.

3. THEORY

A solution for our relative information fixed-pairing treatment, must acknowledge two fundamental characteristics of the game. These are the symmetries that are built into the game, and potential differences in players' abilities to recognize when these symmetries have been broken.

Our design ensures that in the first period of our game all three sectors are completely symmetric. Players could not guarantee dispersion even if we permitted them to talk before the game. The fact that we rotate the disc independently across players guarantees that players *de facto* randomize by assigning equal probabilities to all sectors in the first period.

In the second period, players observe which sectors were chosen in the first period. Consider the case where players achieved dispersion in the first period (the other case, in which their choices resulted in congestion, is analyzed analogously). The fact that we spin the disc and that both players' choices are marked identically ensures that players cannot distinguish between their own choice and their partner's choice. Therefore, players are *de facto* precluded from guaranteeing dispersion in the second period by maintaining their first-period choices in the second period.

However, unlike in the first period, the absence of communication is a binding constraint here. If they could communicate, they could agree on one player playing the *odd sector*, the sector not chosen by either player in the first period, and the other player playing one of the first-period choices. In the absence of communication, the fact that players' positions are identical prevents them from coordinating on such asymmetric behavior. Therefore we look for equilibria where in the second period both players put the same probability on the odd sector.

Before the third period (and similarly for subsequent periods) smart players will remember whether in the second period they chose the odd sector, the sector to the left of the odd sector (as viewed from the center of the disc), or the sector to the right of the odd sector. Then, if they manage to achieve dispersion in the third period, they can achieve dispersion in every subsequent period by following the rule of choosing the same sector in relation to the odd sector as in the previous period.

A problem arises because not all players need be smart, in the sense of realizing the possibility of making left-right distinctions on the disc. Players who can only distinguish chosen and unchosen sectors can only guarantee future dispersion if the dispersion realized was such that one player in the previous period chose the odd, unchosen, sector and the other chose one of the two previously chosen sectors. We formalize this problem by allowing for different types of players, who are endowed with different languages, a coarse-language and a fine-language, in which they describe the choice set to themselves.

The distinction between coarse- and fine-language players is as follows. Coarse-language players can only distinguish *chosen* and *unchosen* sectors in any period after the first period. Fine-language players can use the circular structure to enumerate all sectors after the first period. Further, fine-language players can commonly distinguish all sectors in a period after the second period. The reason is that for period three and after, fine-language players can describe each others' choices relative to the odd sector. Already, in period two, a fine-language player can for example choose "the sector to the left of the odd sector." At the beginning of period three, a fine-language player can also see his partner's period-two choice in reference to the odd sector of period one. As a result, fine-language players can maintain dispersion in period three and all subsequent periods.[8]

Player symmetry requires that players use identical strategies. Accordingly, we will focus on equilibria in which players use identical strategies and in which they employ efficient symmetric continuation strategies.

Denote by V_D a player's continuation payoff after players have achieved sustainable dispersion (dispersion in period three or later for fine-language players, and chosen-unchosen dispersion for coarse-language players) and by V_O the continuation payoff otherwise. Denote by p and q the probabilities of each player choosing the odd sector before there is sustainable dispersion, either the sector not chosen if players chose different sectors or the sector chosen if players chose the same sector. Note that the probabilities assigned to the two remaining sectors have to equal $(1 - p)/2$ each for one player and $(1 - q)/2$ each for the other. Of course in a symmetric equilibrium p and q must be the same. Consider the two cases were all players are fine-language players, $\lambda = 1$, or all players are coarse-language players, $\lambda = 0$. Then the payoff from using probability q against probability p equals:

$$\pi(q, p) = pq[0 + V_O] + (p(1-q) + q(1-p))[1 + V_D]$$
$$+ (1-p)(1-q)\left[\frac{1}{2}[1 + \lambda V_D + (1-\lambda)V_O] + \frac{1}{2}[0 + V_O]\right].$$

In equilibrium, the player choosing q must be indifferent among all q. Hence the derivative with respect to q must be zero.

$$\frac{\partial \pi(q, p)}{\partial q} = pV_O + (1-2p)[1 + V_D] - (1-p)\left[\frac{1}{2}[1 + \lambda V_D + (1-\lambda)V_O] + \frac{1}{2}V_O\right]$$
$$= 0.$$

Solving for p, we obtain

$$p = \frac{2V_D + 1 - 2V_O - \lambda[V_D - V_O]}{4V_D + 3 - 4V_O - \lambda[V_D - V_O]}.$$

Hence, if all players are fine-language players, $\lambda = 1$, then

$$p_f = \frac{1}{3}.$$

Fine-language players uniformly randomize across all three sectors through period two and continue to randomize in period three and subsequent periods until dispersion is achieved. Once dispersion is achieved, players coordinate by both choosing left or right of the odd sector or by selecting chosen and unchosen.

If all players are coarse-language players, $\lambda = 0$, then

$$p_c = \frac{2[V_D - V_O] + 1}{4[V_D - V_O] + 3}.$$

Note that p_c is increasing in $V_D - V_O$. We conclude that coarse-language players put more probability on the odd sector than fine-language players. After period one, coarse-language players randomize until they achieve the dispersion outcome of chosen and unchosen sectors. Observe that cognitive differences only matter in the repeated game with at least three periods.

More generally, we can consider the incomplete information game where a player is a fine-language player with probability μ and a coarse-language player with probability $1 - \mu$.[9] Coarse-language players being unaware of their cognitive constraint attach no probability to other players being fine-language players. They play under the presumption that the other player is a coarse-language player with certainty. Therefore, in the incomplete information game, regardless of μ, coarse-language players use the strategy derived for the complete information game above in which all players are coarse-language players.

In contrast, fine-language players are aware of the fact that both types are present and accordingly have beliefs about the type of the player they are facing. Thus, in general, optimal behavior of fine-language players could depend on their beliefs and potentially require complicated updating of beliefs. Fortunately, in the present context, the previously noted strategy for fine-language players, derived above under the assumption that it is common knowledge that all players are fine-language players, remains optimal for *any* belief β by fine-language players that their partner is a fine-language player. To see this, simply note that this strategy is optimal against both fine-language players and coarse-language players. The optimality against fine-language players is immediate.

The optimality against coarse-language players follows from the following facts: (1) against a coarse-language player one can not do better than a coarse-language player; (2) in periods in which a coarse-language player randomizes, *any* form of randomization, including playing the odd sector with probability p_f or repeating an action that led to dispersion the last period is optimal; and (3), trying to maintain dispersion by repeating last period's action is optimal in periods where a coarse-language partner is doing the same.

In the precise information fixed-pairing treatment, all players are fine-language players unless they ignore the information given to them. They can all distinguish among the sector they chose, the sector chosen by the player they are paired with, and the odd sector. All players uniformly randomize until a dispersion outcome is achieved. Once achieved, the dispersion outcome is played for the remainder of the game, both play left or right of the odd sector. As long as there are coarse-language players here, the probability of picking the odd sector is greater than or equal to one-third and the dispersion outcome can also be achieved by the chosen and unchosen selection.

In the self-pairing treatments, relative and precise information, all players uniformly randomize in period one. In period two, all players should achieve a dispersion outcome because there is no coordination problem after the first period. Fine-language players have the option of choosing to the left or right of the odd sector; coarse-language players can only coordinate by focusing on chosen and unchosen sectors.

4. RESULTS

4.1. Dispersion Outcomes

We first present the proportion of dispersion outcomes achieved by period for the four treatments, fixed-pairing with precise and relative information and self-pairing with precise and relative information, Figure 4. First, note that the self-pairing precise information treatment reaches full coordination first. Second, the proportion of dispersion outcomes for the fixed-pairing precise information and self-pairing precise and relative information treatments are indistinguishable. In these three treatments, all players are either fine-language players (fixed-pairing) or should not have a coordination problem when selecting a dispersion outcome (self-pairing). Third,

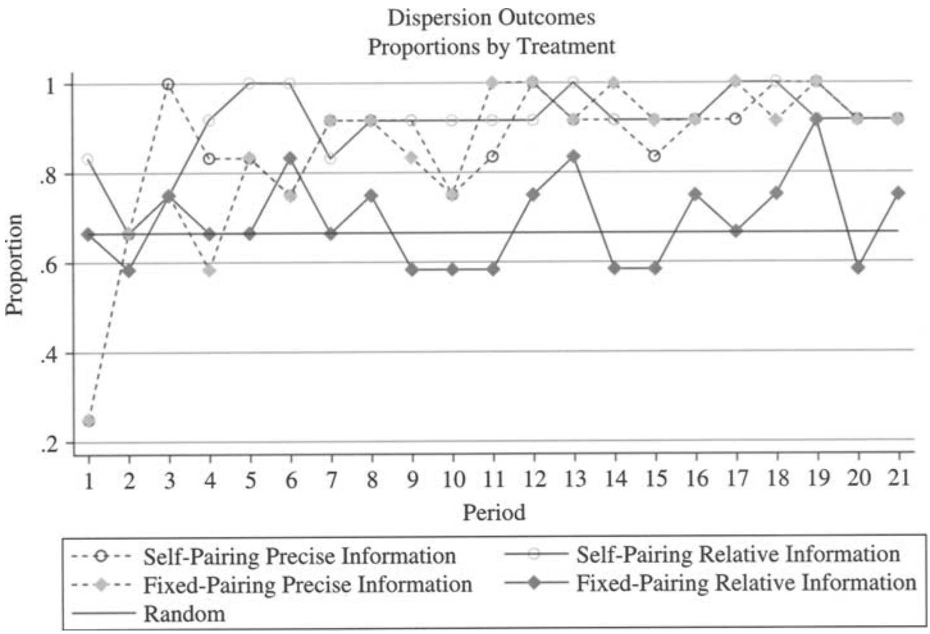


Figure 4.

while the self-pairing precise and relative information treatments do not reach coordination in period two, as predicted, the treatments are well on their way by period three. Finally, the proportion of dispersion outcomes in the fixed-pairing relative information treatment is indistinguishable from the expectation that behavior is random, .67. This result contrasts sharply with the result in Blume and Gneezy [2002], where relative information increased coordination relative to precise information.

4.2. *Individual Player Choices*

Regarding individual player choices, our theory suggests that for fixed-pairing, prior to achieving a dispersion outcome, the probability of selecting the Odd sector is higher in the relative information treatment ($p > 1/3$) than in the precise information treatment ($p = 1/3$). Unfortunately, there are very few observations here, sixteen in period two to be exact, too few for any meaningful analysis across the two treatments.[10] However, aggregating across the two treatments, $p > 1/3$, which is the prediction from theory in the presence of coarse-language players in both treatments.

4.3. *Paired Player Choices*

Paired choices of players are presented in Table 1 for the four treatments and as a basis for comparison, the expectation that behavior is random. The relationship

Table 1. Paired Player Choices

The relationship between paired choices in period t and outcomes in period $t - 1$ is presented by treatment for periods two to twenty-one. Paired choices in period t are broken down by whether the paired choices are Odd/Not Odd or Both Not Odd with the Dispersed outcome in period t, or whether the paired choices are Other combinations that all imply the Matched outcome in period t. Outcomes are broken down by Dispersed and Matched in period $t - 1$. For comparison purposes, outcomes are also presented under the expectation that behavior is random.

Treatment	*Choices Period t:*	*Paired Choices*			
		Dispersed		*Matched*	*Total*
		Odd/ Not Odd	*Both Not Odd*	*Other*	
	Outcome Period t – 1:				
Fixed-Pair					
Precise					
	Dispersed	39	150	13	202
	Matched	15	6	17	38
		54	156	30	240
Relative					
	Dispersed	69	44	52	165
	Matched	44	9	22	75
		113	53	74	240
Random		107	53	80	240
Self-Pair					
	Dispersed	323	71	36	430
	Matched	27	11	12	50
		350	82	48	480
Random		214	106	160	480

between paired choices in period t and outcomes in period $t - 1$ is presented for periods two to twenty-one. Table 1 also presents the outcomes, Dispersed and Matched, for the paired choices for period t. Paired choices in period t are broken down by whether the choices are Odd/Not Odd or Both Not Odd with the Dispersed outcome in period t, or whether the paired choices are Other combinations that all imply the Matched outcome in period t. Outcomes in period $t - 1$ are broken down by Dispersed and Matched.

In the fixed-pairing precise information treatment, all players should be fine-language players and therefore should have access to playing left or right of the odd sector. How successful were the players in achieving a dispersed outcome and in coordinating their Not Odd choices, both choose Left or both Right, to achieve a dispersed outcome? From Table 1, out of a possible 240 outcomes, 210 are dispersed. For the 210 dispersed outcomes, 156 choices in period t were Both Not Odd (which implies both chose Left or both Right) and 54 were Odd/Not Odd (which from our theory implies chosen and unchosen).

Figure 4 suggests a difficult coordination problem in the fixed pair relative information treatment. Table 1 documents this problem. For the 165 dispersed outcomes achieved in period $t - 1$, players failed to capitalize on this success 52 times in period t. Further, for the successes achieved in period t, sometimes players coordinated on Odd/Not Odd, 69, and sometimes Both Not Odd, 44. A similar conclusion holds for the analysis of the 75 matched outcomes in period $t - 1$. Given either a dispersed or matched outcome in period $t - 1$, players face the coordination task in period t of choosing over Odd/Not Odd or Not Odd (with Not Odd presenting a secondary coordination problem of how to coordinate over the two sectors). Player choices are consistent with the expectation that behavior is random.

In the self-pairing treatments, players do not face such a coordination problem. A player can decide him or herself between Odd/Not Odd and Not Odd (both Right or both Left), regardless of the prior period's outcome. Players were very successful at achieving a dispersion outcome, but it is difficult to distinguish between coarse and fine-language players. The results implied by Figure 4 and shown in Table 1 (the two information treatments are combined in Table 1 because of their similar play) document that the number of matched outcomes is lowest in the self-pairing treatments despite the large number of Odd choices by players. Some players coordinated in period t by choosing Right or Left of the Odd sector on both screens, 82 out of 480. However, most players coordinated by Odd/Not Odd, 350 out of 480. This choice, Odd/Not Odd (or from theory, chosen and unchosen) appears to be the "least costly" way to coordinate rather than a statement about coarse and fine-language players.

4.4. *Frequency of Paired Choices by Period*

We next consider how many times player pairs chose a particular set of choices in each period. Figure 5 presents the results for the fixed-pairing precise information treatment. The graph documents the frequency of the paired choices made, Both Not Odd, Odd/Not Odd and Other. To read this graph, note that for Both Not Odd, eight

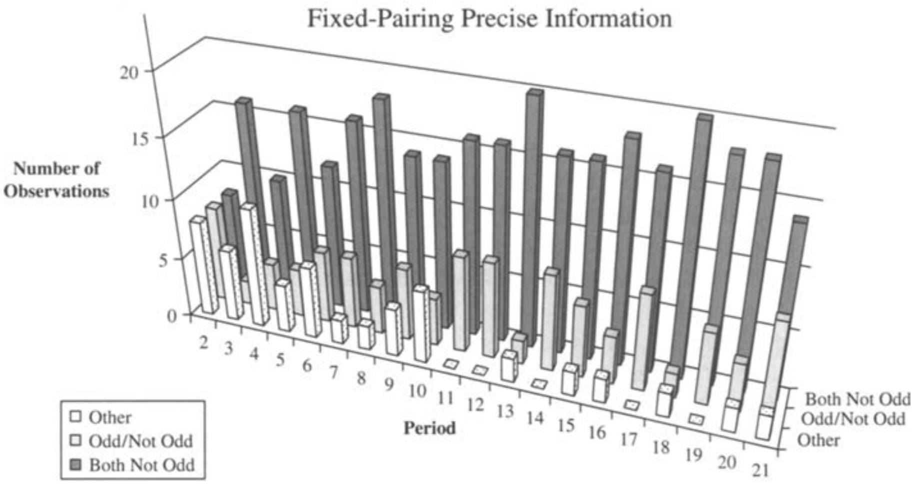


Figure 5. Frequency of Paired Choices.

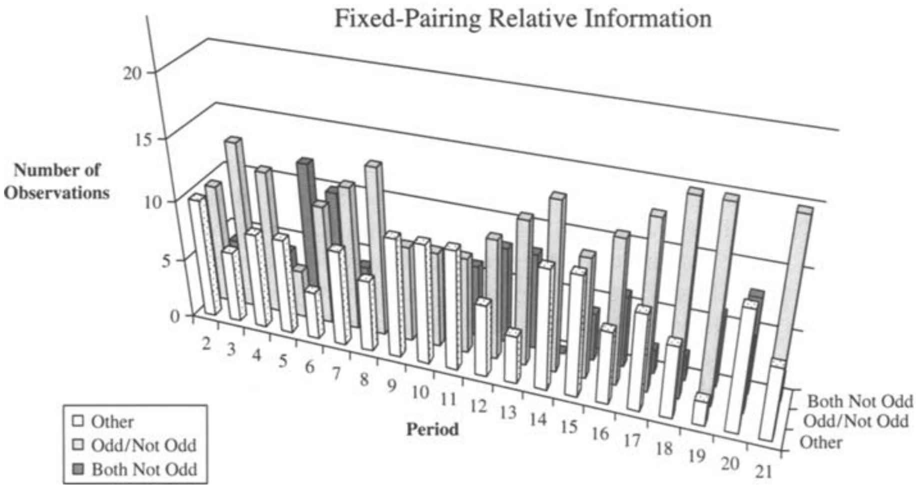


Figure 6. Frequency of Paired Choices.

such paired choices were made in period two and thirteen such choices were made in period twenty-one with the frequencies of Both Not Odd choices similarly graphed for the periods in between. The graph documents not only the high frequency of the Both Not Odd choice and its sustainability but also the demise of the Other category of paired choices.

Figure 6 describes the frequency of paired choices in the fixed-pairing relative information treatment. Again, the figure documents the coordination problem in this

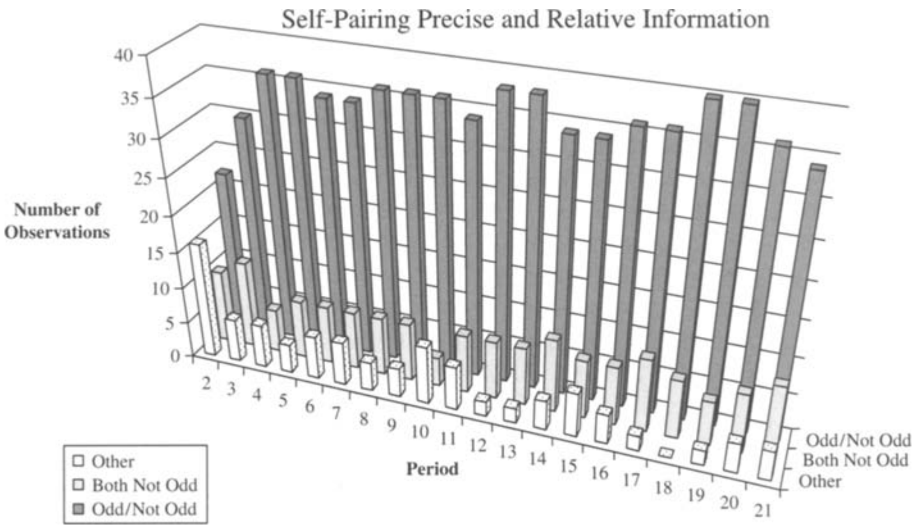


Figure 7. Frequency of Paired Choices.

treatment. All three paired choices, Both Not Odd, Odd/Not Odd and Other, were chosen throughout the treatment.

The self-pairing treatments of precise and relative information are presented in Figure 7; the two information treatments are again combined because of their similar play. The graph documents the high frequency and sustainability of the paired choice of Odd/Not Odd (from theory, chosen and unchosen). The Both Not Odd choice occurs with less frequency but is sustained throughout the treatments. The Other category of choices tends to die off over the treatments.

5. SUMMARY

Spatial dispersion games are characterized by multiple, non-strict equilibria. It is an open question whether players can select and attain an equilibrium in a spatial dispersion game. If equilibrium can be achieved, how long will it take and what are its characteristics. A natural question to also ask is whether the insights from matching games extend to dispersion games?

Our principal finding is that in spatial dispersion games, strategic interaction magnifies the role of cognitive constraints when compared to matching games. Players in the fixed-pairing relative information treatment had a difficult time coordinating their actions in order to achieve a dispersion outcome. This result contrasts with the result in Blume and Gneezy [2002], where in matching games relative information increased coordination compared to precise information, and Blume, DeJong and Maier [2003], where three sector matching games with relative information achieved a high level of coordination.

With these cognitive constraints in the fixed-pairing relative information treatment, pairs of agents failed to solve the dispersion problem that posed little or no problem for individual agents. In the self-pairing treatments, players were very successful in achieving dispersion outcomes. While some players coordinated by choosing right or left of the odd sector on both screens, most players coordinated by selecting the "least costly" way to coordinate, selecting the odd and not odd sectors. Thus, in both information treatments with self-pairing, we find that the mode used by individual agents to solve the dispersion problem is the same, odd and not odd.

When we remove the cognitive constraints in our design, pairs of agents solve the same problem just as well as individuals do. The frequency of dispersion outcomes in the fixed-pairing precise information treatment is comparable to both self-pairing treatments. However, the dispersion outcomes were different. Consistent with theory, players essentially coordinated by both players choosing left or right of odd in the fixed-pairing precise information treatment. In the self-pairing treatments, the majority of players picked the least costly way to coordinate, selecting the odd and not odd sectors.

ACKNOWLEDGMENT

The authors thank the National Science Foundation for its support.

NOTES

1 The issue of coordination via dispersion is extensively studied; other examples include Rapoport, Lo and Zwick [2002] and Zwick, Rapoport and Lo [2002] in which agents must disperse across several "locations" where the probability of success is inversely related to the number of agents at a location. Ochs [1999] is another example of spatial coordination in a market entry game.

2 As well as other structures that incorporate relative position, temporal order, size, brightness, modularity, etc.

3 For the sectors to be identical, it is important that the orientation of the disc is not common to both players. This can be achieved by spinning the disc before presenting it to each player pair. Furthermore, we wish to eliminate asymmetries arising from a directional structure on the disc (clockwise vs. counter-clockwise). This can be achieved by having agents in a match choose from opposite sides of the disc, which is presented to each player before each choice.

4 This type of optimal learning has been analyzed by Crawford and Haller [1990] and Blume [2000]; other applications of this idea can be found in Alpern and Reyniers [2002], Bhaskar [2000] and Kramarz [1996].

5 First introduced into the literature by cognitive and language psychologist Karl Bühler [1907, 1908] as Aha-Erlebnis; literally described by the situation in which one encounters a difficult foreign thought, hesitates and then suddenly attains the insight. Köhler [1925] studied the aha-experience experimentally with chimpanzees and Weber [2003] is an application applied to human psychology.

6 For the sectors to be identical, it is important that the orientation of the disc is not common to both players. This is achieved by spinning the disc before presenting it to each player. Furthermore, we wish to eliminate asymmetries arising from a directional structure on the disc (clockwise versus counter-clockwise and up versus down). This can be achieved by having agents in a match choose from opposite sides of the disc, or by randomizing over the side which is presented to each player before each choice. The randomizing scheme is more powerful in preserving symmetries, but for our purposes the opposite-side scheme suffices.

[7] Instructions are available from the authors upon request.

[8] To appreciate the difference between a coarse- and fine-language player, note that there can be two unchosen sectors when players match by selecting the odd sector or there can be one unchosen sector when players achieve a dispersion outcome by not selecting the odd sector.

[9] Others have studied differences in cognition in games; however, in the games considered players interests are not necessarily perfectly aligned. Nagel [1995] studied players' ability to reason through iterative dominance in the guessing game; Stahl and Wilson [1995] and Costa-Gomes, Crawford and Broseta [2001] studied players' varying abilities in dominance solvable games and games with unique equilibria.

[10] Choices in period one are *de facto* random. Thus, period two is the first period in which to observe player choices.

REFERENCES

Alpern, S. and D. Reyniers. (2002). "Spatial Dispersion as a Dynamic Coordination Problem." *Theory and Decision*, 53, 29–59.

Alpern, S. and S. Gal. (2003). "The Theory of Search Games and Rendezvous." Kluwer Academic Publishers, Boston, Massachusetts.

Bacharach, M. (1993). "Variable Universe Games," in K. Binmore, A. Kirman, and P. Tani (eds), *Frontiers of Game Theory*, MIT Press, Cambridge, Massachusetts.

Bhaskar, V. (2000). "Egalitarianism and Efficiency in Repeated Symmetric Games," *Games and Economic Behavior*, 32, 247–262.

Blume, A. (2000). "Coordination and Learning with a Partial Language." *Journal of Economic Theory*, 95, 1–36.

Blume, A., D. DeJong and M. Maier. (2003). Learning Strategic Sophistication, University of Iowa working paper.

Blume, A. and U. Gneezy. (2000). "An Experimental Investigation of Optimal Learning in Coordination Games." *Journal of Economic Theory*, 90, 161–172.

Blume, A. and U. Gneezy. (2002). Cognitive Forward Induction and Coordination without Common Knowledge: Theory and Evidence, University of Pittsburgh and Technion Working Paper.

Bühler, K. (1907–1908). "Tatsachen und Probleme zu einer Psychologie der Denkvorgänge." *Archiv fur die Gesamte Psychologie*, 9, 297–365 and 12, 1–92.

Costa-Gomes, M., V. Crawford and B. Broseta. (2001). "Cognition and Behavior in Normal Form Games: An Experimental Study." *Econometrica*, 69, 1193–1235.

Crawford, V. and H. Haller. (1990). "Learning How to Cooperate: Optimal Play in Repeated Coordination Games." *Econometrica*, 58, 581–596.

Dekel, E., B. Lipman and A. Rustichini. (1998). "Standard State-Space Models Preclude Unawareness." *Econometrica*, 66, 159–173.

Downs, A. (1957). *An Economic Theory of Democracy*, Harper and Row.

Geanakoplos, J. (1992). "Common Knowledge." *The Journal of Economic Perspectives*, 6, 53–82.

Hotelling, H. (1929, "Stability in Competition." *Economic Journal*, 41–57.

Köhler, F. (1925). "The Mentality of Apes," Harcourt, Brace, New York, New York.

Kramarz, F. (1996). "Dynamical Focal Points in N-Person Coordination Games." *Theory and Decision*, 40, 277–313.

Li, J. (2003). "Unawareness," University of Wisconsin-Madison Working Paper.

Nagel, R. (1995). "Unraveling in Guessing Games: An Experimental Study." *American Economic Review*, 85, 1313–1326.

Ochs, J. (1999). "Coordination in Market Entry Games," in D. Budescu, I. Erev and R. Zwick (eds) Games and Human Behavior: Essays in Honor of Amnon Rapoport, Lawrence Erlbaum Associates, Inc., Mahwah, New Jersey.

Palfrey, T. (1984). "Spatial Equilibrium with Entry," *Review of Economic Studies*, 51, 139–156.

Rapoport, A., A. Lo, R. Zwick, (2002). "Choice of Prizes Allocated by Multiple Lotteries with Endogenously Determined Probabilities." *Organizational Behavior and Human Decision Processing*, 87, 180–206.

Salop, S. (1979). "Monopolistic Competition with Outside Goods." *The Bell Journal of Economics*, 10, 141–156.

Schipper, B.C. (2002). "A State-Space Structure with Unawareness," University of Bonn Working Paper.

Stahl, D. and P. Wilson. (1995). "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10, 218–54.

Weber, R.A. (2003). "Learning and Transfer of Learning with No Feedback: An Experimental Test Across Games," Carnegie Mellon University working paper.

Zwick, R., A. Rapoport and A. Lo, (2002). "Behavioral Strategies in Repeated Coordination Games," in R. Zwick and A. Rapoport (eds) Experimental Business Research. Kluwer Academic Publishers, Boston, Massachusetts.

Chapter 9

COGNITIVE HIERARCHY: A LIMITED THINKING THEORY IN GAMES

Juin-Kuan Chong
Sungkyunkwan University and National University of Singapore

Colin F. Camerer
California Institute of Technology

Teck-Hua Ho
University of California-Berkeley

Abstract

Strategic thinking, best-response, and mutual consistency (equilibrium) are three key modeling principles in non-cooperative game theory. Camerer, Ho and Chong (2004) relax mutual consistency to predict how players are likely to behave in one-shot games before they can learn to equilibrate. They introduce a one-parameter cognitive hierarchy (CH) model to predict behavior in one-shot games, and initial conditions in repeated games. The CH approach assumes that players use k steps of reasoning with frequency $f(k)$. Camerer, Ho and Chong (2004) assume $f(k)$ to be a one-parameter Poisson distribution.

This paper investigates and lends support to the generality and precision of this Poisson CH model in three ways: 1. an unconstrained general distribution CH model is found to offer only marginal improvement in fit over its Poisson cousin and hence this suggests that the Poisson approximation is reasonable; 2. the steps of thinking players used in games are found to correlate with response time and schools they attend which suggests that cognitive hierarchy captures realistically a reasoning mechanism that goes on in the brain of these players; and 3. several classes of interesting economic problems, including asset pricing and business entry can be explained by the iterated reasoning of the Poisson CH model. When compared to the Quantal Response Equilibrium model which relaxes the best-response assumption of equilibrium theory, the better fit of Poisson CH model seem to suggest that mutual consistency is a more plausible assumption to relax in explaining deviation from equilibrium theory.

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 203–228.
© 2005 *Springer. Printed in the Netherlands.*

1. INTRODUCTION

Most theories of behavior in games assume that players think strategically (that is, they form beliefs by analyzing what others might do) and choose optimal responses given their beliefs. Precision often comes from the further assumption that each player's belief is consistent with what the other players actually plan to choose – i.e., players are in equilibrium.

The modern view in game theory is that equilibrium arises from adaptation, evolution, communication, or imitation. Since these processes take time, equilibration should not occur instantly in one-shot games. Indeed, mutual consistency is routinely violated in one-shot game experiments. In Bertrand pricing games, for example, players always have an incentive to undercut, so price should immediately go to marginal cost. But players do not immediately price at marginal cost (though they learn to over time; e.g., Capra, 1999). Another example is the "p-beauty contest" game. In this game players pick numbers from 0 to 100 and the player whose number is closest to p times the average wins a fixed prize. The equilibrium is zero, but average choices in one-shot games with $p = 2/3$ are typically from 20 to 35, reflecting lack of mutual consistency.

The goals of a non-equilibrium theory are to explain data more accurately while weakening mutual consistency, and to retain the precision and generality which makes equilibrium theories so useful. Camerer, Ho and Chong (2004) propose one such approach, which is called "cognitive hierarchy" (CH). The CH model weakens the assumption of mutual consistency by having the players form different beliefs of what others will do. The differences in belief arise from the different iterations of strategic thinking that players perform – steps of thinking. Formally, the CH model has two components – (1) *decision rules* for players doing k steps of thinking, and (2) a *frequency distribution* of steps $f(k)$.

The *decision rules* reflect an iterated process of strategic thinking: a player who does k steps of thinking figures out what players who perform less thinking are likely to do, and best-responds given those beliefs. The iterative process begins with step-0 types who choose randomly across all possible strategies with equal probability. The *distribution* $f(k)$ characterizes the distribution of thinking steps in the population of players. Camerer, Ho and Chong (2004) assume $f(k)$ to be Poisson, which has only one parameter τ, its mean and variance. The estimates of τ in Camerer, Ho and Chong (2004) generally fell between 1 to 2 and fit actual choices reasonably well, and were usually better than equilibrium predictions. These estimates were robust because fixing a common τ across games fits almost as well as estimating τ separately for each game.

In games with mixed-strategy equilibria, frequencies were often surprisingly close to the mixture rates predicted by equilibrium. The Poisson CH model fit these data fairly well (with median $\tau = 1.78$). Equilibrium theory was less successful at games like beauty contest where the games stopped at an average around 30, rather than converging to the equilibrium of zero. In a separate empirical analysis of 24 p-beauty contest data sets by Camerer, Ho and Chong (2003), the median estimate $\tau = 1.61$. The conclusion that people appeared to do only one or two steps of thinking, on average, explained why the convergence process in these games stopped at around 30. Therefore, the Poisson CH model appears to explain limited equilibration in games like beauty contests, *and* the surprising degree of "instant equilibration" that occurs in entry and mixed-equilibrium games, with similar values of τ.

Camerer, Ho and Chong (2003, 2004) offer mounting empirical evidence why the Poisson CH model should be taken seriously as a general and precise model of limited thinking. In this paper, we lend additional support in the following ways:

1. The Poisson assumption of the $f(k)$ distribution offers good approximation to the distribution of thinking steps. To this end, we compare and contrast fits and accuracy measure of a general distribution CH model with that of the Poisson CH model. The result supports the choice of Poisson to be a parsimonious approximation. Human subjects are constrained by limited working memory and hence as k rises, fewer and fewer players will do the next level of thinking. Poisson expresses this constraint in reduced form.

   We also contrast with the empirical result for Quantal Response Equilibrium (QRE) model, which relaxes best-response assumption of equilibrium theory, to show that limited thinking is a more plausible explanation for the empirical deviation from equilibrium theory.
2. The steps of thinking players do in the cognitive hierarchy capture reduced-form outputs of some cognitive mechanism. In other words, τ capture the essence of cognitive skills of the players. The theory can be tested with cognitive data such as self-reports, tests of memory, response times, measures of eye gaze and attention (Camerer et al., 1994; Costa-Gomes, Crawford and Broseta, 2001), or even brain imaging (cf. Camerer, Loewenstein, and Prelec, 2002). To this end, we link the thinking steps of players to their demographic details and response time. In general, we find that subjects tend to improve their thinking steps over time and this is accompanied by an increase in response time.
3. The Poisson CH model of iterated thinking can be applied to several interesting problems in economics, including asset pricing, speculation, competition neglect in business entry, incentive contracts, and macroeconomics. We show how the CH model can account for two patterns of broad economic interest – speculation and money illusion.

   The CH model is also a useful complement to other business research such as tax compliance (Kim and Waller, 2004), investment in collaborative networks (Amaldoss and Rapoport, 2004) and internet congestion (Friedman and Huberman, 2004), where mutual *in*consistency and heterogeneity in limited cognition could be the key factors that drive the empirical observation.

The paper is organized as follows. The next section describes briefly the CH model. Section III reports the empirical analysis of the general distribution CH model and the QRE model and contrasts it with the Poisson CH model. Section IV presents the analysis of response time and thinking steps. Section V focuses on the application of CH model to speculation and money illusion. Section VI concludes and sketches future research.

2. THE COGNITIVE HIERARCHY (CH) MODEL

A precise thinking steps theory needs two components: Decision rules for what players using k step of thinking in the cognitive hierarchy will do, and a distribution of thinking steps $f(k)$. Let players be indexed by i and strategies by j and j'. Player i has m_i strategies denoted by s_i^j; denote other players' (denoted) strategies by $s_{-i}^{j'}$, and player i's payoffs by $\pi_i(s_i^j, s_{-i}^{j'})$.

We assume that 0 step players are not thinking strategically at all; they randomize equally across all strategies. Other simple rules could be used to start the cognitive hierarchy process off, but equal randomization has some empirical and theoretical advantages.[1] Zero-step thinkers may also be "unlucky" rather than "dumb". Players who start to analyze the game carefully but get confused or make an error might make a choice that appears random and far from equilibrium (much as a small algebra slips in a long proof can lead to a bizarre result). Denote the choice probability of step k for strategy S^j by $P_k(S^j)$. So, we have $P_0(s_{-i}^{j'}) = 1/m_{-i}$ in a 2-player game.[2]

Players doing one or more steps of thinking are assumed to not realize that others are thinking as "hard" as they are (or harder), but they have an accurate guess about the relative proportions of players using fewer steps than they do. Formally, players at step k know the true proportions $f(0), f(1), \ldots, f(k-1)$. Since these proportions do not add to one, they normalize them by dividing by their sum. That is, step-k players believe the proportions of players doing h steps of thinking are $g_k(h) = f(h)/\sum_{l=0}^{k-1} f(h)$, $\forall h < k$ and $g_k(h) = 0$, $\forall h \geq k$. Given these beliefs, the expected payoff to a k-step thinker from strategy is

$$E_k(s^j)\sum_{j'=1}^{m_{-i}} \pi_i(s_i^j, s_{-i}^{j'})*\left[\sum_{h=0}^{k-1} g_{k(h)} \cdot P_h(s^{j'})\right]$$

We assume players best-respond (or randomize equally if two or more strategies have identical expected payoffs).

The normalized-beliefs assumption $g_k(h) = f(h)/\sum_{l=0}^{k-1} f(l)$, has an interesting property we call "increasingly rational expectations". To see what this means, first note that the absolute total deviation of step-k's beliefs and true frequencies is $D(k) = \sum_{h=0}^{\infty} |f(h) - g_k(h)|$. Then consider how large this total deviation is for players at different levels of the cognitive hierarchy.

Zero-step thinkers have no beliefs at all. One-step thinkers believe everyone is doing 0 steps of thinking (i.e., $g_1(0) = 1$); since only $f(0)$ are doing 0 steps of thinking the one-step beliefs are wrong by a total absolute deviation of $D(1) =$

$1 - f(0) + \sum_{h=1}^{\infty} |f(h) - 0|$ which comes to $D(1) = 2 \cdot [1 - f(0)]$. Two-step thinkers believe $g_2(0) = f(0)/[f(0) + f(1)]$ and $g_2(1) = f(1)/[f(0) + f(1)]$. Since the actual frequencies are $f(0)$ and $f(1)$ the sum of the deviations between their beliefs and the true frequencies is $D(2) = g_2(0) - f(0) + g_2(1) - f(1) + \sum_{h=2}^{\infty} |f(h) - 0|$. A little algebra shows that this total deviation is $D(2) = 2 \cdot [1 - f(0) - f(1)]$, which is smaller than the size of the 1-step thinkers' belief error, $D(1)$. In fact, it is easy to show[3] that the total deviation $D(k)$ falls monotonically as k increases. The reason is that the "missing" belief $\sum_{h=k}^{\infty} f(h)$ which is reallocated by the step-k thinker to the lower-step types shrinks as k grows large. The k-step thinkers' beliefs gradually come closer and closer to the truth. (In Stahl and Wilson's (1995) terminology the highest-step thinkers become "worldly".)

The fact that beliefs converge as k grows large has another important implication: As the missing belief grows small, players who are doing k and $k + 1$ steps of thinking will have approximately the same beliefs, and will therefore have approximately the same expected payoffs. While we have endogenized the mean number of thinking steps, this convergence property is a clue about why players will do only a few steps of thinking – it doesn't pay to think too hard, because doing k steps and $k + 1$ steps yields roughly the same expected payoff. If the number of steps of thinking is endogenized by some kind of comparison of benefits (marginal expected payoffs from thinking more) and cognitive costs, the fact that the expected payoffs of higher-step thinkers converges will lead to a natural truncation which limits the amount of thinking.

2.1. *Alternative specifications of limited thinking*

Once the mutual consistency of choices and beliefs is relaxed, there are many ways to specify choices and beliefs that are *not* consistent. The cognitive hierarchy model in which k-step thinkers believe everyone else does $k - 1$ or fewer steps is one specification, but others spring to mind.

One alternative specification is to assume that step-k thinkers believe everyone else is doing $k - 1$ steps (i.e., $g_k(h) = I(k - 1, h)$ where $I(x, y)$ is an identity function). Call this the "$k - 1$" specification. Preliminary estimates showed that the $k - 1$ model fits about as accurately as our specification in three sets of matrix games.[4] However, the $k - 1$ model has some unfortunate properties. The $k - 1$ model is a freeze-frame version of Cournot dynamics, in which a player always believes others will repeat their choices in the most recent period and best responds to that belief. Since it is possible for play to cycle endlessly in Cournot, the $k - 1$ model can cycle too.

Furthermore, in the $k - 1$ model, players doing more and more steps of thinking do not become more worldly – in fact, their beliefs diverge from rational expectations as k increases.[5] The belief deviations in the $k - 1$ model are also larger than in our specification; in a sense, the $k - 1$ thinkers have "less rational" expectations than in our approach.[6] Furthermore, because k-step thinkers' beliefs do not converge to

Table 1. Asymmetric Matching Pennies

	L	R
T	(x, 0)	(0, 1)
B	(0, 1)	(1, 0)

the correct distribution in the $k - 1$ model, their beliefs embody a double dose of overconfidence. Two-step thinkers, for example, think that all players are doing one step of thinking, and think that all the one-step thinkers are completely deluded in thinking there are 0-step thinkers.

It is also easy to find games in which the $k - 1$ specification fits data very poorly. In the market entry games discussed in section IV below, the $k - 1$ model predicts a step function – the rate of entry into a capacity-constrained market will depend only on whether the capacity is less than half of the number of entrants, or more than half. But the data are surprisingly monotonic in the capacity, so the predicted step function is a poor approximation.[7] Another example is asymmetric matching pennies, shown in Table 1.

For $x > 1$ the model in which players think everyone is one step below them makes the same prediction for every value of x. But row players actually choose T more often when x is larger, a fact which is anomalous for the $k - 1$ specification but is predicted by the CH model specification.[8]

In the CH model, players who do k steps of thinking are not aware that others might be thinking like they do (or even thinking harder). An alternative approach is to make players "self-aware" so that k-step players' beliefs include the possibility that there are others doing k steps like themselves (e.g., $g_k(c) = f(c)/\Sigma_{l=0}^{k} f(l)$, for $0 \leq c \leq k$ and $g_k(c) = 0$ otherwise).

Selten (1998) argues that the "circular" reasoning implicit in self-awareness is cognitively less plausible than a purely sequential procedure. Self-awareness is deliberately excluded from the CH model for that reason, and several others. One other reason is that overconfident players will doubt others are thinking as much as they themselves are. If players all think they are "smarter" (or harder-thinking) than others then they *will* neglect the possibility that others are thinking at their step. This sort of overconfidence about relative skill is well-documented in some economic domains (e.g., Roll, 1984; Camerer and Lovallo, 1999; Benabou and Tirole, 2002).

Including self-awareness also leads to a model which is very much like a noisy equilibrium or quantal response equilibrium (QRE) model, because including self-awareness reintroduces an element of the mutual consistency which is precisely what the CH approach jettisons. To see this, first note that the relative proportion of 0 and 1-step thinkers, conditional on thinking only up to 1 step, is $f(0)/(f(0) + f(1)) = 1/(1 + t)$. For large t this fraction will be small, which means 1-step thinkers believe most others are 1-step thinkers too (with a small fraction of 0-step randomizers thrown in the mixture). The 1-step thinkers' optimal choices will be the solution to a recursive equation which requires (approximate) mutual consistency – bringing us back near Nash equilibrium. But the point of the cognitive hierarchy approach was to *improve* on the predictive accuracy of Nash equilibrium for one-shot games; circling back towards equilibrium defeats the purpose of creating a different approach.

Self-awareness also adds computational difficulty, compared to the CH specification, because it requires solving for fixed points. This is especially cumbersome in games with large strategy spaces or many players. Finally, and perhaps most importantly, in earlier work we *did* compare CH models with and without self-awareness. Adding self-awareness always reduced explanatory power, often substantially.[9]

3. ESTIMATION AND MODEL COMPARISON

This section estimates a general distribution CH model in which the frequencies of k-step thinkers, $f(k)$, are not constrained to satisfy the Poisson distribution (and truncated at six steps)[10]; and the QRE model, and compares the results to the Poisson CH model reported in Camerer, Ho and Chong (2004).

We use maximum likelihood (MLE) techniques to estimate the general CH model and the QRE model. We fit six data sets: 33 matrix games with 2–4 strategies from three data sets; 22 games with mixed equilibria (new data); the binary entry game described above (new data)[11]; and 7 sender-receiver signaling games.

The matrix games are 12 games from Stahl and Wilson (1995), 8 games from Cooper and Van Huyck (2003) (used to compare normal- and extensive-form play), and 13 games from Costa-Gomes, Crawford and Broseta (2001). All these games were played only once with feedback, with sample sizes large enough to permit reliable estimation.

The 22 games with mixed-equilibria are taken from those reviewed by Camerer (2003, chapter 3), with payoffs rescaled so subjects win or lose about \$1 in each game (see Appendix for details). These games were run in four experimental sessions of 12 subjects each, using the "playing in the dark" software developed by McKelvey and Palfrey. Two sessions used undergraduates from Caltech and two used undergraduates from Pasadena City College (PCC), which is near Caltech.

The binary entry game is the one described above. In the four experimental sessions, each of 12 players simultaneously decides whether to enter a market with announced capacity c. If c or fewer players enter the entrants earn \$1; if more than c enter they earn nothing. Not entering earns \$.5. In this simple structure, risk-neutral players care only about whether the expected number of entrants will be less than $c - 1$.[12] Subjects were shown five capacities c – 2, 4, 6, 8, 10 – in a fixed random order, with no feedback. Finally, the 7 sender-receiver signaling games were studied by Banks, Camerer and Porter (1994) to explore which signaling game refinements (intuitive criterion, divinity, and universal divinity, etc.) predict best when there are multiple Nash equilibria.

3.1. Which models fit best?

How does the CH Poisson specification compare to the general distribution CH and QRE? Table 2 shows log likelihoods (LL) and mean-squared deviations for these models estimated game-by-game or with common parameters across games in a dataset. This Table answers several questions. Focusing first on the CH Poisson model, moving from game-specific estimates of τ to common within-column estimates only degrades fit badly in the Stahl-Wilson data; in the other samples imposing a common τ fits about as well as letting τ vary in each game.

Table 2 also reports fits from the general CH model. Except for the Stahl-Wilson data (once again), imposing this 6-parameter general specification degrades fit very little compared to the Poisson distribution. Table 3 shows the fractions of players estimated to use each level in the general specification; these fractions are reasonably close to those constrained by the Poisson distribution. This suggests that the Poisson distribution is a highly plausible approximation of the distribution of thinking steps.

The CH Poisson model also fits substantially better than QRE (and hence, better than Nash), or about as well, except in the Stahl-Wilson games when common parameters are imposed. This result does *not* mean QRE research (which imposes mutual consistency but relaxes optimization) should be abandoned in favor of the CH approach (which does the opposite, relaxing consistency and retaining optimization); our view is that both approaches should be explored further. But the relative success of CH in many games is an indication that mutual consistency is not *necessary* to produce a model that fits data from one-shot games reasonably well.

The CH model retains optimization but relaxes mutual consistency. Quantal response equilibrium is a complementary approach, which retains mutual consistency but relaxes optimization (Rosenthal, 1989; Chen, Friedman and Thisse, 1996; McKelvey and Palfrey, 1995, 1998; Goeree and Holt, 1999). QRE weakens the best-response property in Nash equilibrium and substitutes statistically-rational expectations in the sense that a player's belief about the distribution of play by others matches the actual distribution.

In empirical applications, QRE and CH will usually predict deviations in the same direction from Nash equilibrium, and they should be treated as alternative paths which both deserve exploration. However, keep in mind that in the results above, QRE fit and predicted a little less accurately than CH (except in the Stahl-Wilson matrix games). Because they will often predict similar deviations, it is useful to carefully distinguish how they differ. QRE will generally make different predictions when games are subject to "inessential transformations" (see Dalkey, 1953; Ho and Weigelt, 1996). For example, in QRE "cloning" strategies (adding precisely equivalent strategies) will generally increase the frequency of play of the set of cloned strategies (because players who noisily best-respond will play these strategies equally often).[13] In CH, in contrast, cloning strategies will only increase how often the cloned strategy (set) is played for 0-step thinkers. If the cloned strategy has the highest expected payoffs, then higher-step thinkers are assumed to randomize across the set of equally-good (and best) responses so they will play a *set* of best responses

Table 2. Model Fit (Log Likelihood LL and Mean Squared Deviation MSD)

Data set	*Stahl & Wilson (1995)*	*Cooper & Van Huyck (2003)*	*Costa-Gomes et al. (2001)*	*Mixed*	*Entry*	*Signal*
Within-dataset Forecasting						
Log-likelihood						
Cognitive Hierarchy (Game-specific τ)[1]	–360	–838	–264	–824	–150	–191
Quantal Response (Game-specific λ)	–395	–871	–281	–897	–150	–199
Cognitive Hierarchy (Common General Distribution $f(k)$)	–388	–867	–271	–866	–150	–204
Cognitive Hierarchy (Common τ)	–458	–868	–274	–872	–150	–205
Quantal Response (Common λ)	–421	–916	–292	–1005	–151	–207
Mean Squared Deviation						
Cognitive Hierarchy (Game-specific τ)	0.0074	0.0090	0.0035	0.0097	0.0004	0.0480
Quantal Response (Game-specific λ)	0.0153	0.0157	0.0111	0.0203	0.0001	0.0489
Cognitive Hierarchy (Common General Distribution $f(k)$)	0.0125	0.0142	0.0090	0.0175	0.0004	0.0481
Cognitive Hierarchy (Common τ)	0.0327	0.0145	0.0097	0.0179	0.0005	0.0489
Quantal Response (Common λ)	0.0248	0.0276	0.0180	0.0456	0.0022	0.0506
Cross-dataset Forecasting						
Log-likelihood						
Cognitive Hierarchy (Common General Distribution $f(k)$)	–422	–918	–338	–884	–154	–227
Cognitive Hierarchy (Common τ)	–469	–956	–293	–884	–154	–238
Quantal Response (Common λ)	–431	–990	–374	–1402	–166	–216
Mean Squared Deviation						
Cognitive Hierarchy (Common General Distribution $f(k)$)	0.0265	0.0270	0.0554	0.0194	0.0043	0.0759
Cognitive Hierarchy (Common τ)	0.0416	0.0335	0.0237	0.0215	0.0046	0.0803
Quantal Response (Common λ)	0.0275	0.0506	0.0702	0.0487	0.0216	0.0632

Note 1: The scale sensitivity parameter λ for the Cognitive Hierarchy models is set to infinity. The results reported in Camerer, Ho & Chong (2001) presented at the Nobel Symposium 2001 are for models where λ is estimated.

Table 3. Probability Distribution of Thinking Levels for the General Cognitive Hierarchy Models

Data set	*Stahl & Wilson (1995)*	*Cooper & Van Huyck (2003)*	*Costa-Gomes et al. (2001)*	*Mixed*	*Entry*	*Signal*
Frequency Estimates of the General Cognitive Hierarchy Models with Constraints[1]						
Thinking Levels						
0	0.25	0.42	0.21	0.20	0.50	0.79
1	0.12	0.44	0.21	0.38	0.40	0.09
2	0.12	0.11	0.27	0.23	0.08	0.09
3	0.12	0.03	0.19	0.08	0.01	0.03
4	0.12	0.01	0.09	0.04	0.00	0.00
5	0.12	0.00	0.03	0.04	0.00	0.00
6	0.12	0.00	0.01	0.01	0.00	0.00
7 and Higher	0.00	0.00	0.00	0.00	0.00	0.00
Frequency of the Poisson Cognitive Hierarchy Models						
Thinking Levels						
0	0.21	0.44	0.18	0.23	0.48	0.80
1	0.33	0.36	0.31	0.34	0.35	0.18
2	0.25	0.15	0.27	0.25	0.13	0.02
3	0.13	0.04	0.15	0.12	0.03	0.00
4	0.05	0.01	0.07	0.05	0.01	0.00
5	0.02	0.00	0.02	0.01	0.00	0.00
6	0.00	0.00	0.01	0.00	0.00	0.00
7 and Higher	0.00	0.00	0.00	0.00	0.00	0.00

Note 1: The constraints imposed are: (1) if $f(k) < f(k-1)$ then $f(k+x) \leq f(k)$ where $f(k)$ is the frequency estimate for thinking level k; (2) $f(6) \geq f(7$ and higher).

just as often as if one strategy was a uniquely best response. A similar property arises if strategies are amalgamated rather than cloned. Mookherjee and Sopher (1997) found that amalgamating strategies did not change how frequently they were played, which goes against QRE and is more consistent with CH.

Another subtle contrast between the two models is when some strategies are nearly dominated. For example, if one strategy yields ε less than another strategy, than as $\varepsilon \to 0$ the QRE frequencies of the two strategies will become equal. Since the CH approach assumes best responses, for any $\varepsilon > 0$ the predicted frequency of the dominated strategy will be lower than the predicted frequency of the strategy which dominates it. This contrast sharpens the difference between the two approaches. If subjects spot dominated strategies (regardless of the degree of dominance) and never play them, the CH approach will predict better than the QRE approach; oppositely, if subjects do not notice or care about small degrees of dominance then they will play nearly-dominated strategies relatively often consistent with QRE.

Another difference is that the CH model naturally generates heterogeneity – "spikes" which can potentially match spikes in data. The p-beauty contest is an example.[14] QRE, in contrast, predicts a smooth statistical distribution with no spikes (the same is true of Capra's, 1999, and Goeree and Holt's, 2004, models). In the same way, CH can easily explain endogeneous purification but the simplest form of QRE cannot (in QRE each player mixes with the same statistical distribution across strategies).

Finally, our analysis of mixed games shows that MLE estimation recovers the correct τ parameters in modest samples (around 50) when the true model is CH (see Table 8). However, when samples are small, sampling error is 'accurately' fit by QRE with a low response sensitivity λ. So we suspect that MLE and other techniques will generally underestimate the true value of λ (i.e., estimates are biased downward) in small or medium samples.

3.2. *Predicting across games*

Good theories should predict behavior in new situations. A simple way to see whether the CH model can do this, within a large sample of games, is to estimate the value of τ on $n - 1$ games and forecast behavior in each holdout sample separately. (This is a roundabout way to test how stable τ appears to be across games, and also whether small variations in estimated τ create large or small differences in predicted choice frequencies.) The bottom panel of Table 2 reports the result of this sort of cross-game estimation. Both the CH Poisson and QRE models fit cross-game a little less accurately than when estimates are common within games. QRE degrades particularly badly in the Costa-Gomes et al. and mixed-equilibrium games. This is not surprising since the free parameter in the QRE model is a response sensitivity which is sensitive to changes in payoff scales (e.g., McKelvey, Palfrey and Weber, 2000).

4. THINKING STEPS, RESPONSE TIME AND COGNITION

The CH model makes predictions about the kinds of algorithms that players use in thinking about games. This means that cognitive data other than choices can – like belief-prompting, response times, information lookups, or even brain imaging – in principle, be used to test the model.[15]

Several studies show that prompting players for beliefs about what others will do actually changes their choices, typically moving them closer to equilibrium. A simple example was first demonstrated by Giovanna Devetag and Eldar Shafir and replicated by Warglien, Devetag and Legrenzi (1998). Their game and results are shown in Table 4. If players think others are step 0 (randomizing), choosing X yields an (expected) payoff of 60 rather than 45 from choosing Y. When players simply choose (with financial incentives) 70% of the row players choose X. When subjects are prompted to articulate a belief about what the column players will do before they

Table 4. How belief-prompting promotes dominance-solvable choices by row players (Warglien, Devetag and Legrenzi, 1998)

row move	*column player*		*without belief prompting*	*with belief prompting*
	L	*R*		
X	60,20	60,10	0.70	0.30
Y	80,20	10,10	0.30	0.70

choose, 70% then choose the dominance-solvable equilibrium choice *Y* (see also Croson, 2000; and Hoffman et al., 2000).[17]

In the CH model, the fractions of *X* play are fit perfectly by $\tau = .58$ without belief-prompting and $\tau = 2.20$ after belief-promoting. This suggests that the effect of belief prompting is to encourage strategic thinking among players who might otherwise be 0-step and shift the entire distribution up by about a step and a half of thinking.

Another approach is to directly measure the information subjects acquire in a game by forcing subjects to "look up" payoffs in games (as in Camerer et al (1994), Costa-Gomes, Crawford, and Broseta (2001) and Johnson et al (2002)). Information lookups are another cognitive measure which we expect to be correlated to thinking steps. Johnson et al. show that how much players look ahead to future "pie sizes" in alternating-offer bargaining is correlated with the offers they make. Costa-Gomes et al show that lookup patterns are clearly correlated with choices that result from various (unobserved) decision rules. These patterns are not proof that models based on steps of thinking are correct, but they do illustrate a fresh prediction that results from these models.

One should not expect the average amount of thinking τ to have a universally constant value. It is like a risk-aversion parameter or a discount factor. Values of those parameters are typically not derived from first principles and are not expected to be constant. Discounting and risk-aversion vary across people (and even across a person's life; children are measurably more impatient than adults) and situations; τ probably does too. The hope is simply to find a range of τ values which are plausible, and regular enough to permit us to make guesses about behavior in new games with some confidence. In fact, the estimates of τ generally fall between 1 and 2 (see Table 5 for the estimates from both Camerer, Ho and Chong (2004) and this paper). And because the CH model has cognitive detail, τ *should* change in response to certain kinds of treatment effects. For example, people who are more analytically skilled or have special training in game theory will probably exhibit higher values of τ (more strategic thinking), just as people who are treated for fear of flying act as if a parameter characterizing their aversion to flight risk was changed by therapy.

Table 5. Parameter Estimate t for Cognitive Hierarchy Models

Data set	*Stahl & Wilson (1995)*	*Cooper & Van Huyck (2003)*	*Costa-Gomes* et al. *(2001)*	*Mixed*	*Entry*	*Signal*	*Guyer & Rapoport Rapoport (1972)*	
Game-specific τ								
Game 1	2.93	15.90	2.28	0.98	0.70	0.44	1.43	0.95
Game 2	0.00	1.07	2.27	1.71	0.85	0.00	1.71	0.61
Game 3	1.40	0.18	2.29	0.86	–	0.83	1.71	1.08
Game 4	2.34	1.28	1.26	3.85	0.73	1.08	1.2	0.69
Game 5	2.01	0.52	1.80	1.08	0.70	0.18	1.56	0.84
Game 6	0.00	0.82	1.67	1.13		0.02	1.71	0.53
Game 7	5.37	0.96	0.88	3.29		0.00	1.31	0.87
Game 8	0.00	1.54	2.18	1.84			1.66	1.07
Game 9	1.35		1.89	1.06			1.66	1.02
Game 10	11.33		2.26	2.26			1.66	1.71
Game 11	6.48		1.23	0.87			1.17	3.48
Game 12	1.71		1.03	2.06			1.2	2.39
Game 13			2.28	1.88			1.24	8.4
Game 14				9.07			1.35	0.81
Game 15				3.49			1.08	1.09
Game 16				2.07			0.67	
Game 17				1.14			0.75	
Game 18				1.14			4.13	
Game 19				1.55			1.66	
Game 20				1.95			1.66	
Game 21				1.68			4.54	
Game 22				3.06			1.83	
Median τ	1.86	1.01	1.89	1.77	0.71	0.18		1.31

If the algorithmic reasoning in the CH model is taken seriously as a model of human cognition, then the model can be tested by jointly estimating both choices *and* cognitive variables. As a first step, we need to ascertain the thinking steps that subjects used in games. This can be done using the posterior thinking step probability distribution.

4.1. Posterior Thinking Types of Players

Consider the 22 mixed equilibria games (Camerer, 2003, Chapter 3) collected using Play in the Dark software where there are 48 subjects playing the same sequence of 22 mixed games. We can derive the posterior thinking step probabilities (via Bayesian updating) for each subject in each game. Using these posterior probabilities, we can assign each subject a thinking type, i.e., the most likely thinking step the subject uses in the game. Let us describe how we posteriorly assign a type to each subject for each game. There are several ways one can assign a type to each subject in a game; for example we could assign a type based on either the mean, median or mode of the posterior type distribution. We use a more elaborate process to assign type using the posterior type distribution. In particular, we maximize the posterior likelihood of a subject being a type in a game with certain constraints. To be precise, let us define the constrained likelihood maximization problem.

First, we define some notations. We denote subject i for $i = 1, \ldots, N$ and type or level k where $k = 0, \ldots, K$. h_{ik} is the posterior probability of individual i belonging to level k (in game g, we drop the index g for notational simplicity.). x_{ik} is 1 if i is assigned type k and 0 otherwise. We want to solve the following problem for each game:

$$\text{Max } \Pi_i^N \Pi_{k=0}^K h_{ik} \cdot x_{ik},$$

subject to:

$$\Sigma_{k=0}^K x_{ik} = 1, \forall i, \qquad \text{Constraint 1}$$

$$\Sigma_{i=1}^N x_{ik} = \Sigma_{i=1}^N h_{ik}, \forall k, \qquad \text{Constraint 2}$$

where constraint 1 ensures that each subject can only be assigned to one type; and constraint 2 ensures that the sum of all assignments have to be the same as the posterior aggregate.

Note that maximizing $\Sigma_i^N \Sigma_{k=0}^K \ln(h_{ik}) \cdot x_{ik}$ is equivalent. Hence, we take advantage of this and convert the problem into a linear integer programming problem. If the right hand sides of the second set of constraints $\Sigma_{i=1}^N h_{ik}$ are integer (Observe that the sum of the RHSs of the second set of constraints is N, i.e., $\Sigma_{k=1}^K \Sigma_{i=1}^N h_{ik} = N$), we have a transportation problem which can solved relatively fast and easy. As such, we adopt the following procedure to ensure the integrality of the right hand sides. (for expositional convenience, let's denote $\Sigma_{i=1}^N h_{ik}$ as H_k.

We first round H_k to the nearest integers $\bar{H}_k$. It could happen that after rounding, we may end up with the situation that $\sum_{k=0}^{K} \bar{H}_k < N$. If this happens, we assign the residuals to $\bar{H}_k$ such that the resulting value, $\hat{H}_k$, has the property that $\max_k\{\hat{H}_k - N \cdot f_k\}$ is minimized. The rationale behind this procedure is that if we observe enough subjects, then the posterior type distribution should converge to the prior.

4.2. Thinking Type and Response Time

Rubinstein (2004) reports evidence where response time correlates with chosen action; longer response time accompanies more cognitive choice. The set of actions is divided into more instinctive or more cognitive response. The division is based on the depth of reasoning required to arrive at a particular choice action. This action division scheme is closely related to our thinking type where there is a best response associated with each type. Our analysis of the response time and thinking type yields some interesting differences compared to the findings of Rubinstein (2004). Our response time analysis looks at the impact of thinking type on response time and the evolution of thinking types as the subjects go through the sequence of games.

We use the posterior thinking types derived in the subsection IV.1 in the analysis. To proceed with the analysis, let us define the revealed type variable $t_{ig} = k$ if $x_{ik} = 1$ for game g. The mean of t_{ig} within each game should converge to the game-specific τ by construction and appear to be so in our analysis.[16]

We also consider three demographic characteristics in relation to both thinking type and response time: gender, school and a cognitive measure derived from a row-averaging exercise administered prior to the experiments. The Gender variable equals to 1 if male and 0 if female. The School variable equals to 1 if Caltech and 0 if Pasadena Community College (PCC).

We are interested in both the revealed type t_{ig} and response time r_{ig}. Consider the following:

$$r_{ig} = \alpha_t + \beta_{rt} \cdot t_{ig} + \beta_{gt} \cdot Gender + \beta_{st} \cdot School + \beta_{ct} \cdot Cognitive \qquad \text{(IV.2)}$$

Table 6 reports the regression results. We found that school and thinking type are the significant variables to predict response time. We find $\beta_{st} < 0$ which suggests that Caltech subjects take less time to respond, and $\beta_{rt} < 0$ which means that higher level subjects take shorter time to respond. A check on the correlation between types and response time reveals significant negative correlation. The fact that higher types require to perform more steps in their choice does not seem to lengthen the response time.[18] At first sight, this seems at odd with Rubinstein (2004). We argue that higher types are smarter and hence think faster; in other words, they are more efficient. Different subjects think at different speed; this argument is further verified by CalTech subjects who are generally higher types and take shorter time. It seems obvious that

Table 6. Analysis of Revealed Thinking Types and Response Time

Dependent Variable, Y	*Thinking Level[1], Y = L*		*Response Time, Y = R*		*Response Time, Y = R* (Level 0 excluded)	
	Coefficient	*p-value*	*Coefficient*	*p-value*	*Coefficient*	*p-value*
Independent Variables						
1. Regression $Y = \alpha + \beta_l^*L + \beta_g^*G + \beta_s^*S + \beta_c^*C$						
Level, β_l	–	–	−0.34	(0.00)	−0.10	(0.49)
Gender G, β_G	0.16	(0.19)	0.11	(0.82)	0.71	(0.20)
School S, β_S	0.53	(0.00)	−1.63	(0.00)	−1.13	(0.04)
Cognitive C, β_C	0.00	(0.74)	0.00	(0.35)	0.00	(0.14)
2. Regression $Y = \alpha + \beta_l^*L + \beta_{p1}^*P1 + \beta_{p2}^*P2 + \beta_{c2}^*C2$						
Caltech 7/17/2002 (Base Case)						
Level, β_l	–	–	−0.33	(0.00)	−0.14	(0.33)
PCC 7/29/2002 P1, β_{P1}	−0.71	(0.00)	−1.35	(0.03)	−1.97	(0.00)
PCC 7/31/2002 P2, β_{P2}	−0.84	(0.00)	3.31	(0.00)	2.42	(0.00)
Caltech 7/25/2002 C2, β_{C2}	−0.44	(0.01)	−0.98	(0.10)	−1.36	(0.04)

Note 1: The revealed thinking level is determined by maximizing the posterior probability of thinking type for all players subject to the constraint that the proportion of thinking levels conforms to the average posterior proportion.

Table 7. Comparison of Revealed Thinking Types and Response Time in First and Second Half of Games

	Thinking Level			Response Time		
	First Half	*Second Half*	*Diff*	*First Half*	*Second Half*	*Diff*
Total Subject Pool (Sample Size)	48			48		
Mean	1.24	1.93	0.69	10.00	12.68	2.68
Standard Deviation	0.43	0.69	0.08	2.84	4.59	0.55
p-value (H0: Difference > 0)			0.00			0.00
Male Subjects (Sample Size)	32			32		
Mean	1.31	2.01	0.70	9.62	13.03	3.41
Standard Deviation	0.88	1.36	0.20	6.22	9.31	1.40
p-value (H0: Difference > 0)			0.00			0.02
Female Subjects (Sample Size)	16			16		
Mean	1.11	1.77	0.66	10.77	11.98	1.22
Standard Deviation	1.04	1.65	0.35	9.62	10.44	2.51
p-value (H0: Difference > 0)			0.08			0.64
Caltech Subjects (Sample Size)	24			24		
Mean	1.48	2.25	0.78	9.92	11.11	1.19
Standard Deviation	1.14	1.74	0.30	7.74	8.35	1.64
p-value (H0: Difference > 0)			0.02			0.48
PCC Subjects (Sample Size)	24			24		
Mean	1.01	1.61	0.59	10.08	14.25	4.17
Standard Deviation	0.81	1.31	0.22	7.82	11.77	2.04
p-value (H0: Difference > 0)			0.01			0.05

if one takes extra step in one's thinking process, one will be slower to respond. A more appropriate validity test of the thinking type will be to examine the situation when there is an improvement in thinking step, the subject takes longer time to respond.

But, does the thinking step of subjects improve through the sequence of 22 games? We divide the sequence of 22 games into first and second half. A comparison between the first half of the games and the second half shows that the increase in thinking step is significant. In addition, this increase in thinking step is accompanied by a corresponding increase in response time, finally suggesting a deeper reasoning process consistent with findings of Rubinstein (2004). The result is reported in Table 7. There are some interesting departures from this aggregate result when examining some subgroups of our subjects. While both male subjects and PCC student subjects exhibit same result as the total subject pool, female subjects do not show any significant increase in thinking steps and the corresponding response time. CalTech student subjects show an improvement in thinking step but the corresponding increase in response time is not significant.

5. ECONOMIC IMPLICATIONS OF LIMITED STRATEGIC THINKING

Models of limited thinking can be applied to several interesting problems in economics, including asset pricing, speculation, competition neglect in business entry, incentive contracts, and macroeconomics.

Asset pricing: As Keynes pointed out (and many commentators since him; e.g., Tirole 1985; Shleifer and Vishny, 1990), if investors in stocks are not sure that others are rational (or will price assets rationally in the future) then asset prices will not necessarily equal fundamental or intrinsic values.[19] A precise model of limited strategic thinking might therefore be used to explain the existence and crashes of price bubbles.

Speculation: The "Groucho Marx theorem" says that traders who are risk-averse should not speculate by trading with each other even if they have private information (since the only person who will trade with you may be better-informed). But this theorem rests on unrealistic assumptions of common knowledge of rationality and is violated constantly by massive speculative trading volume and other kinds of betting, as well as in experiments.[20] Speculation will occur in CH models because 1- and higher-step players think they are sometimes betting against random (0-step) bettors who make mistakes.

Competition neglect and business entry: Players who do limited iterated thinking, or believe others are not as smart as themselves, will neglect competition in business entry, which may help explain why the failure rate of new businesses is so high (see Camerer and Lovallo, 1999; Huberman and Rubinstein, 2000). Simple entry games are studied below. Theory and estimates from experimental data show that the CH model can explain why the amount of entry is monotonic in market capacity, but too many players enter when capacity is low. Managerial hubris, overconfidence, and self-serving biases which are correlated with costly delay and labor strikes in the lab

(Babcock et al., 1995) and in the field (Babcock and Loewenstein. 1997) can also be interpreted as players not believing others always behave rationally.

Incentives: In his review of empirical evidence on incentive contracts in organizations, Prendergast (1999) notes that workers typically react to simple incentives as standard models predict. However, firms usually do not implement complex contracts which *should* elicit higher effort and improve efficiency. This might be explained as the result of firms thinking strategically, but not believing that workers respond rationally.

Macroeconomics: Woodford (2001) notes that in Phelps-Lucas "islands" models, nominal shocks can have real effects, but their predicted persistence is too short compared to actual effects in data. He shows that imperfect information about higher-order nominal GDP estimates – beliefs about beliefs, and higher-order iterations – can cause longer persistence which matches the data, and Svensson (2001) notes that iterated beliefs are probably constrained by computational capacity. In CH models, players' beliefs are not mutually consistent so there is higher-order belief inconsistency which might explain the longer persistence of shocks that Woodford noted.

6. CONCLUSION

Camerer, Ho and Chong (2004) propose a model of iterated limited thinking and suggest that the amount of strategic thinking is sharply constrained by working memory. This is consistent with a simple axiom which implies a Poisson distribution of thinking steps that can be characterized by one parameter τ (the mean number of thinking steps, and the variance). This paper explores the generality and precision of this Poisson CH model.

We find that this Poisson CH model offers good approximation to an unconstrained general distribution CH model where the degradation in fit is minimal. In addition, the Poisson CH model fits better than QRE in most cases which implies that mutual consistency is not a necessary condition to explain behaviors in most experimental games. Put it differently, this seems to suggest that mutual consistency is a more plausible assumption to relax, compared to the best-response assumption, in explaining deviation from equilibrium theory. We also found evidence that cognitive hierarchy may have captured the reasoning mechanism given the high correlation between the number of thinking steps implied by players' choice behavior, and both response time and schools they attend.

There are many challenges in future research. An obvious one is to endogenize the mean number of thinking steps τ, presumably from some kind of cost-benefit analysis in which players weigh the marginal benefits of thinking further against cognitive constraint (cf. Gabaix and Laibson, 2000). It is also likely that a more nuanced model of what 0-step players are doing would improve model fits in some types of games.

The model is easily adapted to incomplete information games because the 0-step players make choices which reach every information set, which eliminates the need to impose delicate refinements to make predictions. Explaining behavior in signaling games and other extensive-form games with incomplete information is therefore workable and a high priority for future work. (Brandts and Holt, 1992, and Banks, Camerer, and Porter, 1994, suggest that mixtures of decision rules in the first period, and learning in subsequent periods, can explain the path of equilibration in signaling games; the CH approach may add some bite to these ideas.)

Another important challenge is repeated games. The CH approach will generally underestimate the amount of strategic foresight observed in these games (e.g., players using more than one step of thinking will choose supergame strategies which always defect in repeated prisoners' dilemmas). An important step is to draw a sensible parametric analogy between steps of strategic foresight and steps of iterated thinking is necessary to explain observed behavior in such games (cf. Camerer, Ho and Chong, 2002a,b).

ACKNOWLEDGMENT

This research was supported by NSF grant SES-0078911. Special thanks to C. Monica Capra, Haitao Cui, Paul Glimcher, Roger Myerson, Matthew Rabin, Amnon Rapoport and Rami Zwick for helpful comments and suggestions. Ye Li, Lisa Wang, Ming Hsu and Brian Rogers provided excellent research assistance. Useful comments were also received from seminar participants at Caltech, Chicago, New York University, Pittsburgh, the Nobel Symposium in Sweden (December 2001), Columbia, Berkeley and Experimental Business Research Conference at HKUST (December 2003).

NOTES

1 Equal randomization implies that all strategies are chosen with positive probability. This is helpful for empirical work because it means all strategies will have positive predicted probabilities, so there is no zero-likelihood problem when using maximum likelihood estimation. This also liberates us to assume best response by players using more steps of thinking (rather than stochastic response). For theoretical work, having all strategies chosen with positive probability solves two familiar problems–eliminating incredible threats (since all threats are "tested") as subgame perfection does; and eliminating *ad hoc* rules for Bayesian updating after zero probability events (since there are no such events).

However, equal randomization may be replaced with some heuristics which subjects might used to reduce the set of strategies they would consider. This is especially plausible when the number of strategies is large or when some strategies are obviously inferior even by a casual glance from a untrained subject.

2 In a n-player game, the step-0 probability is a $n - 1$ multinomial expression.

3 Use the sum of the absolute deviations to measure the distance of the normalized distributions from the true distribution. The total absolute deviation is

$$D_{k-1}(k) = \sum_{h=0}^{k-1} \left| f(h) \Big/ \sum_{l=0}^{l=k-1} f(l) - f(h) \right| + \sum_{h=k}^{\infty} |0 - f(h)|$$

Algebra shows that this is $D(k) = 2 \cdot (1 - \sum_{h=0}^{k-1} f(h))$. $D(k)$ is decreasing in k – so beliefs get closer and closer to the truth – and $\lim_{k\to\infty} D(k) = 0$ because $\sum_{h=0}^{\infty} f(h) = 1$.

4 The $k - 1$ assumption is easy to work with theoretically because the sequence of predicted choices can be computed by working up the hierarchy without using any information about the true distribution $f(k)$.

[5] The total absolute deviation for the $k-1$ model is

$$D_{k-1}(k) = \sum_{h=0}^{k-2} f(h) + 1 - f(k-1) + \sum_{h=k}^{\infty} f(h) = 2 \cdot (1 - f(k-1)).$$

This figure falls as k approaches the distribution mode (τ) then rises again, which means the beliefs of the highest step thinkers (beyond τ) are *furthest* from the truth.

[6] Recall that in our approach, the sum of absolute belief deviations is $D(k) = 2\ (1 - \Sigma_{h=0}^{k-1} f(h))$. This is smaller than $D_{k-1}(k)$ for any $k > 0$ because $2 \bullet (1 - \Sigma_{h=0}^{k-1} f(h)) < 2\ (1 - f(k-1))$.

[7] In the entry games, as you increase k the $k-1$ model decision rules alternate back and forth between entering at low c (i.e., c less than half the number of entrants) and staying out at high c, and the opposite pattern. Aggregating these decision rules will produce a step function in which the rate of entry is constant for $c < .5N$ then switches to a higher rate, which does not look at all like the monotonicity in the data.

[8] For column players, the 0- and 1-step thinkers randomize equally over L and R, 2- and 3-step thinkers choose R, 4- and 5-step thinkers choose L, and so forth in a two-step cycle. For row players, 0-step thinkers randomize equally over T and B, 1- and 2-step thinkers choose T, 3- and 4-step thinkers choose B, and so forth in a two-step cycle. These best-response cycles do not depend on $f(k)$ or on x.

[9] The log-likelihood values for the CH models with and without self-awareness are –1265 vs – 1115 for Stahl and Wilson (1995), –1802 vs –1740 for Cooper and Van Huyck (2003) and –570 vs –555 for Costa-Gomes et al. (2001).

[10] The frequencies $f(k)$ are constrained in a small way to improve identification, which is essentially harmless in terms of fit. The constraint imposed is that the $f(k)$ function should be inverted-U shaped in k. That is, if $f(k) < f(k-1)$ for a particular k – that is, the distribution function turns downward – then $f(k+x) < f(k)$ for any positive integer x – i.e., once the $f(k)$ distribution turns downward it cannot rise up again. This constraint is necessary because when the estimates are not constrained in this way, it is possible to have, say, a large fraction of 0, 1, and 2 subjects, but no 3-step subjects. But 4-step subjects who have (normalized) beliefs about this distribution will simply ignore the 3-step types. As a result, they will choose the same strategy a 3-step type would choose. So the unconstrained estimation can place zero $f(k)$ values anywhere in the distribution and produce precisely the same pattern of best responses (and hence, fit) as an alternative specification in which the zero is removed. In econometric language, there is a severe identification problem. One way to eliminate the possibility of these unidentified insertions of zero $f(k)$ types is to force the distribution to *not* wake up again after a zero $f(k)$ and produce positive values of $f(k+1)$. Happily, imposing this no-inverted-U constraint degrades LL very little. Across the four data sets, the reduction in LL is only 40, 0, 1, 14, and 0 points so the constraint is essentially harmless.

[11] This game shares certain properties with the market entry games studied by Zwick and Rapoport (2002).

[12] This structure suppresses the effect of overconfidence actual business entrants might have in a game in which more skilled entrants earn more (e.g., Camerer and Lovallo, 1999).

[13] This is because the logit has the *Independence from Irrelevant Alternatives* property. One could presumably develop a hierarchical equivalent that does not exhibit this property.

[14] Camerer, Ho and Chong (2003) estimate Poisson CH model that reproduces this heterogeneity in data well for 24 p-beauty contest games, where the median τ is 1.5. The 24 p-beauty contest games were taken from previously published results (Nagel, 1995; 1999; Ho, Camerer and Weigelt, 1998), from 2- and 3-player games conducted in the four 12-subject sessions with Caltech and PCC students, from unpublished data collected by Ho, Camerer and Weigelt, and from convenience samples collected by author Camerer with various audiences playing for $20. Montier (2004) also reports similar spike mimicking ability by CH model in his contest with a subject pool of 1000! The estimated τ is 2.8.

[15] See Camerer, Loewenstein, and Prelec, 2002. The use of brain imaging will sound farfetched to most economists. But Glimcher, 2002, chapter 13, reports the existence of "equilibrating" neurons in monkeys which fire in rough proportion to expected payoffs of strategies, as the monkeys play a 2×2 "work-or-shirk" game with a mixed equilibrium against a computerized opponent. When play is out of

equilibrium, neurons fire more actively when the monkey plays the strategy with the higher expected payoff. This activation guides the monkey to play the "better" strategy more often, which eventually produces equilibration. After equilibration, when both strategies have equal expected payoffs, the neurons fire at the same rate after each of the two strategies are played, so that the brain is "recognizing" equilibration. Since the human brain is essentially the monkey brain with some extra cortex, which is used largely for planning and understanding social structure and language, it is likely that humans have a similar neural circuitry which encodes expected payoffs and guides equilibration in simple games.

[16] Schotter et al. (1994) found a similar effect of display and timing in games with Nash equilibria which are not subgame-perfect. In the simultaneous matrix form more players chose the Nash equilibrium, as if they did not reason through what others would do. Note that these display effects can be interpreted as focusing players' attention in different ways, altering the number of thinking steps they are doing or what players think at different steps. We also observed a belief-prompting effect in beauty contest games (unpublished). When players simply made choices, 25% chose numbers above 50 in the first period. When forced to guess what the average choice would be, this figure fell to 15%. The samples were small so the effect is not significant but it goes in the same direction as the effects above.

[17] Results of the revealed types t_{ig} for the 48 subjects in the 22 mixed games are available from the authors upon requests.

[18] When level 0 players are excluded in the regression, the negative correlation becomes insignificant. The result of this regression is also reported in Table 6.

[19] Besides historical examples like Dutch tulip bulbs and the $5 trillion tech-stock bubble in the 1990s, experiments have shown such bubbles even in environments in which the asset's fundamental value is controlled and commonly-known. See Smith, Suchanek and Williams, 1988; Camerer and Weigelt, 1993; and Lei, Noussair and Plott, 2001.

[20] See Sonsino, Erev and Gilat, 2000; Sovik, 2000.

REFERENCE

Amaldoss, Wilfred and Amnon Rapoport (2004). "Effects of Idiosyncratic Investments in Collaborative Networks: An Experimental Analysis," In Zwick and Rapoport (Eds.), Experimental Business Research: Volume 3, pp. 81–112.

Babcock, Linda, George Loewenstein, Samuel Issacharoff, and Colin F. Camerer (1995). "Biased Judgments of Fairness in Bargaining." *American Economic Review*, 85, 1337–43.

Babcock, Linda, and George Loewenstein (1997). "Explaining bargaining impasses: The role of self-serving biases." *Journal of Economic Perspectives*, 11, 109–26.

Banks, Jeffrey, Colin F. Camerer, and David Porter (1994). "Experimental Tests of Nash Refinements in Signaling Games." *Games and Economic Behavior*, 6, 1–31.

Benabou, Roland and Jean Tirole (2002). "Self-Confidence and Personal Motivation." *Quarterly Journal of Economics*, 117, 871–915.

Binmore, Kenneth, Joe Swierzbinski, and Chris Proulx (2001). "Does maximin work? An experimental study." *Economic Journal*, 111, 445–464.

Bloomfield, Robert (1994). "Learning a mixed strategy equilibrium in the laboratory." *Journal of Economic Behavior and Organization*, 25, 411–436.

Brandts, Jordi and Charles Holt (1992). "An Experimental Test of Equilibrium Dominance in Signaling Games." *American Economic Review*, 82, 1350–65.

Camerer, Colin F. (2003). *Behavioral Game Theory: Experiments on Strategic Interaction*, Princeton: Princeton University Press.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong (2002a). "Strategic Teaching and Equilibrium Models of Repeated Trust and Entry Games," CalTech Working Paper, http://www.hss.caltech.edu/~camerer/camerer.html

Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong (2002b). "Sophisticated EWA Learning and Strategic Teaching in Repeated Games." *Journal of Economic Theory*, 104, 137–188.

Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong (2003). "Models of Thinking, Learning and Teaching Games." *American Economic Review*, 93(2) (May), 192–195.

Camerer, Colin F., Teck-Hua Ho and Juin-Kuan Chong (2004). "A Cognitive Hierarchy Theory of Games." *Quarterly Journal of Economics*, 119(3), 861–898.

Camerer, Colin F., Eric J. Johnson, Talia Rymon, and Sankar Sen. (1994). "Cognition and Framing in Sequential Bargaining for Gains and Losses," in K. Binmore, A. Kirman, and P. Tani (Eds.), *Frontiers of Game Theory*, Cambridge: MIT Press, 27–47.

Camerer, Colin F., George Loewenstein, and Drazen Prelec (2002). "Gray Matters: How Neuroscience can Inform Economics" under revision for *Journal of Economic Perspectives*.

Camerer, Colin F., and Daniel Lovallo (1999). "Overconfidence and Excess Entry: An Experimental Approach." *American Economic Review*, 89, 306–18.

Camerer, Colin F. and Keith Weigelt (1993). "Convergence in Experimental Double Auctions for Stochastically Lived Assets," in D. Friedman and J. Rust (Eds.), *The Double Auction Market: Theories, Institutions and Experimental Evaluations*, Redwood City, CA: Addison–Wesley, 355–396.

Capra, C. Mònica (1999). *Noisy Expectation Formation in One-shot Games*, Ph.D. thesis.

Chen, Hsiao-Chi, James W. Friedman, and Jacques-Francois Thisse (1996). "Boundedly Rational Nash Equilibrium: A Probabilistic Choice Approach." *Games and Economic Behavior*, 18, 1832–54.

Cooper, David and J. Van Huyck (2003). "Evidence on the Equivalence of the Strategic and Extensive Form Representation of Games." *Journal of Economic Theory*, 110(2), 290–308.

Costa-Gomes, M. and V. Crawford and B. Broseta (2001). "Cognition and Behavior in Normal-form Games: An Experimental Study." *Econometrica*, 69, 1193–1235.

Croson, Rachel T. A. (2000). "Thinking Like A Game Theorist: Factors Affecting the Frequency of Equilibrium Play." *Journal Of Economic Behavior And Organization*, 41(3), 299–314.

Dalkey, Norman (1953). "Equivalence of information patterns and essentially determinate games" Contributions to the Theory of Games II, *Annals of Mathematical Studies*, 28, Princeton, NJ: Princeton University Press.

Friedman, Daniel and Bernado Huberman (2004). "Internet Congestion Experiments," Proceedings of Second Asian Conference on Experimental Research, Hong Kong.

Gabaix, Xavier and David Laibson (2000). "Bounded rationality and directed cognition," Harvard University working paper.

Glimcher, Paul (2002). *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*, MIT Press.

Goeree, Jacob K. and Charles A. Holt (1999). "Stochastic Game Theory: For Playing Games, Not Just for Doing Theory." *Proceedings of the National Academy of Sciences*, 96, 10564–10567.

Goeree, Jacob K. and Charles A. Holt (2004). "A Theory of Noisy Introspection." *Games and Economic Behavior*, 46(2), 365–382.

Goeree, Jacob, Holt, Charles A. and Thomas R. Palfrey (2003). "Risk Averse Behavior in Generalized Matching Pennies Games." *Games and Economic Behavior*, 45(1), 97–113.

Guyer, M. V. and Rapoport, A. (1972). " 2 × 2 Games Played Once." *Journal of Conflict Resolution*, 16(3), 409–431.

Ho, Teck-Hua, Colin Camerer, and Keith Weigelt (1998). "Iterated Dominance and Iterated Best-response in *p*-Beauty Contests." *American Economic Review*, 88, 947–969.

Ho, Teck-Hua and Keith Weigelt (1996). "Task Complexity, Equilibrium Selection, and Learning: An Experimental Study." *Management Science*, 42, 659–679.

Hoffman, Elizabeth, Kevin McCabe, and Vernon L. Smith (2000). "The impact of exchange context on the activation of equity in ultimatum games." *Experimental Economics*, 3, 5–9.

Huberman, Gur and Ariel Rubinstein (2000). "Correct Belief, Wrong Action and a Puzzling Gender Difference," unpublished manuscript.

Johnson, E., C. Camerer, S. Sen and T. Rymon (2002). "Detecting Backward Induction in Sequential Bargaining Experiments." *Journal of Economic Theory*, 104, 16–47.

Kim, Chung K. and William S. Waller (2004). "A Behavioral Accounting Study of Strategic Interaction in a Tax Compliance Game," In Zwick and Rapoport (Eds.), Experimental Business Research: Volume 3, pp. 35–56.

Lei, Vivian, Charles Noussair and Charles Plott (2001). "Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality." *Econometrica*, 69(4), 831–859.

Lieberman, Bernhardt (1962). "Experimental studies of conflict in some two-person and three-person games," in J. H. Criswell, H. Solomon, and P. Suppes (Eds.), *Mathematical Models in Small Group Processes*, Stanford: Stanford University Press, 203–220.

McKelvey, Richard D. and Thomas R. Palfrey (1995). "Quantal Response Equilibria for Normal-form Games." *Games and Economic Behavior*, 7, 6–38.

McKelvey, Richard D. and Thomas R. Palfrey (1998). "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics*, 1, 9–41.

McKelvey, Richard, Thomas Palfrey and Roberto Weber (2000). "The Effects of Payoff Magnitude and Heterogeneity on Behavior in 2 × 2 Games with Unique Mixed Strategy Equilibria." *Journal of Economic Behavior and Organization*, 42(4), 523–548.

Messick, David M. (1967). "Interdependent Decision Strategies in Zero-sum Games: A Computer-controlled Study." *Behavioral Science*, 12, 33–48.

Montier, James (2004). "Global Equity Strategy." *Dresdner Kleinwort Wasserstein Research*, 17 February 2004.

Mookherjee, Dilip and Barry Sopher (1997). "Learning and Decision Costs in Experimental Constant-sum Games." *Games and Economic Behavior*, 19, 97–132.

Nagel, R. (1995). "Unraveling in Guessing Games: An Experimental Study." *American Economic Review*, 85, 1313–1326.

Nagel, R. (1999). "A Survey of Research on Experimental Beauty-contest Games," in D. Budescu, I. Erev, and R. Zwick (Eds.), *Games and Human Behavior: Essays in Honor of Amnon Rapoport*, Mahwah, NJ: Lawrence Erlbaum.

Ochs, Jack (1995). "Games with Unique, Mixed Strategy Equilibria: An Experimental Study." *Games and Economic Behavior*, 10, 202–217.

O'Neill, Barry (1987). "Nonmetric Test of the Minimax Theory of Two-person Zero-sum Games." *Proceedings of the National Academy of Sciences*, 84, 2106–2109.

Prendergast, Canice (1999). "The Provision of Incentives in Firms." *Journal of Economic Literature*, 37, 7–63.

Rapoport, Amnon and Wilfred Amaldoss (2000). "Mixed Strategies and Iterative Elimination of Strongly Dominated Strategies: An Experimental Investigation of States of Knowledge." *Journal of Economic Behavior and Organization*, 42, 483–521.

Rapoport, Amnon and Richard B. Boebel (1992). "Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis." *Games and Economic Behavior*, 4, 261–283.

Roll, Richard (1984). "Orange Juice and Weather." *American Economic Review*, 74, 861–880.

Rosenthal, Robert W. (1989). "A Bounded Rationality Approach to the Study of Noncooperative Games." *International Journal of Game Theory*, 18(3), 273–92.

Rubinstein, Ariel (2004). "Instinctive and Cognitive Reasoning: Response Times Study." unpublished manuscript.

Schotter, A., K. Weigelt and C. Wilson (1994). "A Laboratory Investigation of Multiperson Rationality and Presentation Effects." *Games and Economic Behavior*, 6, 445–468.

Selten, Reinhard (1998). "Features of Experimentally Observed Bounded Rationality." *European Economic Review*, 42, 413–436.

Shleifer, Andrei and Robert W. Vishny (1990). "Equilibrium Short Horizons of Investors and Firms." *American Economic Review Papers and Proceedings*, 80 (May), 148–153.

Smith, Vernon L., Gerry Suchanek and A. Williams (1988) "Bubbles, Crashes and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica*, 56, 1119–1151.

Sonsino, D., Ido Erev and S. Gilat (2000). "On the Likelihood of Repeated Zero-sum Betting by Adaptive (Human) Agents." Technion, Israel Institute of Technology.

Sovik, Ylva (2000). "Impossible Bets: An Experimental Study," University of Oslo Department of Economics.

Stahl, D. O. and P. Wilson (1995). "On Players Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10, 213–54.

Svensson, L. E. O. (2001). "Comments on Michael Woodford paper," presented at Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps. Columbia University, October 5–6, 2001.

Tang, Fang-Fang (2001). "Anticipatory learning in two-person games: Some experimental results." *Journal of Economic Behavior and Organization*, 44, 221–32.

Tirole, J. (1985). "Asset Bubbles and Overlapping Generations." *Econometrica*, 53, 1071–1100 (reprinted 1499–1528).

Warglien, M., M. G. Devetag and P. Legrenzi (1998). "Mental Models and Naïve Play in Normal Form Games," Universita Ca' Foscari di Venezia working paper.

Woodford, M. (2001). "Imperfect common knowledge and the effects of monetary policy," presented at Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps, Columbia University, October 5–6, 2001.

Zwick, Rami and Amnon Rapoport (2002). "Tacit Coordination in a Decentralized Market Entry Game with Fixed Capacity." *Experimental Economics*, 5, 253–272.

APPENDIX: MODEL RECOVERY, MIXED GAMES, AND BOOTSTRAPPED STANDARD ERRORS

The 22 mixed games were taken from the review in Camerer (2003, chapter 3). They are (in order of presentation to the subjects): Ochs (1995), (matching pennies plus games 1–3); Bloomfield (1994); Binmore et al. (2001) Game 4; Rapoport and Almadoss (2000); Binmore et al. (2001), games 1–3; Tang (2001), games 1–3; Goeree, Holt, and Palfrey (2003), games 2–3; Mookherjee and Sopher (1997), games 1–2; Rapoport and Boebel (1992); Messick (1965); Lieberman (1962); O'Neill (1987); Goeree, Holt, and Palfrey (2003), game 1. Four games were perturbed from the original payoffs: The row upper left payoff in Ochs's original game 1 was changed to 2; the Rapoport and Almadoss (2000) game was computed for $r = 15$; the middle row payoff in Binmore et al. (2001) game 2 was 30 rather than –30; and the lower left row payoff in Goeree, Holt and Palfrey's (2003) game 3 was 16 rather than 37. Original payoffs in games were multiplied by the following conversion factors: 10, 10, 10, 10, 0.5, 10, 5, 10, 10, 10, 1, 1, 1, 0.25, 0.1, 30, 30, 30, 5, 3, 10, 0.25. Currency units were then equal to $10.

Table 8 below shows estimates of τ recovered from simulated data, created using the CH model, to see how well the estimation procedure recovers τ when the actual value is known. Each line shows a different value of "true" τ, for different sizes of simulated samples (n, either 20, 48, or 100), across a 2×2, 4×4, and 6×6 game, and then averaged over the three games. There is little bias in recovering the actual τ values (except for a slight upward bias when τ is small), although the 95% confidence intervals are rather wide when samples are of size 20, and for the 2×2 game. The key lesson is that small samples do not have much power, and 2×2 games are not very useful for estimating CH models. The problem is that each level of thinking, above 0, picks a distinct strategy; so when there are only two strategies several different levels all pick the same strategy, which means it is hard to identify how many levels are being used.

Table 8. Estimates of t from model recovery simulations

		2 × 2		4 × 4		6 × 6		average across games	
True τ	*n*	*mean*	*90% CI*	*mean*	*90% CI*	*mean*	*90% CI*	*mean*	*90% CI*
0.5	20	0.62	(0.05, 2.00)	0.53	(0.20, 0.90)	0.51	(0.30, 0.90)	0.56	(0.19, 1.23)
	48	0.52	(0.15, 0.95)	0.51	(0.30, 0.75)	0.50	(0.35, 0.65)	0.51	(0.28, 0.77)
	100	0.52	(0.25, 0.80)	0.51	(0.35, 0.70)	0.50	(0.40, 0.60)	0.51	(0.34, 0.68)
1.0	20	1.08	(0.50, 2.00)	1.02	(0.55, 1.55)	0.97	(0.60, 1.45)	1.04	(0.57, 1.60)
	48	1.04	(0.60, 1.60)	1.01	(0.70, 1.35)	0.96	(0.70, 1.20)	1.01	(0.69, 1.37)
	100	1.01	(0.70, 1.30)	1.00	(0.80, 1.20)	0.96	(0.75, 1.05)	1.00	(0.78, 1.21)
1.5	20	1.55	(0.95, 2.00)	1.52	(1.00, 2.15)	1.45	(1.05, 1.55)	1.48	(0.96, 1.83)
	48	1.55	(1.10, 2.00)	1.51	(1.15, 1.85)	1.48	(1.40, 1.50)	1.50	(1.22, 1.75)
	100	1.54	(1.20, 1.95)	1.50	(1.25, 1.75)	1.50	(1.50, 1.50)	1.50	(1.30, 1.70)
2.0	20	1.96	(0.85, 2.85)	2.05	(1.45, 2.70)	2.06	(1.55, 2.65)	2.01	(1.31, 2.68)
	48	2.02	(1.25, 2.50)	2.02	(1.65, 2.50)	2.01	(1.65, 2.40)	2.01	(1.59, 2.43)
	100	2.05	(1.80, 2.35)	2.01	(1.75, 2.30)	2.01	(1.75, 2.30)	2.01	(1.77, 2.30)
2.5	20	2.39	(1.20, 3.00)	2.45	(1.90, 2.90)	2.52	(1.90, 3.00)	2.46	(1.74, 2.98)
	48	2.49	(2.00, 3.00)	2.47	(2.05, 2.70)	2.53	(2.05, 3.00)	2.49	(2.06, 2.91)
	100	2.51	(2.15, 2.90)	2.49	(2.20, 2.65)	2.51	(2.20, 2.85)	2.50	(2.21, 2.80)

Chapter 10

PARTITION DEPENDENCE IN DECISION ANALYSIS, RESOURCE ALLOCATION, AND CONSUMER CHOICE

Craig R. Fox
University of California at Los Angeles

David Bardolet
University of California at Los Angeles

Daniel Lieb
Duke University

Abstract

In this chapter we explore a wide range of judgment and decision tasks in which people are called on to allocate a scarce resource (e.g., money, choices, belief) over a fixed set of possibilities (e.g., investment opportunities, consumption options, events). We observe that in these situations people tend to invoke *maximum entropy heuristics* in which they are biased toward even allocation. Moreover, we argue that before applying these heuristics, decision makers subjectively partition the set of options into groups over which they apply even allocation. As a result, allocations vary systematically with the particular partition that people happen to invoke, a phenomenon called *partition dependence*. We review evidence for maximum entropy heuristics and partition dependence in the following domains: (1) decision analysis in which degree of belief and importance weights must be distributed among possible events and attributes, respectively; (2) managerial decision making in which money and other organizational resources are allocated among risky projects, divisions, and organizational stakeholders; and (3) consumer choice in which individuals make selections among various consumption goods and consumption time periods.

1. INTRODUCTION

Imagine that three graduate students with identical tastes enter a corner grocery store, each planning to purchase 12 frozen dinners to consume over the next couple

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 229–251.
© 2005 *Springer. Printed in the Netherlands.*

of weeks. This store carries three varieties of frozen dinner: chicken parmesan, beef ravioli, and Szechwan beef. The first student decides to spread out her consumption over different flavors by choosing four meals of each variety. The second student notices that the first two dishes are Italian cuisine and the third is Chinese. Hence this student purchases six Italian meals (three chicken parmesan, three beef ravioli) and six Chinese meals (Szechwan beef). The third student notes that the first dish is chicken whereas the latter two dishes are beef. This student purchases six chicken dishes (chicken parmesan) and six beef dishes (three beef ravioli, three Szechwan beef). In this example all three graduate students apply a "maximum entropy heuristic," spreading out their choices evenly over different kinds of meals. In this case the heuristic reflects a desire to seek variety. However, the implication of this strategy depends crucially on how each student subjectively groups the set of options (by meal type, by cuisine, or by meat). Thus, these hypothetical graduate students exhibit preferences that are *partition-dependent*, varying with their subjective partition of the option set.

In this paper we review a number of contexts in which people must allocate a scarce resource (e.g., money, choices, belief) over a fixed set of possibilities (e.g., investment opportunities, consumption options, events). We argue that people act as if they rely on a combination of a maximum entropy heuristic and allocation according to their innate preferences and beliefs. To the extent that they rely at all on a maximum entropy heuristic their allocations will be partition-dependent, varying systematically with the way in which they subjectively partition the set of possibilities. We use the term "maximum entropy" to refer to an even allocation over all possibilities that have been identified (i.e., reflecting maximum disorder or minimum information). We use the term "heuristic" to refer to cognitive strategies that people use to simplify otherwise complex judgments (note that this differs from the definition of heuristics as "attribute substitutions" as in Kahneman & Fredrick, 2002).

To provide evidence for this account we review a number of disparate allocation domains that are of interest to management scholars: (1) decision analysis in which degree of belief and importance weights must be distributed among possible events and attributes, respectively; (2) managerial decision making in which money and other organizational resources are allocated among risky projects, divisions, and organizational stakeholders; and (3) consumer choice in which individuals make selections among various consumption goods and consumption time periods. For each domain we first characterize the particular maximum entropy heuristic and its possible psychological sources. Next, we review studies that have documented partition dependence by experimentally manipulating the relative accessibility of alternative partitions. Finally, we review any moderators of partition dependence that have been identified. We conclude with a general discussion in which we summarize these results, address the possibility of demand effects in the studies cited, distinguish partition dependence from related phenomena, discuss other dependencies implied by the present account, and identify priorities for future research.

2. REVIEW OF EMPIRICAL WORK

2.1. Partition dependence in decision analysis

Decision analysis is used in a wide range of industries to address a broad array of important organizational decisions such as financial forecasting and budget allocation, R&D project selection, bidding and pricing, and more general corporate strategic planning (Keefer, Kirkwood & Corner, 2004). The use of such techniques in organizations is influenced by such factors as managerial attitudes, organizational structure, and industry characteristics (See and Clemen, 2003).

Tools of decision analysis typically allow considerable discretion in how to structure the assessment task. The analyst must define the problem space including: relevant options under consideration (decision nodes), key uncertainties (chance nodes), and relevant attributes of potential consequences. Once a decision problem is explicitly defined, the analyst, sometimes with the assistance of an expert, elicits a utility function as well as probabilities of all relevant events and weights of all relevant attributes. Whereas utility functions are typically derived from a series of choices (e.g., Wakker & Deneffe, 1996; Abdellaoui, 2000), probabilities and attribute weights are typically evaluated directly through an explicit allocation among possible events or attributes. We assert that these assessments are distorted by the particular partition of the event or attribute space that is identified by the decision analyst.

2.1.1. Judged probability

Most problems that are submitted to decision analysis involve consequences that are uncertain. These problems therefore require an assessment of the subjective probabilities of events, such as a failure of important equipment or a rise in the price of electricity. Practical procedures for assessing subjective probabilities were characterized by Spetzler and Stael Von Holstein (1975) and similar procedures are still in use today (Clemen, 1996). Typical protocols for probability elicitation are described in Keeney & von Widterfeldt (1991) and Morgan & Henrion (1990). Although these authors provide guidance concerning a number of important steps of the assessment procedure (e.g., identifying and selecting experts, training them in probability elicitation, the probability assessment itself), they have little to say concerning the process of choosing which events should be assessed.

Fischhoff, Slovic, and Lichtenstein (1978) found that when novices and experts judged probabilities of various reasons why a system might fail (e.g., reasons why a car would fail to start) their responses were affected by which causes were explicitly identified (e.g., "dead battery," "ignition system") and which were pruned from the tree and relegated to a residual catch-all category ("all other causes"). In particular, they found that when branches were pruned from the tree, the judged probability of the now more inclusive catch-all category (as assessed by a new group of participants) increased by less than the probabilities of the pruned branches; this probability was instead distributed among the other remaining branches of the tree.

These investigators and most that followed have attributed pruning bias to the availability heuristic: people can more easily recall or imagine instances of a particular category when that category is described in greater detail – thus the causes that are out of site are out of mind. Fox and Clemen (2004) observed that the availability account of pruning bias cannot explain more general instances of partition dependence. For instance, they showed that unpacking an event (e.g., the next top-ranked business school according to *Business Week* will be a school other than Wharton) into a disjunction of obvious constituents and a catch-all (Chicago or Harvard or Kellogg or Stanford or another school other than Wharton) does not lead to higher judged probabilities, whereas assigning these constituent events to separate branches that are assessed individually does lead to a significant increase in aggregate judged probabilities.

Fox and Clemen (2004) argued that the pruning bias is an instance of a more general pattern of partition dependence in probability assessment that extends beyond the domain of fault trees and discrete categories of events. They argue further that pruning bias can be attributed in part to a tendency to first allocate probabilities evenly over all events under consideration, and then adjust in response to an evaluation of how those events differ. The tendency to anchor on an "ignorance prior" probability of $1/n$ for each of n branches may reflect an intuitive application of the fallacious "principle of insufficient reason" (e.g., Laplace, 1776; cited in Hacking, 1975, p. 132) according to which events that cannot be distinguished are treated as equally likely. Because adjustment is typically insufficient (Tversky & Kahneman, 1974) probabilities are biased toward even allocation and assessments are therefore partition-dependent.

Evidence for this "ignorance prior" model and partition dependence has been steadily mounting in studies of judgment under uncertainty. Fox and Rottenstreich (2003) found that the language of the probability query can facilitate either a two-fold "case" partition {the target event occurs; the target event fails to occur} or an n-fold "class" partition of similar events {event 1 occurs; event 2 occurs; . . . ; event n occurs} and corresponding biases toward 1/2 or $1/n$. For instance, when asked to judge the probability that "the temperature on Sunday will be higher than any other day next week" the median response of University of Chicago undergraduates was .30 (as if they had anchored their judgments on 1/2 and adjusted somewhat), but when asked to judge the probability that "next week, the highest temperature of the week will occur on Sunday" the median response was .15 (as if they had anchored their judgments on 1/7 and found little need to adjust). Moreover, responses of 1/2 were more common under the former wording whereas responses of 1/7 were more common under the latter wording. Taking a different approach, See, Fox, and Rottenstreich (2004) presented participants with objects flashing on a computer screen that could take on one of four shapes (triangle, circle, square, diamond) and one of two colors (gray, black). Subsequent estimates of relative frequency were biased toward 1/4 for shapes and 1/2 for colors. Finally, evidence of partition dependence in judgment of conditional probabilities is provided by Fox & Levav (in press).

The interpretation of pruning bias in terms of anchoring on the ignorance prior suggests that this phenomenon will be observed in a wide array of situations, including those that involve assessing probabilities for dimensional spaces. For instance, in one study Fox and Clemen (2004) asked members of the Decision Analysis Society (a group of academics and practitioners who study decision analysis) to assess probabilities that the total number of members in the society five years in the future would fall into various ranges. Half assessed the probabilities for the ranges {400 or less, 401–600, 601–800, 801–1,000, more than 1,000} whereas half assessed the probabilities for the ranges {1,000 or less, 1,001–1,200, 1,201–1,400, 1,401–1,600, more than 1,600}. The median judged probability that the membership would total more than 1,000 was 12% for the first group, in which this was one category out of five (ignorance prior = 20%), but it was 27% for the second group, in which this was the sum of four categories out of five (ignorance prior = 80%).

Although the foregoing example suggests that even experts are not immune to partition dependence in judging probabilities, the ignorance prior model suggests that this bias will be less pronounced when people have more knowledge or information that they can use to distinguish among events. This prediction seems to be supported by experimental data. Fox and Clemen (2004) asked MBA students in a decision models class (in which they had received prior training in probability theory and the use of decision trees) to judge the likelihood that the Jakarta Stock Index (JSX) and the NASDAQ index would close in various ranges at the end of the current calendar year. Participants evaluated the probabilities of the ranges {below 1,000, 1,000–2,000, 2,001–4,000, above 4,000} for one index and the ranges {4,000 or below, 4,001–8,000, 8,001–16,000, above 16,000} for the other index. Each participant assigned himself to an experimental condition based on the last digit of his local telephone number: if the number was odd (even), the participant was asked to write "NASDAQ" over the first (second) tree and "JSX" over the second (first) tree. Results demonstrated both pronounced partition dependence and a pronounced knowledge effect: for the unfamiliar JSX (median knowledge rating was 0 on a 0–10 scale), the median respondent reported probabilities that coincided precisely with the ignorance prior, judging the probability that the index would close at 4,000 or below to be 25% if this was a single branch (i.e., respondents with an odd telephone number) but 75% if it was the sum of three branches (respondents with an even telephone number). For the familiar NASDAQ (median knowledge rating was 7 on a 0–10 scale), the median respondent judged the probability that the index would close at 4,000 or below to be 25% if this was a single branch but 50% if this was the sum of three branches – a striking inconsistency, but significantly less pronounced than the corresponding effect for JSX. Similar knowledge effects were documented by See, Fox & Rottenstreich (2004) using their learning paradigm: when participants were given an enhanced opportunity to learn the frequency of objects (a training period that featured more repetitions of objects that were presented more slowly) partition dependence diminished significantly, though it did not disappear entirely.

2.1.2. Attribute weighting

A second discretionary judgment in decision analysis is the allocation of weights to attributes. For example, a consulting firm choosing among different potential sites for a new branch office might consider the cost of living, weather, health care, transportation, and so forth in each of the cities it is considering. Attribute weighting entails two main tasks. First, the analyst must identify the relevant objectives (attributes) to be considered (see Keeney and Raiffa, 1976; Keeney, Renn and von Winterfeldt, 1987). Usually the analyst has flexibility concerning the level of specificity or detail to include in the assessment. For instance, our consulting firm might consider cost of living as a single attribute or might break this attribute down into separate ratings of housing costs, utility costs, food costs, etc. Second, one must attach a numerical weight to each of the attributes that have been identified (for a review of techniques see von Winterfeldt and Edwards, 1986). Although both tasks (attribute identification and attribute weighting) have been separately addressed in the literature, there has been only a smattering of attention to the interaction between them. For example, some studies have found that varying the structure and formulation of attributes can have a substantial impact on the measured multiattribute utility function (see, e.g., Fischer, Damodaran, Laskey and Lincoln, 1987). However, none of these studies have examined factors influencing how people allocate weights across attributes.

Weber (1983) explored the influence of attribute selection on attribute weighting. He asked participants to place weights on various attributes to be used in deciding which automobile to purchase. One group of participants was presented with a set of four attributes whereas a second group was presented with a set of five attributes in which one of the four initial attributes (cost per mile) was split into two more specific attributes (depreciation per mile and yearly operating costs). Results showed that the sum of weights on the more specific attributes was significantly higher than the weight of the original attribute from which the specific attributes were derived. The weights attached to the other common attributes (i.e., those that were not split in either condition) were identical.

Weber, Eisenführ and von Winterfeldt (1988) replicated and extended this finding. These researchers asked business students to evaluate hypothetical future jobs. Participants were provided with a value tree that included three basic objectives (job security, income and career opportunities) that could be split into two more specific attributes. This setting allowed the experimenters to manipulate the level of detail of the value tree by presenting eight groups of respondents with eight different combinations of general and specific attributes. Results of this study document a strong bias in which splitting attributes leads to an increase in aggregate weight. In one group, for example, respondents assigned a weight of .30 to the attribute "job security," but when this attribute was split a second group assigned weights of .20 and .24, respectively, to the attributes "stability of the firm" and "personal job security." Weber et al. (1988) demonstrated the robustness of this "overweighting bias" using a variety of weight elicitation methods (Edwards, 1977; Green and Srinivasan, 1978; von Winterfeldt and Edwards, 1986).

Although participants in the studies of Weber et al. did discriminate among attributes (about 80% of the respondents had a ratio "largest vs. smallest weight" larger than 3 when there were 4 or more attributes to evaluate) there is clear evidence of partition dependence in their allocation of weights. Apparently, these respondents tended toward maximum entropy when evaluating attributes. The nature and psychological sources of this behavior (and whether the tendency reflects anchoring on even allocation or hedging in that direction) have yet to be identified. Langer and Fox (2004) obtained further evidence of naïve diversification in allocation among two simple three-outcome lotteries whose returns both depended on the roll of a particular die.

2.2. Partition dependence in resource allocation

One of the fundamental decisions faced by managers is how to allocate resources among diverse projects, firm units, or stakeholders. First, managers must often allocate discretionary funds among uncertain prospects (e.g., research and development projects) or fixed capital budgets among different divisions of a firm. In these cases they may trade risk against reward, factor in such considerations as relative need and projected return on investment, and/or take into account the perceived correlation among returns in the portfolio of projects. Second, managers must frequently allocate benefits and burdens among organizational actors in as just a manner as possible. Because all of these allocation decisions entail distribution of a scarce resource over a fixed set of alternatives, maximum entropy heuristics may be implicated and partition dependence may be observed. We consider each domain in turn.

2.2.1. Allocation among risky and uncertain projects

Relatively little descriptive research has been completed to date on allocation decisions by managers among risky or uncertain projects. However, there has been some illuminating and relevant work on personal investment decisions. Samuelson and Zeckhauser (1988, pp. 31–33) reported that about a half of a large sample of university employees allocated their retirement funds equally between stocks and bonds. These authors argued that employees rely on equal division in addition to or in place of more fundamental concerns, such as security or growth potential. Benartzi and Thaler (2001) showed that investors rely more generally on "naïve diversification" strategies in which they allocate $1/n$ of their investment savings among the n investment opportunities offered to them. These researchers asked UCLA employees to allocate hypothetical retirement savings between two funds, called A and B. In one manipulation, fund A was a stock fund and fund B was a bond fund. In another manipulation, A was a stock fund but B was a "balanced" fund that invested half of its assets in stocks and half in bonds. In a third condition, A was the "balanced" fund and B was a bond fund. In every condition participants exhibited a strong tendency toward naïve diversification, allocating close to half of their savings to each fund with little regard to distinguishing features of these funds. In fact, precisely even allocations were chosen by 34% of the respondents in the first condition, 21% in the second condition and 28% in the third condition, and even allocation was the modal response of all three groups.

Benartzi and Thaler documented a similar trend in field data from actual investments in 170 retirement savings plans obtained from Money Market Directories. In addition to confirming the experimental results, the empirical data allowed Benartzi and Thaler to observe the use of the $1/n$ heuristic when the number of available investment opportunities is larger than two. The average number of funds offered to any individual investor in their sample was 6.8, of which 62% were equity funds. The data showed that, in fact, nearly 62% of the total investment was allocated to equities. Moreover, the correlation between the percentage of investment options in a plan that were equity funds and the proportion of savings in the plan allocated to equities was positive and statistically significant. Langer and Fox (2004) obtained further evidence of naïve diversification in allocation among two simple three-outcome lotteries whose returns both depended on the roll of a particular die.

If people are biased to allocate investment funds evenly over the options that have been identified, then the particular way in which the investment space is partitioned should influence the resulting distribution of funds. For instance, a typical 401(k) savings plan partitions the investment space by the particular investments that happen to be offered. In one survey of UCLA employees, Benartzi and Thaler (2001) found that participants who were presented with a stock investment and a bond investment allocated a mean 54% of their retirement savings to stocks. However, participants who were offered a stock investment and a mixed stock/bond investment allocated a mean 46% to the first fund, which implies an investment of 73% of their savings in stocks.

Langer and Fox (2004) extended the notion of partition dependence in risky allocation by presenting participants with hierarchical allocation tasks. In one study MBA students were asked to allocate 401(k) savings among stocks (a passively managed S&P 500 fund), bonds (long-term US Treasury bills) and real estate (a geographically diversified real estate investment trust). Two of these investments were assigned to one fictional vendor and the third investment was assigned to a second fictional vendor. Following the description of all investments, participants were asked to allocate funds first to vendors, then to specific investments. If participants allocate savings evenly to the vendors then evenly to the funds offered by a given vendor, one would expect 50% of savings to go to the fund offered by the singleton vendor and 25% to each of two funds offered by the other vendor. Indeed, Langer and Fox found that participants allocated dramatically more money to a particular investment if it was assigned to the singleton vendor than if it was assigned to the other vendor. For instance, participants allocated a median of 38% of their savings to real estate when it was assigned to the singleton vendor, but they allocated a median of only 20% or 24% to real estate if it was paired with bonds or stocks, respectively. The authors replicated this effect using simple well-defined chance lotteries and incentive-compatible payoffs.

Thus far we have reviewed evidence of partition dependence in decisions made by individuals allocating personal funds among investment opportunities and chance lotteries. The question arises whether this phenomenon would extend to decisions made by large firms allocating investment funds among divisions. There is some empirical support in field data for the notion that firms are biased toward even allocation across divisions. For instance, Scharfstein (1999) examined capital budgeting data from divisions of 165 large conglomerates and documented a tendency to underinvest in well performing divisions and overinvest in poorly performing divisions (as compared with stand-alone industry peers). Other papers (Lamont, 1997; Shin and Stulz, 1997; Berger and Ofek, 1995) offer further evidence of cross-subsidization between divisions. Some authors attribute this "equal allocation" pattern to agency problems and "corporate socialism" in the budgeting process (Scharfstein and Stein, 2000). Although we acknowledge that social factors may contribute to such effects we conjecture that the tendency to spread out funds among divisions may in fact be driven by a more cognitive instinct to rely on naïve diversification with insufficient adjustment on the basis of factors that distinguish divisions.

New studies by Bardolet, Fox & Lovallo (2004) provide preliminary evidence that: (1) the bias toward equal allocation of capital resources among divisions occurs even in the absence of social factors; and (2) the procedure for budgeting and the hierarchical organization of the firm can give rise to dramatic partition dependence in budgeting decisions. These authors asked executive MBA students to take the role of the manager in charge of the capital allocation process in a hypothetical corporation. This firm had three main product divisions (Home Care, Beauty Care, and Health Care), each with a different number of regional subdivisions (Home Care subdivisions were located in the U.S., Europe and Latin America; Beauty Care subdivisions were located in the U.S. and Europe, and Health Care had a single subdivision located in the U.S.). Participants were provided with a brief description of the different divisions and subdivisions, together with some data concerning past performance and future prospects. One group of respondents (representing a firm with centralized decision making) was asked to divide the available capital among the six subdivisions. A second group (representing a firm with decentralized decision making) was asked to divide the capital only among the three major divisions (Home Care, Beauty Care and Health Care). Note that in both experimental conditions firm characteristics were held constant and the "Health Care (U.S.)" subdivision was one of the groups to which capital was to be assigned. Responses exhibited pronounced partition dependence: the median allocation to "Health Care (U.S.)" was 33% in the decentralized firm (in which it was one of three major functional divisions among which capital was divided) but only 20% in the centralized firm (in which it was one of six subdivisions among which capital was divided).

The psychological basis of naïve diversification has not yet been uniquely identified. It may be that people view spreading out their investments and contributions as a "safe" or risk-averse decision that usually reduces variance in the probability distribution over outcomes. Alternatively, naïve diversification could be viewed as an obvious and defensible default in the face of innumerable allocation possibilities; indeed, previous studies have documented a host of situations in which people choose according to "reason-based" decision rules rather than which option offers the highest perceived value (Shafir, Simonson, & Tversky, 1993). If this is the case one might expect people to shift from naïve diversification to an alternative decision rule when the former becomes difficult or impractical to apply – for instance, when the number of investment funds is rather large. Although Benartzi and Thaler (2001) observed no change in the behavior of investors in plans with more available options, they noted that the number of options in the plans in their sample was small and hypothesized that people would stop naively diversifying if they were offered a larger number of options. Indeed, Huberman and Jiang (2004) recently reported that when the number of available funds in 401(k) plans rises above a manageable quantity, people shift from naïve diversification to concentrating their investment in a small number of relatively safe funds.

Several factors seem to moderate the extent to which people rely on naïve diversification and therefore exhibit partition dependence. First, as mentioned above, people seem to rely less on these strategies as the number of options increases beyond a manageable number (Huberman & Jiang, 2004). Second, knowledge of relevant markets or investments may moderate the magnitude of partition dependence, just as substantive expertise seems to moderate partition dependence in probability judgment. We note that Langer and Fox (2004) found no significant knowledge effect when they asked respondents to rate their own knowledge in one of their experiments; however, their participants were MBA students whose financial and statistical backgrounds did not vary widely. Third, reliance on naïve diversification may diminish with increasing motivation of the decision maker. For example, Scharfstein (1999) observed that multi-business conglomerates were less prone to "corporate socialism" (i.e., a bias toward diversification) when their managers had a larger equity stake in the company. Finally, we speculate that the magnitude of adjustment from even allocation may be influenced by variation in an investor's confidence in his or her ability to predict the return on particular investments. Past research suggests that decision makers are more willing to act on domains of uncertainty about which they feel knowledgeable or competent (Heath & Tversky, 1991) relative to salient standards of comparison (Fox & Tversky, 1995; Fox & Weber, 2002). We suspect that when making allocation decisions among uncertain prospects, each prospect is compared against others in the option set and/or potential opportunities in the same investment category (e.g., other possible stock indices). Indeed, French and Poterba (1991) reported that investors in the USA, Japan and the UK allocate 94%, 98% and 92% of their overall equity investment, respectively, to domestic equities, showing a strong "home bias" that is difficult to defend on normative grounds (see also Kilka & Weber, 2000). Similarly, investors in regional telephone companies tend to invest overwhelmingly in companies located in their home state (Huberman, 2001) and Finnish investors tend to invest in companies whose headquarters are located closer to their homes and whose CEO shares their ethnicity (Grinblatt and Keloharju, 2001).

2.2.2. *Fair division of benefits and burdens*

Managers must frequently make decisions concerning the allocation of benefits and burdens among organizational actors. A great deal of research on distributive justice has found that people are sensitive to the perceived fairness of distributions of both tangible resources and working conditions (e.g., Adams, 1965; Deutsch, 1985; Leventhal, 1976; Rescher, 1966). Whether an allocation is perceived to be fair is highly context-dependent and such assessments can vary with the type of resource being allocated (e.g., monetary versus nonmonetary; benefits versus costs) and the distributional norm that the judge invokes (Deutsch, 1985). Common allocation norms include (but are not limited to) merit, effort, ability, need, equity, or equality (Deutsch, 1985; Rescher, 1966). In many allocation settings, equality (i.e., equal division among parties) is the most obvious and simple rule to apply (Messick and Schell, 1992; Messick, 1993). Yaari and Bar-Hillel (1984; see also Bar-Hillel & Yaari, 1993) examine participants' intuitions concerning just allocation of divisible entities (i.e., benefits or burdens) among individuals who have no prior claims on those entities. They argue that "equal treatment of equals" is considered the default distribution but that people find departures from even allocations warranted in response to differences in needs, tastes, or beliefs of the individuals in question.

To the extent that people apply the equality heuristic in assigning benefits and burdens, the final allocation should depend crucially on the way in which the set of people is partitioned. Indeed, Fox, Ratner and Lieb (2004) asked participants to imagine that they are executors of an estate, charged with allocating money to the deceased's grandchildren, two of whom were children of one son and four of whom were children of a second son. Respondents in the *hierarchical* condition were first asked how much they would allocate to the children of each son, and then were asked how much they would allocate to each of the grandchildren. Note that if the equality heuristic is applied at the level of sons and then at the level of grandchildren, 1/4 of the money will be allocated to each of the two children of the first son ($1/2 \times 1/2 = 1/4$) and 1/8 of the money will be allocated to each of the four children of the second son ($1/2 \times 1/4 = 1/8$). Respondents in the *non-hierarchical* condition were simply asked how much they would allocate to each of the grandchildren. Note that the equality heuristic in this case implies that 1/6 of the money will be allocated to each of the six grandchildren. Thus, the authors predicted that participants would be more likely to allocate money evenly on a *per stirpes* basis (by son) in the hierarchical condition and they would be more likely to allocate money evenly on a *per capita* basis (by grandchild) in the non-hierarchical condition. Indeed, respondents were about three times as likely to allocate an equal amount to each family of children (i.e., unevenly across grandchildren) in the hierarchical than non-hierarchical condition (25% versus 8.5%). Likewise, 84% of participants allocated equally to each grandchild (i.e., unevenly to the two son's families) in the non-hierarchical condition, whereas only 67% allocated equally to each grandchild in the hierarchical condition.

Apparently more than 90% of participants in the foregoing study relied on some application of the equality heuristic to the level of sons and/or grandchildren. The question arises whether partition dependence will be observed in situations where alternative fairness norms prevail. Fox, Ratner and Lieb (2004) asked participants to allocate financial aid among entering college freshmen whose family household incomes fell into various ranges. Respondents in the "low-partition" condition assigned percentages to the ranges {< $15,000; $15,001–$30,000; $30,001–$45,000; $45,001–$60,000; $60,001–$75,000; >$75,000}. Respondents in the "high-partition" condition assigned percentages to the ranges {< $75,000; $75,001–$85,000; $85,001–$100,000; $100,001–$120,000; $120,001–$145,000; >$145,000}. Participants were explicitly told that they should "feel free to indicate 0% or 100% for any of the categories below as these income categories were chosen arbitrarily." First, the results showed strong evidence of need-based fairness norms: mean allocation percentages were largest for the lowest income category in both conditions and allocation percentages decreased monotonically as income level increased. Second and more important, participants were sensitive to the stated categories: the mean percentage of financial aid allocated to families with incomes less than or equal to $75,000 was 96% in the low-partition condition (in which this comprised five of six income categories), but only 48% in the high-partition condition (in which this was one of six income categories). Thus, participants seemed to rely on both need-based allocation and even allocation over all specified categories.

We suspect that partition dependence will be less pronounced in situations where people have: (1) clearer criteria to distinguish among individuals or groups of individuals in allocating resources, and (2) greater cognitive resources to distinguish among individuals or subgroups. In support of this latter assertion, Roch, Lane, Samuelson, Allison and Dent (2000) argued that when people request a share of some common resource pool they tend to anchor first on an equal division (i.e., invoke the equality heuristic) then adjust their request in a self-serving manner (e.g., following norms that are based on the order of picking, number of people sharing the resource pool, etc.). The authors found that when participants were placed under cognitive load by being asked to remember a long string of numbers they were less likely to access rationale for adjustment and therefore made requests that were closer to the default equal allocation.

2.3. *Partition dependence in consumer choice*

Consumers are often called on to make multiple selections of goods and services from a menu of possibilities. Rational choice theory assumes that decision makers select sets of options that maximize their aggregate utility of consumption. Thus, normatively equivalent procedures for eliciting preferences should not affect consumers' choices. Recent investigations of multiple choices by consumers have revealed instead that they often sacrifice pleasure of consumption in order to obtain assortments with greater variety. Simonson (1990) observed that when students were asked to choose three snacks to be consumed one-at-a-time over the following three weeks, they tended to choose a variety of different items, but when they were asked on three consecutive weeks to choose a single snack to be consumed immediately they instead tended to request the same item each time. More recently, Ratner, Kahn, and Kahneman (1999) found that even when choices were made sequentially people often chose less-preferred items in order to secure greater variety. Several explanations have been advanced to explain variety-seeking behavior, including: concerns about satiation (e.g., McAlister, 1982), a desire for novelty and change (e.g., Venkatesan, 1973), and risk aversion due to uncertainty concerning future preferences (Kahn & Lehmann, 1991; Simonson, 1990; for an early review see McAlister & Pessemier, 1982).

Recent work has demonstrated that the implications of variety-seeking behavior depend crucially on the way in which the set of options is subjectively grouped. For instance, in one study Fox, Ratner, and Lieb (2004) asked participants to choose three films from a list of six. Participants were told that some of them would receive free video rentals of all of their choices. For each film participants received information concerning the title, actors in starring roles, classification (drama, action, or comedy), country of origin (Australia, Canada, or Britain), and a brief plot synopsis. Participants in the "genre partition" condition saw the movies grouped together by genre (Action, Comedy, Drama), with two films for each category. Participants in the "country partition" condition saw the same six movies grouped together by country of origin (Canada, Britain, Australia), with two films for each category. The authors hypothesized that participants would seek more variety over the attribute that is made accessible through the grouping manipulation (genre versus country). The results reveal strong evidence of partition dependence: 47% of participants chose videos from all three genres in the genre partition condition but only 20% did so in the country partition condition; similarly, 63% of participants chose videos from all three countries in the country partition condition but only 47% did so in the genre partition condition.

Similar results were obtained in a follow-up study using a more subtle manipulation of physical grouping. Four familiar varieties of candy (smarties, bubble gum, tootsie rolls, starlight mints) were displayed in three large plastic bowls. For all participants, one bowl contained two flavors (in separate piles), and the other two bowls each contained a single flavor. When participants selected five candies to take home with them they acted as if they were diversifying over bowls and were roughly 50% more likely to choose a type of candy when it was placed on a bowl by itself than when it was placed in a bowl with another type of candy. Interestingly, this tendency to spread out consumption over bowls was significantly diminished among participants who were asked to remember an eight-digit number while making their selections, suggesting that cognitive load may interfere with higher cognitive motives to diversify.

Partition dependence in consumer choice appears to be moderated by the strength and accessibility of preferences among options. Fox, Ratner and Lieb (2004) replicated the aforementioned video selection study by asking graduate students to choose three different bottles of white wine from a list of six that were either grouped by grape (Chardonnay, Pinot Grigio, Sauvignon Blanc) or by region of origin (Australia, California, Italy). In addition, participants were asked to indicate the number of bottles of white wine that they had purchased in the previous twelve months. The results revealed partition dependence that was more pronounced among respondents who had purchased fewer bottles of wine, thereby supporting the notion that expertise moderates reliance on variety-seeking over accessible categories. More direct evidence for the notion that strength and salience of preferences, rather than expertise per se, moderates partition dependence was obtained in a follow-up study in which MBA students were asked to complete two tasks: (1) choose three items from a list of eight snacks available in their student kiosk that were grouped into three categories, {cookies, crackers, fruits & veggies} or {cookies & crackers, fruits, veggies}; (2) rate how attractive they found the prospect of consuming each of these items. Results revealed partition dependence that was less pronounced among participants who had rated items before choosing (so that hedonic preferences were more accessible when they subsequently chose), and among participants who exhibited stronger hedonic preferences (as measured by a higher variance in attractiveness ratings over items).

3. DISCUSSION

The studies reported in this paper support the notion across a wide range of domains that when people allocate scarce resources (belief, attribute weights, money, choices) among a fixed set of options (events, attributes, projects, individuals or groups, consumption goods) they tend to invoke *maximum entropy heuristics* in which they distribute the resource evenly across all options and adjust to the extent that they distinguish among them. Such even allocation strategies require people to first subjectively partition the set of options into groups. In many (if not most) situations there is no single canonical partition and the relative accessibility of alternative partitions is influenced by spurious factors such as the elicitation procedure (e.g., hierarchical versus non-hierarchical), convenient category cutoffs, or physical groupings. Thus, allocations tend to exhibit *partition dependence*, varying systematically with the partition that happens to be most accessible to the decision maker. Moreover, we have seen that the relative accessibility of alternative partitions can be manipulated experimentally. The heuristics reviewed in this paper and associated manifestations of partition dependence are summarized in Table 1. We conclude this chapter with a discussion of possible demand effects, related phenomena, extensions of the present work, and areas for future research.

3.1. Demand effects

We have argued that partition dependence violates rational choice theory because it gives rise to allocations that differ across strategically equivalent elicitation modes. However, one might be concerned that the method of eliciting allocations or describing possibilities could, in fact, communicate information to participants in the studies cited. For instance, rational probability assessors may infer that all events for which they are asked to assign probabilities must have a nontrivial likelihood of occurrence otherwise they would not be asked about these events. Similarly, a rational judge might infer that different income ranges that have been identified by the experimenter for allocating financial aid must have approximately equal

Table 1. Summary of Research on Maximum Entropy Heuristics and Partition Dependence.

Domain	*Scarce Resource*	*Allocation Space*	*Maximum Entropy Heuristic*	*Representative References*
1. Decision Analysis	Belief	Event Space	Principle of Insufficient Reason	Fox & Clemen (2004)
	Importance Weights	Attributes	?	Weber, Eisenführ & von Winterfeldt (1988)
2. Organizational Resource Allocation	Money	Investments	Naïve Diversification	Benartzi & Thaler (2001) Langer & Fox (2004)
	Benefits and burdens	Group Members	Equality Heuristic	Messick (1993) Fox, Ratner & Lieb (2004)
3. Consumer Choice	Fixed number of choices, money	Consumption Options	Variety-seeking	Simonson (1988) Fox, Ratner & Lieb (2004)

representation in the population. This argument suggests that partition dependence is a demand effect whereby a participant considers the assessment as an implicit conversation with the experimenter in which the experimenter is expected to adhere to accepted conversational norms, including the notion that any contribution should be relevant to the aims of the conversation (Grice, 1975; Orne, 1962).

Although we agree that in some instances norms of conversational implicature may play a role in partition dependence, we assert that they do not provide an adequate explanation of this phenomenon for several reasons. First, several of the studies reviewed here demonstrate the robustness of partition dependence even in the presence of monetary incentives (e.g., participants allocated money over chance lotteries in which some of them were to play their choices for real money) or real choices (e.g., all participants selected candies to take home). Second, some studies demonstrate the robustness of partition dependence when participants are explicitly told that the categories into which the space is partitioned are arbitrary and that they should feel free to allocate 0% or 100% to any of the categories identified (e.g., when participants allocated financial aid to families of different income ranges). Third, some studies demonstrate the robustness of partition dependence when participants are explicitly made aware of the different partitions that are being presented to different groups (e.g., when participants assigned themselves on the basis of their telephone number to different partitions of the closing values of stock indices, for which they judged the probability). Fourth, the "demand effect" interpretation cannot readily explain the finding in several studies that participants very often provide precisely even allocations (e.g., the median allocation of probabilities for the Jakarta Stock Index was precisely 1/4 for each of four ranges, regardless of partition condition). It could be argued that conversational norms would imply that an allocation question is relevant only if the experimenter thinks that reasonable respondents will distinguish among the available options. Finally, even if respondents do surmise that there is information conveyed by the particular partition with which they are presented, we assert that people may draw such conclusions not only in the laboratory but also in more naturalistic settings in which the partition is determined by arbitrary factors.

3.2. Related phenomena

The notion that different descriptions of the same problem facilitate different psychological representations – and in turn different responses – has been observed in a variety of other domains of decision making. In prospect theory (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992) choices and attitudes toward risk are affected by how participants subjectively represent options in terms of losses and gains relative to a reference point, and this representation can be manipulated through variations in the framing of options (e.g., Tversky & Kahneman, 1986). For instance, physicians' recommendations varied as a function of whether the possible outcomes of particular treatments were described in terms of survival versus mortality rates (McNeil, Pauker, Sox, & Tversky, 1982). Researchers have also found that the subjective packaging of consequences ("mental accounting") can influence choices.

For instance, people generally find positive outcomes more attractive when they are segregated and they find negative consequences less unattractive when they are integrated (Thaler, 1985, 1999). For example, most respondents in one study thought that a person who had won $50 in one lottery and $25 in another would be happier than a person who had won $75 in a single lottery; similarly, most respondents thought that a person who had received a letter from the IRS saying that there had been a mistake in his tax return and that he owed $100 and another letter from the state income tax authority saying that he owed $50 would be more upset than another person who only received one letter saying that he owed $150 to the IRS due to a mistake in his tax return.

Description-dependence has also been observed in other domains of judgment and decision making, as well as reasoning and problem solving. Events (e.g., precipitation in Chicago next April 1) are often judged to be more likely when they are described as disjunctions of constituent events (e.g., rain or sleet or snow or hail in Chicago next April 1), a phenomenon known as "unpacking" (Rottenstreich & Tversky, 1997; but see Sloman et al., 2004). Similar effects have been observed in pricing of prospects: in one study, participants were willing to pay more for an insurance policy that covered "hospitalization for any accident or disease" than they were for a policy that covered "hospitalization for any reason" (Johnson, Hershey, Meszaros, & Kunreuther, 1993). People are more successful verifying logical statements when they are described in terms of familiar and concrete events than when they are described in more abstract terms (Wason and Johnson-Laird, 1972). People also have an easier time solving challenging puzzles when they are represented using isomorphs that impose lower working memory demands (Kotovsky, Hayes & Simon, 1985).

3.3. Extensions

Although we have argued that partition dependence derives from reliance on maximum entropy heuristics, we assert that partition dependence may also be observed in situations where people apply *minimum* entropy heuristics. If decision makers in some contexts prefer not to diversify but instead mass their allocations on a single category, then the way in which they partition the set of available options will also influence their resulting choices. For instance, suppose that an uncle shopping for holiday gifts for four nieces wishes to give them all the same "kind" of present so that they are not envious of one another. He enters a toy store that sells books, stuffed animals, crafts, and games. If this is the way the uncle categorizes the options then he may select books for all four nieces. If instead he partitions the option set into educational gifts (books and crafts) versus non-educational gifts (stuffed animals and games) he may stick to educational gifts but pick out some books and some crafts; if he partitions the set instead by gifts for individual use (books and stuffed animals) versus gifts for interactive use (crafts and games) then he may favor interactive toys and purchase some crafts and some games. Thus, even in a situation where the uncle invokes a minimum entropy heuristic, the way in which he partitions the option set influences his allocation of choices.

Reliance on maximum entropy heuristics can lead not only to partition dependence but also other systematic dependencies that cannot be easily reconciled with rational choice theory. First, the implications of even allocation may depend on the units being allocated. Langer and Fox (2004) presented Duke University graduate students with two $20,000 portfolios that consisted of Apple and IBM stocks – one portfolio in which an equal number of shares of each stock were purchased and one portfolio in which an approximately even number of dollars were invested in each stock. All participants were also told the prevailing price at which each stock traded (shares of IBM stock were much more expensive than shares of Apple stock). One group was asked to choose between these portfolios described in terms of the number of dollars invested in each stock; another group was asked to choose between the same portfolios described in terms of the number of shares invested in each stock. The results were striking: most participants preferred the equal dollars portfolio when both portfolios were described in terms of the number of dollars invested, whereas most participants preferred the equal share portfolio when both portfolios were described in terms of the number of shares invested. Langer and Fox replicated this finding using chance gambles and monetary incentives.

Likewise, the decision to allocate benefits or burdens equally can have different implications depending on the particular units that are being allocated. Harris and Joyce (1980; see also Messick & Schell, 1992) presented participants with a scenario in which five partners took turns selling plants at a stand in a flea market. Partners accumulated joint expenses for running the business and each generated different amounts of revenue from their individual efforts. Some participants were asked to allocate a "fair share" of the joint *profits* among the partners. In this case 43% of participants allocated profits equally and only 1% of participants allocated expenses equally. In contrast, a second group was asked to allocate a "fair share" of the *expenses* among the partners. In this case no participant recommended that profits be allocated equally and 38% of participants recommended that expenses be allocated equally.

In addition to partition- and unit-dependence, the impact of maximum entropy heuristics may depend on the particular procedure that is used to elicit judgments or preferences. Benartzi and Thaler (2001) noted that the tendency to allocate retirement savings evenly across investment instruments was much less pronounced if participants were instead asked to choose among different portfolios that were mixtures of base investments. In one study, they asked a first group of participants to allocate savings between a stock fund and a bond fund, and provided these participants with 27 yearly rates of return for each fund, depicted graphically. They asked a second group to choose among five funds that were (unbeknownst to the participants) mixtures of those two base funds, and provided these participants with yearly returns of each mixture as well the average aggregate return for each mixture. The difference between these elicitation modes was dramatic: participants in the "allocation" condition assigned 56% of their hypothetical savings to the stock fund, whereas participants in the "choice" condition assigned 75% of their hypothetical savings to stocks. Langer and Fox (2004) obtained a similar result using well-specified chance gambles and monetary incentives.

3.4. Areas for future research

We have provided evidence that people rely on a variety of maximum entropy heuristics in a number of arenas of judgment and decision making. We have argued further that the way in which the set of events, attributes, or options are described may influence the partition that people subjectively invoke, and therefore the pattern of judgments or choices that they make. Further research is needed to answer several important follow-up questions.

First, it would be useful to further explore moderators of partition dependence in order to better understand the psychological factors that underlie this phenomenon. We have seen that in some cases people adjust more from maximum entropy distributions and are less susceptible to partition dependence when they are more knowledgeable concerning the events, attributes, or options in question. This pattern was observed in studies of subjective probability assessment (Fox & Clemen, 2004) and consumer choice (Fox, Ratner & Lieb, 2004). It would be interesting to see whether this tendency for knowledge or information to moderate partition dependence extends to other domains such as attribute weighting and resource allocation decisions.

In some situations we surmise that maximum entropy heuristics are spontaneous, associative assessments or default strategies that entail a minimum of conscious attention or reflection, and adjustment entails more conscious elaboration of belief or preference. In such cases we suspect that cognitive load manipulations will tend to suppress the adjustment process and exacerbate partition dependence (as in the fairness study of Roch, et al. 2000). In other situations it may be that maximum entropy heuristics represent a conscious tendency to hedge away from spontaneous, associative assessments of underlying belief or preference that discriminate among possibilities. In these cases cognitive load may instead mitigate partition dependence (as in the bowls of candy study of Fox, Ratner & Lieb, 2004). Likewise, we surmise that time pressure will tend to curtail cognitive elaboration and may in some cases exacerbate or mitigate partition dependence. Also, nonconscious priming of motives may activate or compete with a particular maximum entropy heuristic (see e.g., Bargh & Chartrand, 1999). For instance, variety-seeking in consumer choice seems to be more pronounced in public than in private choice settings (Ratner & Kahn, 2002; Ariely & Levav, 2000), suggesting that activation of a social norm may play a role. Perhaps nonconscious priming of words such as "daring," "novelty," "satiation," or "uncertainty" will exacerbate variety-seeking and likewise resulting partition dependence.

In most of the studies reviewed in this chapter alternative partitions have been made more accessible through experimental manipulations. A second topic for future research is to better understand features of natural environments that influence the partitions that people normally invoke. We surmise that in some cases people are influenced by exogenous factors such as decision trees in which partitions are defined by the analyst. Likewise, subjective grouping may be influenced by physical features of an environment, such as the grouping of product varieties on a supermarket shelf. We suspect that in other cases partitions are determined by entirely endogenous factors and that research literature on learning and categorization may be helpful in better understanding the subjective partitions that people naturally project on their environment.

Finally, we believe that future research ought to address prescriptive methods for overcoming partition dependence. Several approaches might be fruitful. First, to the extent that one might identify a canonical partition, this ought to be made explicit. For instance, a firm trying to assess the probability that a competitive bid will be accepted by a client may find it natural to assess the probabilities that each firm in the running will have their bid accepted, thereby partitioning by firm. On the other hand, in most situations a single, canonical partition cannot be identified. For example, in judging the probabilities of possible future interest rates one year from today, there is no single, obvious criterion by which to parse the event space. In such cases we recommend that managers and consumers invoke multiple partitions and multiple elicitation methods, attempting to actively reconcile any discrepancies that arise. For instance, participants might assess probabilities for various partitions of the same space and also assess confidence intervals, then try to integrate the output of these disparate methods. Finally, we suggest that the relevant judgment or decision making process might be formally modeled so that the extent of the bias across partitions might be measured and thereby subtracted from the relevant assessment. Future research is needed to identify proper parameterizations of such models and assess their validity.

ACKNOWLEDGMENT

Completion of this project was supported in part by NSF grant SES-0099209 to Craig Fox. We thank Kelly See for helpful discussions.

REFERENCES

Abdellaoui, M. (2000). "Parameter-free elicitation of utilities and probability weighting functions." *Management Science*, 46, 1497–1512.

Adams, J. S. (1965). "Inequity in social exchange." In Berkowitz, L. (ed.), *Advances in Experimental Social Psychology* (Vol. 2, pp. 267–299). New York: Academic Press.

Ariely, D. & Levav, J. (2000). "Sequential choice in group settings: Taking the road less traveled and less enjoyed." *Journal of Consumer Research*, 27, 279–290.

Bardolet, D., Fox, C. R. & Lovallo, D. (2004). "Partition dependence in capital budgeting." Unpublished data, Anderson School of Management, UCLA.

Bargh, J. A. & Chartrand, T. L. (1999). "The unbearable automaticity of being." *American Psychologist*, 54, 462–479.

Bar-Hillel, M. & Yaari, M. (1993). "Judgments of distributive justice." In B. A. Mellers and J. Baron (eds.), Psychological perspectives on justice: Theory and applications, pp. 56–84. New York: Cambridge University Press.

Benartzi, R. & Thaler, R. (2001). "Naïve diversification strategies in retirement saving plans." *American Economic Review*, 91, 475–482.

Berger, P. & Ofek, E. (1995). "Diversification's effect on firm value." *Journal of Financial Economics*, 37, 39–65.

Clemen, R. T. (1996). *"Making Hard Decisions (2nd ed.)."* Belmont, CA: Duxbury Press.

Deutsch, M. (1985). "Distributive justice: A social-psychological perspective." New Haven: Yale University Press.

Edwards, W. (1977). "How to use multiattribute utility theory for social decision making." *IEEE Trans. Systems, Man, Cybernet.*, 7, 326–340.

Fischer, G., Damodaran, N., Laskey, K. & Lincoln, D. (1987). "Preferences for proxy attributes: The overweighting bias." *Management Science*, 33, 198–214.

Fischhoff, B., Slovic, P. & Lichtenstein, S. (1978). "Fault trees: Sensitivity of estimated failure probabilities to problem representation." *J. Experimental Psychology: Human Perception and Performance*, 4, 330–334.

Fox, C. R. & Clemen, R. T. (2004). "Partition dependence in subjective probability assessment." Unpublished manuscript, Anderson School of Management, UCLA.

Fox, C. R. & Levav, J. (in press). "Partition-edit-count: Naïve extensional reasoning in judgment of conditional probability." *Journal of Experimental Psychology: General*, forthcoming.

Fox, C. R., Ratner, R. K. & Lieb, D. (2004). "How subjective grouping of options influences choice and allocation: Diversification bias and the phenomenon of partition dependence." Unpublished Manuscript, the Anderson School at UCLA.

Fox, C. R. & Rottenstreich, Y. (2003). "Partition priming in judgment under uncertainty." *Psychological Science*, 13, 195–200.

Fox, C. R. & Tversky, A. (1995). "Ambiguity aversion and comparative ignorance." *Quarterly Journal of Economics*, 110, 585–603.

Fox, C. R. & Weber, M. (2002). "Ambiguity aversion, comparative ignorance, and decision context." *Organizational Behavior and Human Decision Processes*, 88, 476–498.

French, K. & Poterba, J. (1991). "Investor diversification and international equity markets." *American Economic Review*, 81(2), 222–226.

Graham, J. & Harvey, C. (2001). "The theory and practice of corporate finance: evidence from the field." *Journal of Financial Economics*, 60, 187–243.

Green, P. & Srinivasan, V. (1978). "Conjoint analysis in consumer behavior: Status and outlook." *J. Consumer Res.*, 5, 103–123.

Grice, H. P. (1975). "Logic and conversation." In P. Cole & J. L. Morgan (eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41–58). New York: Academic Press.

Grinblatt, M. & M. Keloharju (2001). "How distance, language, and culture influence stockholdings and trades." *Journal of Finance*, 56, 1053–1073.

Hacking, I. (1975). *The Emergence of Probability.* Cambridge, England: Cambridge University Press.

Harris, R. J. & Joyce, M. A. (1980). "What's fair? It depends on how you phrase the question." *Journal of Personality and Social Psychology*, 38, 165–179.

Heath, C. & Tversky, A. (1991) Preference and beliefs; Ambiguity and competence in choice under uncertainty." *Journal of Risk and Uncertainty*, 4, 5–28.

Huberman, G. (2001). "Familiarity breeds investment, *Review of Financial Studies*, 14, 659–680.

Huberman, G. & Jiang, W. (2004). The *1/n* heuristic in 401(k) plans. Unpublished manuscript, Columbia University Graduate School of Business, New York, NY.

Johnson, E. J., Hershey, J., Meszaros, J. & Kunreuther, H. (1993). "Framing, probability distortions, and insurance decisions." *Journal of Risk and Uncertainty*, 7, 35–51.

Kahn, B. E. & Lehmann, D. R. (1991). "Modeling choice among assortments." *Journal of Retailing*, 67 (Fall), 274–299.

Kahneman, D. & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. pp. 49–81 *in* T. Gilovich, D. Griffin, and D. Kahneman (eds.). *Heuristics & Biases: The Psychology of Intuitive Judgment.* New York. Cambridge University Press.

Kahneman, D. & Tversky, A. (1994). "Choices, values and frames." *American Psychologist*, 39, 341–350. (Reprinted as Ch. 1 in Kahneman, D., and Tversky, A. (eds.), Choices, Values and Frames. New York: Cambridge University Press and the Russell Sage Foundation, 2000).

Kahneman, D. & Tversky, A. (1979). "Prospect theory: An analysis of decisions under risk." *Econometrica*, 47, 263–291.

Keefer, D. L. Kirkwood, C. W. & Corner, J. L. (2004). Perspective on decision analysis applications, 1990–2001, Decision Analysis, 1, 5–24.

Keeney, R. L. & Raiffa, H. (1976) *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*, New York: Wiley.

Keeney, R. L., Renn, O. & von Winterfeldt, D., (August 1987). "Structuring Germany's energy objectives." *Energy Policy*, 352–362.

Keeney, R. & von Winterfeldt, D. (1991). "Eliciting probabilities from experts in complex technical problems." *IEEE Transactions on Engineering Management*, 38, 191–201.

Kilka, M. & Weber, M. (2000). "Home-bias in International stock return expectations." *Journal of Psychology and Financial Markets*, 1, 176–193.

Kotovsky, K., Hayes, J. R. & Simon, H. A. (1985). "Why are some problems hard?: Evidence from the tower of Hanoi." *Cognitive Psychology*, 17, 248–294.

Lamont, O. (1997). "Cash flow and investment: Evidence from internal capital markets." *Journal of Finance*, 52, 83–109.

Langer, T. & Fox, C. R. (2004). Naïve diversification and partition dependence in investment allocation decisions: An experimental investigation. Unpublished Manuscript, University of Mannheim.

Leventhal, G. S. (1976). "The distribution of rewards and resources in groups and organizations." In Berkowitz, L., and Walster, E. (eds.), *Advances in experimental social psychology*, 9, 91–131. New York: Academic Press.

Loewenstein, G. & Elster, J. (1992). *Choice over time*. New York: Russell Sage Foundation.

Loewenstein, G. F. & Prelec, D. (1993). "Preferences for sequences of outcomes." *Psychological Review*, 100, 91–108.

McAlister, L. (1982). "A dynamic attribute satiation model of variety-seeking behavior." *Journal of Consumer Research*, 9, 141–150.

McAlister, L. & Pessemier (1982). "Variety-seeking behavior: An interdisciplinary review." *Journal of Consumer Research*, 9, 311–322.

McNeil, B. J., Pauker, S. G., Soc, H. G. & Tversky, A. (1982). "On the elicitation of preferences for alternative therapies." *New England Journal of Medicine*, 306, 1259–62.

Messick, D. M. (1993). "Equality as a decision heuristic." In Mellers, B. A. & Baron, J. (eds.), *Psychological perspectives on justice: Theory and applications*. Cambridge series on judgment and decision making. (pp. 11–31). New York: Cambridge University Press.

Messick, D. M. & Schell, T. (1992). "Evidence for an equality heuristic in social decision making." *Acta Psychologica*, 80(1–3), 311–323.

Morgan, G. & Henrion, M. (1990). *Uncertainty, a Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. New York: Cambridge University Press.

Orne, M. T. (1962). "On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications." *American Psychologist*, 17, 776–783.

Ratner, R. K. & Kahn, B. K. (2002). "The Impact of Private versus Public Consumption on Variety-Seeking Behavior." *Journal of Consumer Research*, 29, 246–257.

Ratner, R. K., Kahn, B. E. & Kahneman, D. (1999). "Choosing less-preferred experiences for the sake of variety." *Journal of Consumer Research*, 26, 1–15.

Rescher, N. P. (1966). Distributive justice. New York: Bobbs-Merrill.

Roch, S., Lane, J., Samuelson, C., Allison, S. & Dent, J. (2000). "Cognitive load and the equality heuristic: A two-stage model of resource overconsumption in small groups." *Organizational Behavior and Human Decision Processes*, 83, 185–212.

Rottenstreich, Y. & Tversky, A. (1997). "Unpacking, repacking, and anchoring: Advances in support theory." *Psychological Review*, 104, 406–415.

Samuelson, W. & Zeckhauser, R. (1988). "Status quo bias in decision making." *Journal of Risk & Uncertainty*, 1:1 (March), 7–59.

Scharfstein, D. S. (1999). The Dark side of internal capital markets II: evidence from diversified conglomerates, NBER working paper #6352.

Scharfstein, D. & Stein, J. (2000). "The Dark side of internal capital markets: Divisional rent-seeking and inefficient investment." *Journal of Finance*, 55 (6).

See, K. E. & Clemen, R. T. (2003). Psychological and organizational factors influencing the use of decision analysis. Working paper, Fuqua School of Business, Duke University.

See, K. E., Fox, C. R. & Rottenstreich, Y. (2004). Partition dependence and learning in judgment under uncertainty. Unpublished manuscript, Fuqua School of Business, Durham, NC.

Shafir, E. B., Simonson, I. & Tversky, A. (1993). "Reason-based choice." *Cognition*, 49 (October–November), 11–36.

Shin, H. & Stulz, R. (1997). Are internal capital markets efficient, Ohio State University, working paper, No. 97-4.

Simonson, I. (1990). "The effect of purchase quantity and timing on variety-seeking behavior." *Journal of Marketing Research*, 32, 150–162.

Sloman, S., Rottenstreich, Y., Wisniewski, E., Hadjichristidis, C. & Fox, C. R. (2004). "Typical versus atypical unpacking and superadditive probability judgment." *Journal of Experimental Psychology: Learning, Memory & Cognition*, 30, 573–582.

Spetzler, C. S. & Staël Von Holstein, C.-A. (1975). "Probability encoding in decision analysis." *Management Science*, 22, 340–352.

Thaler, R. (1999). "Mental accounting matters." *Journal of Behavioral Decision Making*, 12, 183–203.

Thaler, R. (1985). "Mental accounting and consumer choice." *Marketing Science*, 4, 199–214.

Tversky, A. & Kahneman, D. (1992). "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty*, 5, 297–323.

Tversky, A. & Kahneman, D. (1986). "Rational choice and the framing of decisions." *Journal of Business*, 59, 251–78.

Tversky, A. & Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and biases." *Science*, 185, 1124–1131.

Venkatesan, M. (1973). "Cognitive consistency and novelty seeking." In Scott Ward and Thomas S. Robertson (eds.), *Consumer Behavior: Theoretical Sources* (pp. 355–384). Englewood Cliffs, NJ: Prentice-Hall.

Von Winterfeldt, D. & Edwards, W. (1986). *Decision Analysis in Behavioral Research*. New York: Cambridge University Press.

Wakker, P. P. & Deneffe, D. (1996). "Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown." *Management Science*, 42, 1131–1150.

Wason, P. & Johnson-Laird, P. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press.

Weber, M. (1983). "An Empirical investigation on multi-attribute decision-making." Chapter in Hansen, P. (ed.), *Essays and Surveys on Multiple Criteria Decision Making*. New York: Springer.

Weber, M., Eisenführ, F. & von Winterfeldt, D. (1988). "The Effect of splitting attributes in multiattribute utility measurement." *Management Science*, 34, 431–445.

Yaari, M. & Bar-Hillel, M. (1984). "On dividing justly." *Social Choice and Welfare*, 1, 1–24.

This page intentionally blank

Chapter 11

GENDER & COORDINATION

Martin Dufwenberg
University of Arizona

Uri Gneezy
University of Chicago

Abstract

Groups of six females or six males play the minimal effort coordination game for ten periods. Small differences in coordination are found in the initial stages but not in the final stages. Besides reporting this result, we raise a methodological issue: Is there a bias in the research community against reporting or publishing results that document the absence of a gender effect? If so, there is a risk of bias in perceptions regarding the magnitude and limits of gender differences.

1. INTRODUCTION

Are there gender differences with respect to economic behavior? Real life observation indicates that males and females are treated differently in the work place (see e.g., McPherson & Hirsch, 1995, and Lazaer & Rosen, 1990). Can the root of such differences be found in different decision-making behavior?

Until recently, relatively few experimental economics papers reported data on such matters, but the past few years this line of research has become more popular. Researchers have examined which of the two sexes is more fair or generous, or compared the discriminatory behavior of men and women, to mention a few examples.[1] The bag of results is somewhat mixed, and since results do not always point in the same direction it is too early to draw far-reaching conclusions regarding the behavioral differences of men and women. More facts are needed in order to move towards the development of a systematic theory. Our paper takes a small step in this direction. We examine whether a single sex group of only males will coordinate differently than a group of only females.

The result may have some interest from the viewpoint of organizational theory. A personnel manager may be interested in knowing whether the gender composition of the group affect productivity. In particular, which team of employees – one with women or one with men – is more productive. Our experiment may generate insights

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 253–262.
© 2005 *Springer. Printed in the Netherlands.*

that shed some light on this issue by analogy. We study behavior in a game which possesses multiple equilibria,[2] that differ in terms of "efficiency". Some equilibria lead to higher payoffs for everyone than other equilibria. The game is useful for investigating which sex is best at coordinating on a good equilibrium, and may thereby indirectly shed light on whether a male or a female team of employees is more productive.

In the next section we introduce the game. Section 3 explains the experimental procedures. Section 4 reports the results. Section 5 contains a concluding discussion. Against the background of our results, we raise a methodological concern regarding the importance of reporting and publishing results no matter their content.

2. THE GAME

We base our study on the Van Huyck, Battalio & Beil (1990) minimal effort coordination game. This game is played by a group of players, each of whom simultaneously chooses an integer from 1 to 7. The payoff to a player depends on that player's choice as well as on the minimal number chosen by any of the other players in the group (including the player her- or himself). The payoff matrix is presented in Table 1.

The game has seven pure strategy equilibria, in each of which every player chooses the same integer. The equilibria are Pareto-ranked. The one in which all players choose 7 results in the highest possible payoff to every player, while the equilibrium at which all players choose 1 gives each player the lowest payoff. Van Huyck et al's results indicate that when the game is played repeatedly the inefficient equilibrium provides an accurate description of behavior. This result has inspired a

Table 1. The payoff matrix

Smallest number chosen by the participants in your group

		7	6	5	4	3	2	1
Your Decision	**7**	130	110	90	70	50	30	10
	6	–	120	100	80	60	40	20
	5	–	–	110	90	70	50	30
	4	–	–	–	100	80	60	40
	3	–	–	–	–	90	70	50
	2	–	–	–	–	–	80	60
	1	–	–	–	–	–	–	70

large discussion regarding the condition under which coordination fails or not (see Ochs, 1995, for a survey).

3. THE DESIGN

The experiment was conducted at the Technion. Students were recruited via posters on campus. On these posters they were promised money for participating in an experiment that would take about half an hour. They were asked to call a phone number, which was written on the poster. When they did this, an answering machine replied asking them to leave their phone number and told them that they would be contacted later.

Six participants were invited by phone to each session: either six females or six males.[3] This fact was never explicitly pointed out to the subjects, but they could see the other participants in the lab. In each session, after all the six students entered the lab, they received a standard-type introduction. Participants then received the instructions for the experiment. The payoff matrix presented in Table 1 was used. The payment was in points, with 10 points worth 1 Shekel (at the time of the experiment, 4 Shekels = $1). They were allowed to ask questions privately. Five sessions with only females and five with only males were run. The coordination game was repeated 10 times, with participants receiving information about the minimum choice in the previous stage before making choices in the current round.

4. THE RESULTS

Table 2 presents the data.

Table 2. The data. Tables 1f to 5f presents the results of only females, and tables 1m to 5m the results of only males

Session 1f	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	4	2	2	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	4	1	2	2	1	2	2	2	1	1
4	5	4	2	2	1	3	1	3	1	2
5	4	4	3	3	2	2	1	1	1	1
6	7	1	1	1	1	1	1	1	1	1
Average	4.3	2.3	2.2	1.8	1.5	1.8	1.8	1.5	1.2	1.2

Table 2. (cont'd)

Session 2f	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	3	3	2	1	3	1	1	1	1	1
2	2	3	1	1	1	1	1	1	1	1
3	4	2	3	1	2	1	1	1	1	1
4	4	2	3	1	1	1	1	1	1	1
5	5	4	4	1	1	1	1	1	1	1
6	7	2	2	3	1	1	1	1	1	7
Average	4.2	2.7	2.5	1.3	1.5	1	1	1	1	2

Session 3f	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	7	2	1	7	7	7	7	1	1	1
2	3	4	2	1	2	1	1	1	1	1
3	3	3	2	1	2	1	1	1	1	1
4	1	5	3	1	1	1	1	1	1	1
5	5	1	1	2	2	1	1	1	1	1
6	3	1	1	1	2	1	1	3	1	1
Average	3.7	2.7	1.7	2.2	2.7	2	2	1.3	1	1

Session 4f	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	6	7	5	3	1	1	1	1	1	1
2	7	4	3	2	2	2	2	1	1	1
3	5	3	4	3	3	3	1	1	1	1
4	3	3	3	2	1	2	1	2	1	1
5	6	3	4	1	1	1	1	1	1	1
6	6	7	4	3	2	2	1	1	1	1
Average	5.5	4.5	3.8	2.3	1.7	1.8	1.2	1.2	1	1

Table 2. (cont'd)

Session 5f	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	7	4	2	2	2	2	2	2	2	2
2	6	3	2	2	1	1	1	1	1	1
3	7	1	1	3	2	1	1	1	1	1
4	2	3	1	1	1	1	1	1	1	1
5	4	2	1	1	1	1	1	1	1	1
6	1	2	2	1	1	1	1	1	1	1
Average	4.5	2.5	1.5	1.7	1.3	1.2	1.2	1.2	1.2	1.2

Session 1m	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	4	2	2	1	1	1	1	1	1	1
2	5	4	4	4	5	1	1	1	1	1
3	7	7	5	1	2	1	1	1	1	2
4	4	2	3	3	2	1	1	1	1	1
5	3	4	2	1	3	1	1	1	1	1
6	7	6	1	6	5	1	1	1	1	1
Average	5	4.7	2.8	2.7	3	1	1	1	1	1.2

Session 2m	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	3	3	2	2	1	1	1	2	1	1
2	4	5	4	2	2	2	1	1	1	1
3	2	2	1	1	3	1	1	1	1	1
4	7	6	5	1	1	1	1	1	1	1
5	2	2	1	1	1	1	1	1	1	1
6	7	4	1	2	2	2	1	1	1	1
Average	4.7	3.7	2.3	1.5	1.7	1.3	1	1.2	1	1

Table 2. (cont'd)

Session 3m	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	3	2	2	1	1	1	1	1	1	3
2	4	4	4	3	2	2	2	1	1	1
3	5	4	3	2	1	1	1	1	1	1
4	5	3	2	1	1	1	3	1	1	2
5	2	2	1	2	1	1	1	1	1	1
6	7	4	3	1	1	1	1	1	1	1
Average	4.3	3.2	2.5	1.7	1.2	1.2	1.5	1	1	1.5

Session 4m	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	5	4	4	4	1	1	1	1	1	1
2	3	4	4	2	2	1	1	2	1	2
3	7	5	5	2	2	1	1	1	1	5
4	3	4	5	1	1	1	1	1	1	1
5	4	4	2	4	3	2	1	1	1	1
6	5	4	4	3	1	1	1	1	1	1
Average	4.5	4.2	4	2. 7	1.7	1.2	1	1.2	1	1.8

Session 5m	**Round 1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Player 1	3	1	3	3	1	1	1	1	1	1
2	7	3	1	2	1	1	1	1	1	7
3	2	1	2	1	1	1	1	1	1	1
4	6	5	2	2	2	1	1	1	1	1
5	7	3	1	1	1	1	1	1	1	1
6	1	3	1	1	1	1	1	1	1	1
Average	4.3	2.7	1.2	1.2	1.2	1	1	1	1	2

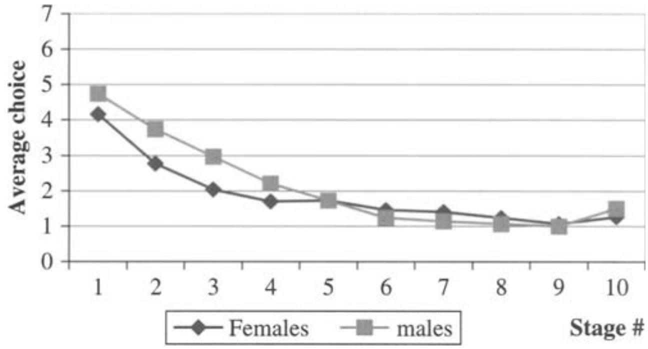


Figure 1. Average choices of all sessions

4.1. Average choice

The average choices of all sessions are presented in Figure 1.

The average of all the choices of females in stage 1 was 4.73 and the variance was 4.07. For males the average was 4.17 and the variance was 3.17.

To test whether the difference in choices between the two treatments is significant, we use the Wilcoxon rank-test. To avoid violating the independence assumption, this statistical test treated each session as a single observation. The difference between the average choices of the different treatments in stage 1 is marginally significant ($z = 1.92$, $p < .06$). Similar results are found using t-test (two-tailed, $t = 2.13$, $p < .07$). This difference is also present in stage 2: The difference is significant both according to the rank-test ($z = 2.02$, $p < .04$) and according to the t-test ($t = 2.31$, $p < .05$). Similar results are also found for stage 3, but in stage 4 the difference becomes insignificant {rank-test ($z = 1.60$, $p < .11$), t-test ($t = 1.87$, $p < .10$)}. From that stage on, the difference is insignificant (e.g. in stage 5 the average for both females and males is equal (1.73)). This difference remains insignificant until stage 10 {($z = .86$, $p < .39$), t-test ($t = .87$, $p < .41$)}.

4.2. Minimal choice

In the minimal effort game the minimal choice in each group is of course very important. Table 3 shows the minimal choice in each session. As can be seen, in stage 1 the minimal choice in three out of the five sessions with only females was 1, as compared with only one session with only males. This difference disappears at the forth stage, however. By that round, the minimum choice in all sessions is 1.

5. DISCUSSION

We have found little difference between groups of men and groups of women, when it comes to their ability to avoid the least efficient equilibrium in a minimum effort game. Our results show some differences in the initial stages of the game, but these

Table 3. The minimal choice in each session per stage

		Minimum choice in stage									
	Session	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Female	**1f**	2	2	1	1	1	1	1	1	1	1
	2f	2	2	1	1	1	1	1	1	1	1
	3f	1	1	1	1	1	1	1	1	1	1
	4f	1	1	1	1	1	1	1	1	1	1
	5f	1	1	1	1	1	1	1	1	1	1
Male	**1m**	3	3	3	1	1	1	1	1	1	1
	2m	3	4	2	1	1	1	1	1	1	1
	3m	3	2	1	1	1	1	1	1	1	1
	4m	2	2	1	1	1	1	1	1	1	1
	5m	1	1	1	1	1	1	1	1	1	1

differences disappear fast and no difference is found in later stages. In the introduction we drew an analogy between behavior in the game we study, and the conduct of teams of employees in a firm. Our results do not suggest a reason why a team of men would be more or less productive than a team of women.

One limitation of our study is that we have not studied interaction in groups that have some other characteristic than gender in common. Croson, Marks & Snyder (2003) study a public goods game with multiple Pareto-ranked equilibria, and find that all female groups do better if the members belong to a sorority than if they are strangers, while all male groups do worse if the members belong to a fraternity than if they are strangers. Thus group identity helps women to coordinate and hurts men. Conceivably a group of females with "high social identity" could escape the bad equilibrium in van Huyck et al's game. We propose these matters for future research.

Finally, we raise a methodological point. The results reported in this paper are not "positive," in the sense that no difference in behavior between females and males was found. We think that in order to truly understand the differences in behavior between genders, one should not only report or publish experiments and results that show positive differences. Such a tack would bias perceptions about the magnitude and the limits of the differences. For example, from the results reported in Bolton & Katok (1995) we can learn that although others have found that females and males behave differently in some bargaining games, in generosity games no difference prevail.

ACKNOWLEDGMENT

We thank the participants at the Second Asian Conference on Experimental Business Research (Hong Kong University of Science & Technology, December 16–19, 2003) for helpful comments.

NOTES

[1] See Brown-Kruse & Hummels (1993) for an early example, and the studies reported in Andreoni & Vesterlund (2001), Ben-Ner, Kong & Putterman (2002), Croson & Buchan (1999), Bolton & Katok (1995), Dufwenberg & Muren (2002, 2004), Eckel & Grossman (1996, 1998, 2001), Fershtman & Gneezy (2001), Gneezy, Niederle & Rustichini (2003), Holm (2000).

[2] Van Huyck, Battalio & Beil's (1990) minimal effort coordination game; cf. section 2.

[3] Van Huyck et al. used groups of fourteen to sixteen participants. Smaller groups were used here since reducing the size of the group increases the probability of coordination success (cf. the discussion in Ochs).

REFERENCES

Andreoni, J. & L. Vesterlund (2001). "Which is the fair sex? Gender differences in altruism." *Quarterly Journal of Economics*, 116, 293–312.

Ben-Ner, A., F. Kong & L. Putterman (2002). "Share and share alike? Intelligence, socialization, personality, and gender-pairing as determinants of giving," unpublished manuscript, University of Minnesota, Peking University and Brown University.

Bolton, G. E. & E. Katok (1995). "An experimental test for gender differences in beneficent behavior." *Economics Letters*, 48, 287–292.

Brown-Kruse, J. & D. Hummels (1993). "Gender effects in laboratory public goods contribution." *Journal of Economic Behavior and Organization*, 22, 255–267.

Croson, R. & N. Buchan (1999). "Gender and Culture: International Experimental Evidence from Trust Games." *American Economic Review Papers and Proceedings*, 89, pp. 386–391.

Croson, R. M. Marks & J. Snyder (2003). "Groups work for women: Gènder and group identity in the provision of public goods." Working Paper, Wharton, University of Pennsylvania.

Dufwenberg, M. & A. Muren (2004). "Gender composition in Teams", Working Paper 04–08, Department of Economics, University of Arizona.

Eckel, C. C. & P. J. Grossman (1996). "The Relative Price of Fairness: Gender Differences in a Punishment Game." *Journal of Economic Behavior and Organization*, 30, 143–158.

Eckel, C. C. & P. J. Grossman (1998). "Are women less selfish than men? Evidence from dictator experiments." *Economic Journal*, 108, 726–35.

Eckel, C. C. & P. J. Grossman (2001). "Chivalry and solidarity in ultimatum games." *Economic Inquiry*, 39, 171–188.

Fershtman, C. & U. Gneezy (2001). "Discrimination in a segmented society: An experimental approach." *Quarterly Journal of Economics*, 351–377.

Gneezy, U., M. Niederle & A. Rustichini (2003). "Performance in Competitive Environments: Gender Differences." *Quarterly Journal of Economics,* August, 1049–1074.

Holm, H. J. (2000). "Gender based focal points." *Games and Economic Behavior*, 32, 292–314.

Lazaer, E. & S. Rosen (1990). "Male-Female Wage Differentials in Job Ladders." *Journal of Labor Economics*, 8, S106–S123.

McPherson, D. & B. Hirsch (1995). "Wages and Gender Composition: Why do Women's Jobs Pay Less?" *Journal of Labor Economics*, 13, 426–471.

Ochs, J. (1995). "Coordination Problems." In *Handbook of Experimental Economics*, A. Roth & J. Kagel (eds.), Princeton University Press, Princeton.

Van Huyck, J., R. Battalio & R. Beil (1990). "Tacit coordination games, strategic uncertainty, and coordination failure." *American Economic Review*, 80, 234–48.

Chapter 12

UPDATING THE REFERENCE LEVEL: EXPERIMENTAL EVIDENCE

Uri Gneezy
University of Chicago

Abstract

Empirical findings suggest that in decisions under uncertainty people evaluate outcomes relative to a reference level: they are risk seeking in the domain of losses and risk averse in the domain of gains. This finding is used in the finance literature to predict/explain the "disposition effect," which is the tendency of investors to sell assets that have gained value ("winners") too early and ride assets that have lost value ("losers") too long. The current experiment was designed to overcome some of the difficulties involved in using real market data to test the disposition effect. One of the main goals was to find evidence on how prior gains and losses influence the risk behavior of people, by shifting the reference level. The results were consistent with the disposition effect hypothesis. Furthermore, it was found that the data are best described by assuming that participants use the historical peak of the process as a reference level.

Keywords: reference level; disposition effect; experiments.
JEL: C91, D81, G11

1. INTRODUCTION

In the traditional theory of decisions under uncertainty (expected utility), utility is determined by final states of wealth. However, empirical evidence suggests that the behavior of individuals is best described by assuming that the significant carriers of utility are not states of wealth but changes relative to a reference level. Another central observation is that people tend to be risk averse when deciding on outcomes above their reference level and risk seeking below this level.[1]

Shefrin and Statman (1985) use these findings to investigate how prior gains and losses influence the risk behavior of people in a dynamic setup. In particular, they find support for the *disposition effect* in financial markets, which is the observation that "investors tend to sell winners too early and ride losers too long" (p. 778).[2] Further support for these findings is reported in Lakonishok and Smidt (1986) and Ferris, Haugen, and Makhija (1988).

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 263–284.
© 2005 *Springer. Printed in the Netherlands.*

The need for experimental evidence arises because it is difficult to test the disposition effect using real market data. Two important exceptions are, first Odean (1998), who tested the disposition effect by analyzing trading records for 10,000 accounts at a large discount brokerage house. Odean found support to the disposition hypothesis using the buying price as a reference level. Second, Heath, Huddart and Lang (1999) study stock option exercise decisions by over 50,000 employees at seven corporations. They find that reference levels depen on the extreme values in the past. In particular, employee exercise activity roughly doubles when the stock price exceeds the maximum price attained during the previous year.

Yet, Very little experimental evidence for the presence of the disposition effect exists. (See, however, the study of Weber and Camerer, 1998, described below.) Using experiments to complement the real markets data is important, because in real markets it is impossible to control for investors' expectations, to observe individual decisions, etc. Using experimental methods enables us to follow an individual's decisions, as well as the history that led to these decisions (see e.g., Rapoport, 1984 and Rapoport, Zwick and Funk, 1988).

In the experiment reported in this paper, participants were endowed with a single asset with a price that follows a random walk. What I observed was the decision to sell this asset, and the history up to that decision. The two questions I investigated were: (i) does the disposition effect prediction outperform the "rational" expected utility prediction? and (ii) what is a loss?

Answering the second question is important because different assumptions about the way gains and losses are coded may result in different predictions of this model: when testing for disposition effect, one actually tests the joint hypothesis of risk attitudes (risk seeking for gains and risk aversion for losses) and a specific reference level formation. Better understanding of the way reference levels are formed may improve our understanding of behavior in markets and increase the descriptive power of the models.

The decision problem is described in section 2. The behavior of an expected utility maximizer is derived in section 3, followed by an analysis of the disposition effect in section 4. In section 5 the experimental design and procedure are described. The results are presented in section 6, and section 7 concludes.

2. THE DECISION PROBLEM

The decision problem used in the experiment is as follows:

> An investor is endowed with a stock whose price is $\$x_0$. Every period the price of the stock either goes up or down by \$1, with probability p and $1 - p$ respectively. The investor may sell her stock at any time, but then she cannot reenter the trade in it. If the investor does not sell before the price reaches either \$0 or \$N the trade will end and she will get the respective price for the stock.

The decision problem, with $p = .6$ and $x_0 = 5$, is presented in Figure 1.[3] I chose to use a random walk process with a positive drift because the traditional finance

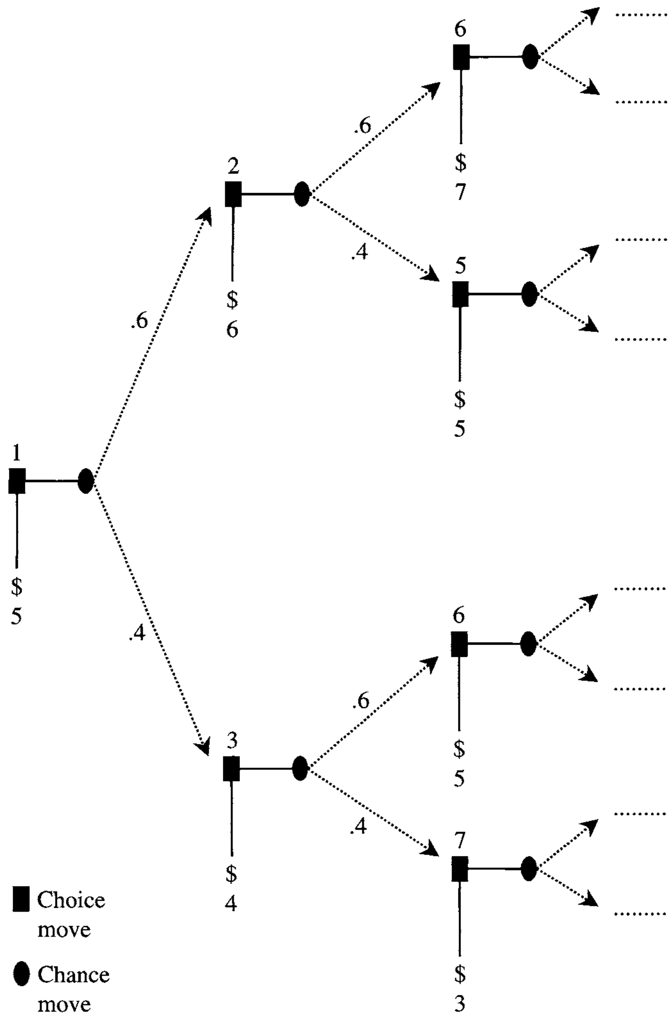


Figure 1. The decision problem.

literature assumes that asset prices in an efficient market follow such a type of processes (Brealey and Myers, 1988).[4]

3. THE BEHAVIOR OF A SUBJECTIVE EXPECTED UTILITY MAXIMIZER

I use Savage's (1954) subjective expected utility (SEU) theory of choice under uncertainty as a benchmark for decision theories. It is commonly accepted that SEU is a good normative theory, so finding the behavior of an SEU maximizer is important. It is also commonly accepted that this theory is not a good descriptive theory, so if actual behavior in the experiment does not coincide with this prediction, one should not be surprised (see e.g., Edwards, 1992).

I assume the standard assumptions about SEU: Let X be the set of all dollar prizes, $x \in X$. $\succ$ is a complete and transitive relation on X. $\exists U$ representing $\succ$, such that $U: X \to R$ and U is strictly increasing. The SEU representation is then: let $p = (p_0, \ldots, p_N)$, $p \succ q$ if and only if $\sum p_x u(x) \geq \sum q_x u(x)$.

An SEU maximizer, i.e., a person who behaves as if she is maximizing her SEU, will make a contingent plan at the beginning of the process, indicating what she will choose whenever a choice is to be made. Using the reduction of compound lotteries axiom, stating that a multi-stage lottery is equally as attractive as the one stage lottery that yields the same prizes with the corresponding multiplied probabilities, the plan can be reduced to a one-stage lottery.[5] The decision-maker has a preference relation over one-stage lotteries, which she uses in order to choose an optimal plan, i.e., a plan that corresponds to the most preferred one-stage lottery implied by the process. The above assumptions also imply that the decision maker is only interested in the ultimate outcomes of the decisions, in particular, the assessment of a strategy is independent of the history of the decision process up to that level. This property of the preference relation is called consequentialism (Hammond, 1988, Machina, 1989).

3.1. A strategy

Any contingent plan a decision-maker uses is called a strategy. There are in general infinitely many strategies. One class of strategies is stationary strategies, which associates a unique choice with each stock price x_t (x_t denotes the price of the stock at period t), independently of the stock's price history up to period t. For example, nodes 1, 5, and 6 of figure 1 correspond to the same stock price, and a stationary strategy would prescribe the same choice at all these nodes.

Claim 1: In the above process, an SEU maximizer uses only stationary strategies.
Proof: See Appendix.

3.2. The value of the process

Define the value of the process as follows: let σ be a stationary strategy. $p(x_T; x_t, \sigma)$ is the probability of price x_T at time T (when T is the time at which the process ends, and may be infinite), starting at time t with x_t and using σ. The value of the process is

$$v(x_t, \sigma) = \sum_{X_T} p(x_T; x_t, \sigma) u(x_T). \tag{3.21}$$

The existence of a value is implied by the fact that every strategy results in a well defined probability distribution over final prices. The optimal value is defined as $v(x_t, \sigma^*)$, when σ^* is an optimal strategy which maximizes 3.21. Formally,

$$v(x_t, \sigma^*) = \max_\sigma v(x_t, \sigma), \tag{3.22}$$

where the maximum is taken over all conceivable strategies σ. A strategy σ^* is said to be optimal if

$$v(x_t, \sigma) = v(x_t, \sigma^*) \quad \text{for all } 0 < x_t < N. \tag{3.23}$$

To find σ^*, note that if σ^* is an optimal strategy, then

(1) if $u(x_t) < pv(x_t - 1, \sigma^*) + (1 - p)v(x_t + 1, \sigma^*)$ then $\sigma^*(x_t) = Continue$ [$\sigma^*(x_t)$ is the choice implied by the optimal strategy at x_t], and
(2) if $u(x_t) = pv(x_t - 1, \sigma^*) + (1 - p)v(x_t + 1, \sigma^*)$ then $\sigma^*(x_t) = Sell$.

This is true because if the strategy says *Sell*, the process ends and its value is $u(x_t)$. In case (1), the decision-maker can increase the value by choosing *Continue* so *Sell* is not optimal. In case (2), choosing *Continue* will yield lower value than $u(x_t)$, so *Continue* cannot be optimal. From this, if σ^* is optimal, then $v(x_t, \sigma^*) = \max\{u(x_t), pv(x_t - 1, \sigma^*) + (1 - p)v(x_t + 1, \sigma^*)\}$. Since this is true for every x_t, and since the strategy is stationary, it must be true for every period. So the optimal value given a stock price x_t, denoted by $v^*(x_t)$ is given by

$$\begin{aligned}
&v^*(N) = u(N)\\
&\vdots\\
&v^*(x_t) = \max\{u(x_t), pv^*(x_t - 1, \sigma^*) + (1 - p)v^*(x_t + 1, \sigma^*)\} \quad \text{for } 0 < x_t < N,\\
&\vdots\\
&v^*(0) = u(0).
\end{aligned}$$

Given $v^*(x_t)$ for all x_t, an equivalent way to choose is applying the following algorithm in every period:

$$\begin{aligned}
&\text{if} && u(x_t) < v^*(x_t) \text{ then } Continue,\\
&\text{and if} && u(x_t) = v^*(x_t) \text{ then } Sell.
\end{aligned}$$

We have now found the class of strategies to be used and the way the strategies are valued. For a given utility function, it is possible by now to find an optimal strategy by solving the system of v. The result of adding the commonly used assumption of non-increasing absolute risk aversion on the utility function is now considered.

3.3. Non-increasing absolute risk aversion

In the literature, the assumption of NIARA is commonly accepted: see e.g., Arrow (1971, p. 96) and Kreps (1988, p. 88). The assumption is also consistent with some empirical findings, such as Binswanger (1981). Assume that u satisfies non-increasing absolute risk aversion (NIARA): let $0 \le \alpha \le 1$ and f, g be monetary prizes. Denote the lottery that yields $(x - f)$ with probability α, and $(x + g)$ with probability $1 - \alpha$ by $[(x - f), \alpha; (x + g), 1 - \alpha]$. The assumption states that[6]

$$\begin{array}{ll} \text{if} & [(x_t - f), \alpha; (x_t + g), 1 - \alpha] < x_t, \\ \text{then for} & x_t < x_s, [(x_s - f), \alpha; (x_s + g), 1 - \alpha] < x_s. \end{array} \quad (3.31)$$

The next proposition gives the general structure of the optimal strategy under the above assumptions.

Proposition 1: For an SEU maximizer with a utility function which satisfies the NIARA assumption, if $v(x_t) > u(x_t)$ and $x_t \leq x_s < N$, then $v(x_s) > u(x_s)$.
Proof: See Appendix.

In hypothesis 2 below the behavior implied by proposition 1 is described.

3.4. *Testing SEU*

We have two hypotheses regarding the behavior of an SEU maximizer. The first is the stationary one in which no extra assumptions about the utility function are made, so the hypothesis is very general.

Hypothesis 1:
The decision-maker uses only stationary strategies.

The observations we have are about the choice to *Sell*; we say that observed behavior is inconsistent with SEU if people choose *Sell* at a price at which they chose *Continue* before.

In the second hypothesis we put some more structure into the utility function. We state the hypothesis of an SEU maximizer with a utility function that satisfies non-increasing absolute risk aversion.

Hypothesis 2:
If the decision-maker chooses *Continue* for x_t, then for x_s, such that $x_t \leq x_s < N$, she also chooses *Continue*.

We say that observed behavior is inconsistent with SEU+NIARA if people choose *Sell* at a price lower than or equal to a price at which they chose *Continue* before. Note that the first hypothesis is more general because in cases in which we can reject the second hypothesis we can also reject the first one, but the reverse does not necessarily hold.

4. BEHAVIOR ACCORDING TO THE DISPOSITION EFFECT

One of the major assumptions in Kahneman and Tversky's (1979) prospect theory is that the significant carriers of utility are changes relative to reference level. (This assumption is not new in economics; see e.g., Markowitz, 1952, for an earlier study.) Moreover, it is assumed that risk attitudes reflect around this reference level, with people being risk averse when deciding about outcomes higher than their reference level (concave value function), and risk seeking when deciding about outcomes lower than their reference level (convex value function). This pattern of risk behavior is supported by empirical evidence: see Fishburn and Kochenberger (1979), Hershey, Kunreuther, and Schoemaker (1982), Kahneman and Tversky (1979), Payne, Laughhunn, and Crum (1980), Thaler and Johnson (1990), Tversky and Kahneman (1992), and the references therein.

Shefrin and Statman (1985) use these findings to predict that people will tend to sell assets that have gained value and keep assets that have lost value. They call this the disposition effect. For example, in the dynamic decision problem discussed in this paper, assume that the decision maker uses the initial price as a reference level. After losing in the first period, she will be in the domain of losses, and hence will be risk seeking – and will not sell the stock which has positive expected value. On the other hand, after winning she will be in the domain of gains, hence she will be risk averse – and may sell the asset. For an example and a more thorough discussion see Weber and Camerer (1998).

But what is the reference level? Any test of the disposition effect tests the joint hypothesis of this pattern of risk attitudes and a particular reference level. In prospect theory, Kahneman and Tversky (1979) consider static choice problems. They analyze the case in which people use the current asset position as a reference level, and add that although this is probably true for most choice problems, there are situations in which gains and losses are coded relative to an expectation or aspiration level that differs from the status quo. They conclude that the location of the reference level, and the manner in which choice problems are coded and edited, emerge as critical factors in the analysis of decisions.

While empirical evidence for reference levels in static choice problems exists, very little is known on how the history of a process influences the reference level. Weber and Camerer (1998) designed an experiment to investigate whether participants would exhibit the disposition effect in laboratory markets. They considered the volume of trade, and found that participants tended to sell fewer shares after the price falls than after it rises. However, they did not investigate the way reference levels are formed, and they stressed the need for this to be done: "... we need to know more about how reference points adapt over time and how multiple reference points are balanced" (p. 182).

Thaler and Johnson (1990) argue that most decision makers are influenced by prior gains and losses, which influence subsequent choices in systematic ways. They ran an experiment with a two-periods setup, in which the first-period gains or losses were exogenously determined. One of their findings was that participants were willing to take great risks when it gave them the chance to "break even" and compensate them for past losses. This finding is of special interest for the current paper since, after losing, it is possible for the decision-maker to "break even" and compensate herself for past losses (i.e., the price of the stock may go up again from any price).

Given these findings, it is expected that participants will not sell the stock with less than their reference level, that is, participants are expected to be reluctant to realize losses. So what one should seek is a level, and the history of the process up to that level, at which participants will choose *Sell*. The first hypothesis in this section is that people are exposed to the disposition effect, and use the initial purchase price as a reference level (this is the assumption made by Shefrin and Statman, 1985).

Hypothesis 3:
The decision-maker will not sell the stock when the price is lower than the initial price.

The second hypothesis of this section is that people use the highest price (the peak) of the process as a reference level.

Hypothesis 4:
The decision-maker will not sell the stock when the price is lower than the historical peak.

Note that this hypothesis is a refinement of hypothesis 3, in the sense that any observation that is consistent with hypothesis 4 is also consistent with hypothesis 3, but the reverse does not hold. As a result, this hypothesis narrows the set of periods in which the theory predicts that a decision to sell may occur.

It is not correct to say that hypothesis 4 describes the behavior of a person who follows prospect theory. The analysis according to prospect theory is much more complex since, if the reference level can change during the process, the decision maker must take this into account in her maximization problem. The hypotheses I use are based on a more "primitive" assumption about behavior, namely that people are risk seeking for losses and risk averse for gains. I do not assume any "value" or "utility" maximization.

5. DESIGN AND PROCEDURE

I start by considering two properties of the design that are relevant for its "attractiveness." The first is the reduced probabilities of earning money, and the second is the duration of the process.

The probability of success: Obviously the process is not a trivial dynamic decision-making problem, in the sense that calculating the probability distribution over prizes implied by continuing more than three or four periods is a complicated task. This implies bounded rationality because decision-makers are not capable of using all the available information correctly. An interesting datum is the reduced probabilities for the extreme case, in which investors do not sell their stocks before reaching \$0 or \$$N$. Denote by p_i the reduced probability of reaching \$$N$ in this case, when starting with $i \in \{0, 1, \ldots, N\}$. Table 1 gives the reduced probabilities for $N = 10$ and $p = .6$.[7]

The duration of the process: It is possible to construct paths of the process such that it will never end. For example, a path in which the investor does not sell until she has to (reaching \$0 or \$$N$), and chance moves are: win, lose, win, lose, etc. The

Table 1. The reduced probability of reaching N, when N = $10 and p = .6 from different starting prices, for a decision-maker who never chooses Sell

p_0	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}
0	.34	.57	.72	.82	.88	.93	.96	.98	.99	1

probability of such paths, however, converges to 0 (see Gneezy, 1996). That is, with probability 1 the process will end in a finite number of periods. But how fast? In particular, how long will the experiment take? A related problem is to convince participants that they are not being cheated. If the duration is unlimited, some decision-makers may not believe (and rightly so) that in any case we are going to continue the entire process. For these two reasons the duration of the process was limited to 100 periods. The probability that for $p = .6$, $N = 10$, and starting from any price, the process will not end within 100 periods is less than .001. (See the Appendix for calculations.)

5.1. Experimental procedure

The parameters used in the experiment are $N = 10$ and $p = .6$. Three experimental treatments were run, with a starting price of 3, 5, and 7 in treatments 1, 2, and 3 respectively.[8] The reason for using three different starting prices instead of just one is to enable us to test hypotheses regarding the formation of the reference levels for different histories.

Sessions 1, 2, and 3 were of treatment 1, sessions 4 to 8 of treatment 2, and sessions 9 and 10 of treatment 3. In each session ten different students participated. The experiment was administered by pen and paper, and held in a seminar room with participants seated far apart. Participants were undergraduate students in economics, recruited in their classes at the University of Haifa. For each session twelve students were invited; ten would participate in the process, one would act as an assistant, and one would serve as a spare in case of a no show.

Upon entering the room, a short standard-type introduction was read to the participants by the experimenter. After the introduction, each participant drew an envelope out of a stack. Ten envelopes contained numbered cards; one envelope contained a note with "assistant," and one had an empty note (the latter envelope was removed when only eleven students showed up). The assistant was told that he would receive a payment equal to the average earnings of the other participants. The student who drew the empty note was paid NIS 20 for showing up and was asked to leave the room.

Instructions were distributed and read aloud. After that, participants could examine the instructions for a few additional minutes, and (privately) ask questions.

Participants were then asked to choose whether to *Continue* or *Sell* in period one. The choice was recorded on a decision form circulated among the participants by the experimenter, such that each participant could not see the choices of the others.

The lottery was conducted by the assistant. To determine whether the price of the stock increased or decreased by 1 in a period, we used a ten-sided die. At the beginning of the experiment, six participants were asked to choose (sequentially) a "winning number," so we had six winning numbers (subjects could not choose a previously chosen number). These numbers were written on the blackboard. After the participants made their choices for the period, the assistant rolled the die. If the number that came up was one of the six winning numbers, the price of the stock increased by 1, otherwise it decreased by 1.

After each period, the new price of the stock was written on the blackboard, next to the history of prices. Then participants made their choices for the next period. Note that once a participant chose *Sell*, this was her choice for the rest of the experiment. The process continued this way until the price of the stock reached either 0 or 10.

At the end of the experiment all participants were paid according to their performance.

6. RESULTS

The results of the experiment are presented in Figure 2. A summary of the decisions to sell as well as the consistency of these decisions with each of the four hypotheses is presented in Table 2. Consider for example session 3 (Figure 2). A decision to sell at the first period when the price is 3 is consistent with all four hypotheses. In period 1 the price of the stock went down to 2. At this period a decision to sell is consistent with hypothesis 1, since there was no previous period at which the price was 2 and the decision-maker decided to continue. It is also consistent with hypothesis 2 since

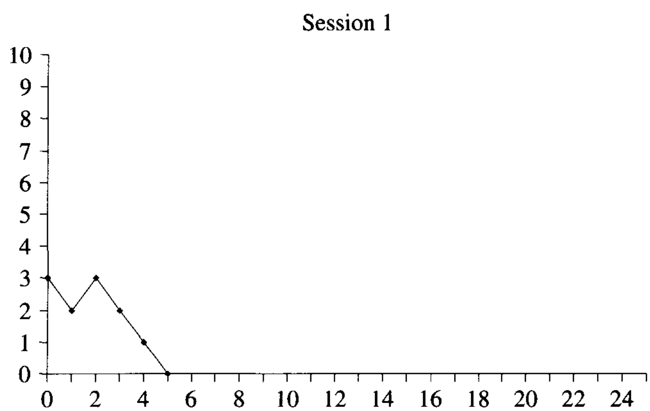


Figure 2. Results.
The processes of the 10 different sessions of the experiment. The vertical axes indicate the price of the stock, and the horizontal axes the stage number. The numbers on the graph represent the decision to Sell. For example, in session 2, three participants chose Sell at period 14, and the price of the stock was 7 at that stage.

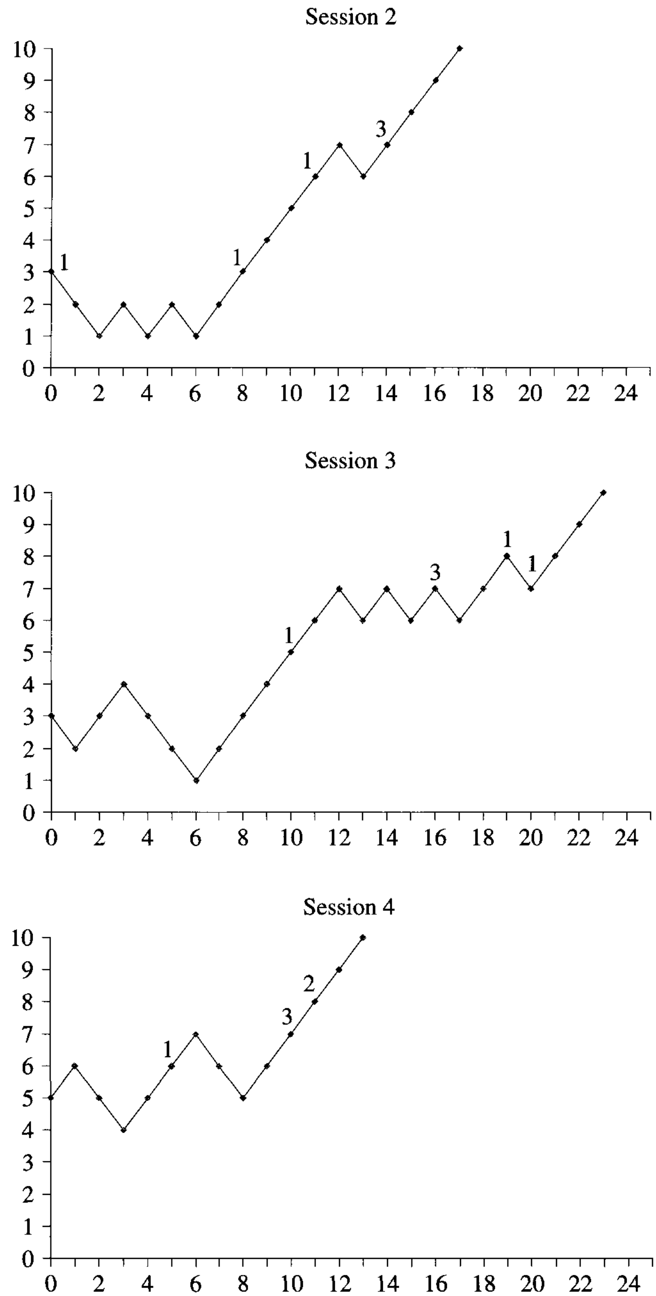


Figure 2. (cont'd)

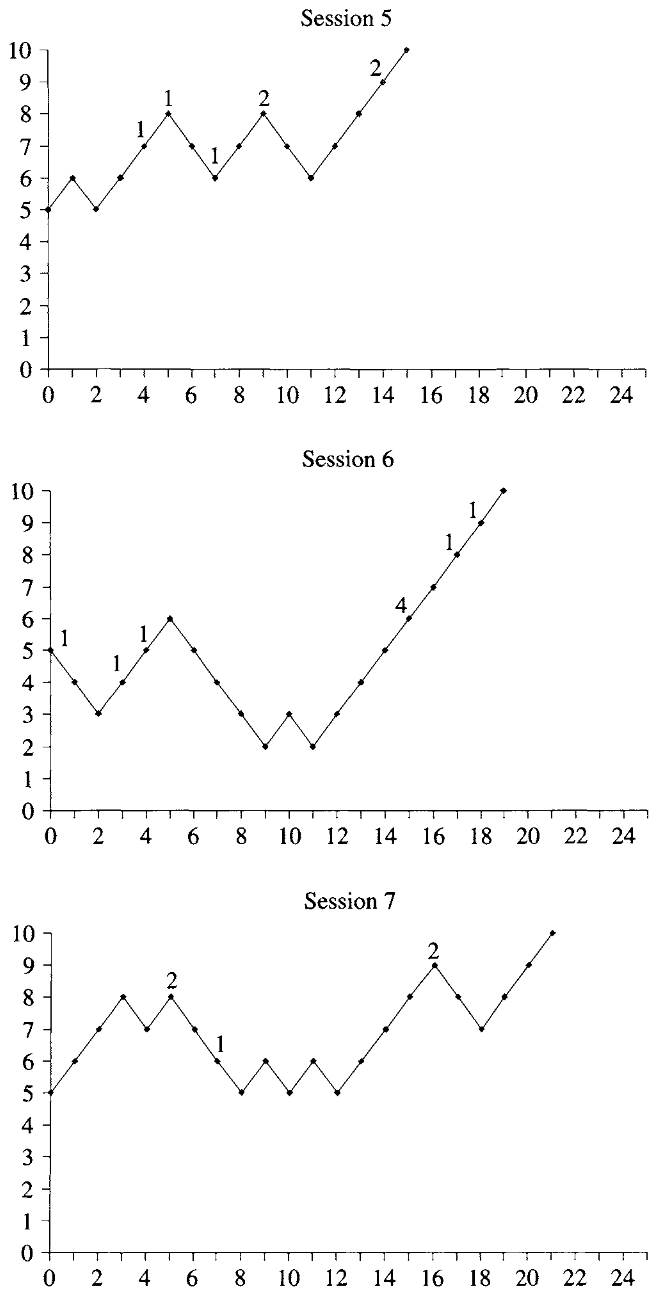


Figure 2. (cont'd)

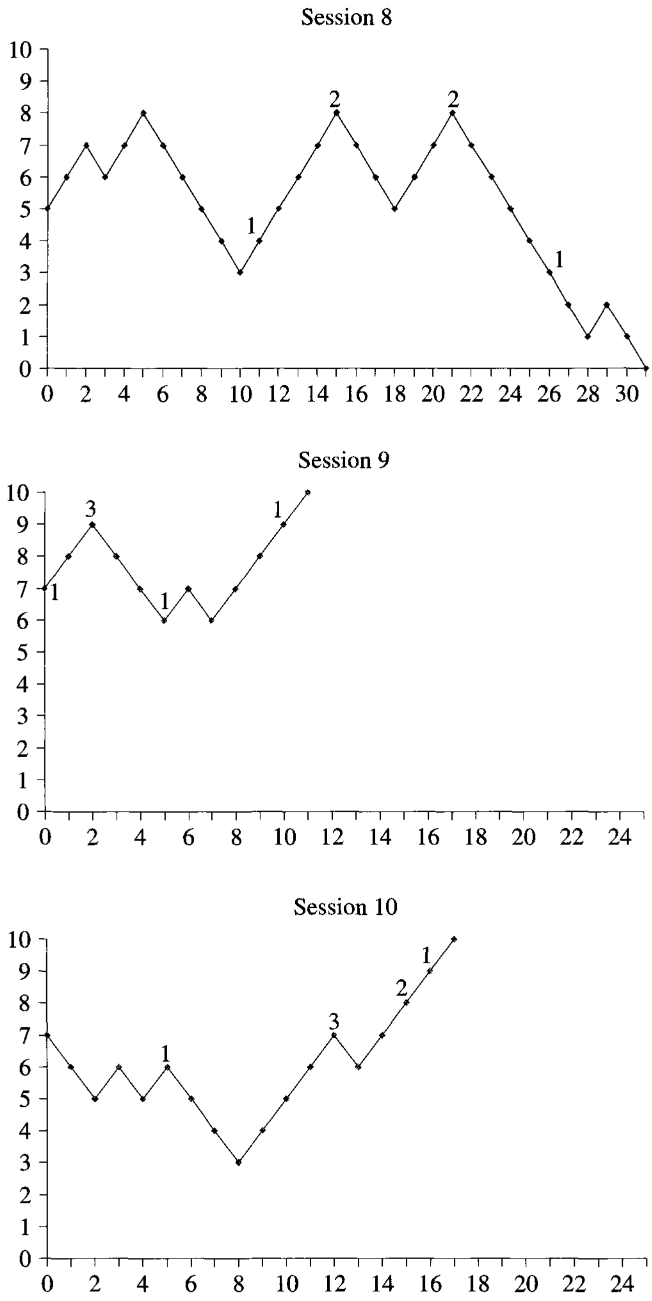


Figure 2. (cont'd)

Table 2. The number of observations that are consistent with each of the four hypotheses, according to sessions

Initial Price	*Session*	*# of Sell*	*Consistent with Hyp.1*	*Consistent with Hyp.2*	*Consistent with Hyp.3*	*Consistent with Hyp.4*
3	1	0	0	0	0	0
3	2	6	2	0	6	6
3	3	6	2	0	6	5
5	4	6	2	0	6	6
5	5	7	4	0	7	6
5	6	8	3	0	7	7
5	7	5	2	0	5	4
5	8	6	0	0	4	4
7	9	6	6	1	5	5
7	10	7	3	0	6	6
Total	57	24	1	52	49	

in no period before was the price of the stock lower than or equal to 2. However, a decision to sell at this period is inconsistent with hypothesis 3, since the initial price 3 is higher than the current price 2. Since the requirements of hypothesis 3 are not met, the decision to sell at this period cannot be consistent with hypothesis 4. In period 2, the price of the stock rose back to 3. A decision to sell at this period is inconsistent with hypothesis 1, since the decision-maker already decided once (at period 1) to continue when the price was 3. The requirements of hypothesis 2, then, must also be violated by a decision to sell at this period. However, a decision to sell at this period is consistent with hypothesis 3 (4) since the price is not lower than the initial price (the highest price in the process) of 3. If we jump now to period 4, we see, for the same reasons as in period 2, that a decision to sell is inconsistent with hypotheses 1 and 2, and is consistent with hypothesis 3. However, unlike period 2, a decision to sell at this period with a price of 3 is inconsistent with hypothesis 4 since the highest price in the process up to that point was 4. If we look further into the process, we see that at period 10 one decision-maker decided to sell when the price was 5. This decision is consistent with hypothesis 1 (since this is the first time in that process that the price was 5), with hypothesis 3 (since the price is higher than the initial price of 3), and with hypothesis 4 (since the price is higher than the peak of process up to the period which is 4). However, the decision to sell at period 10 is inconsistent with hypothesis 2 since the decision-maker decided to continue before with a price lower than 5. Further on we see that in period 16 three participants decided to sell (a decision which is inconsistent with hypotheses 1 and 2 but consistent with hypotheses 3 and 4), and one decision-maker chose to sell in period 19 when the price was 8 (consistent with all hypotheses except hypothesis 2). Finally, one decision-maker decided to sell in period 20, when the price was 7. This decision is inconsistent with all four hypotheses.

More generally, I say that a decision to sell is (in)consistent with a hypothesis whenever the history of the process up to that decision is (in)consistent with the requirements of the hypothesis as stated above. In Table 2, the first column on the left indicates the initial price in the session, the second indicates the number of the session, and the third indicates the number of participants who chose *Sell*. The other four columns indicate how many of these decisions to sell where consistent with each of the four hypotheses.

The benchmark SEU hypotheses (hypotheses 1 and 2) find little support in the data. Only 24 out of the 57 (42%) observations are consistent with the stationary hypothesis. When considering SEU plus the assumption of non-increasing absolute risk aversion, only one observation is consistent with the data. As argued above, this result is not surprising given the accumulated empirical evidence against expected utility theory (Edwards, 1992).

On the other hand, the disposition effect hypothesis, jointly with the assumption that the initial price is the reference level, finds strong support. In 52 out of the 57 observations (91%) the results are consistent with hypothesis 3. Hypothesis 4 is consistent with the data in 49 observations out of the 57 (86%).

In measuring the predictive success of each hypothesis I use the theory introduced in Selten and Krischker (1983) and analyzed in Selten (1991). Selten (1991) defines an *area theory* as a theory that predicts a subset of all possible outcomes. The *hit rate* of a theory is defined as the relative frequency of correct predictions. Selten argues that the hit rate is a measure of accuracy, but accuracy alone cannot be the aim of the area theory since, for example, no area theory can be more accurate than the one that simply predicts the set of all possible outcomes. This theory never fails to predict correctly, but it is useless in view of its complete lack of precision. The precision of an area theory is related to the set of its predicted outcomes. In the current experiment, I say that a hypothesis is more precise as its set of predicted outcomes is smaller. For example, hypothesis 2 is more precise than hypothesis 1.

The relative size of the subset predicted by a hypothesis is called the area of this hypothesis.

I use simulation techniques to find the area of the hypotheses.

The simulation: In each run of the simulation the process was started with a stock prices of 3, 5, and 7, as in the experiment. The process was stopped randomly, with a probability of .1, in each period of the process. That is, the computer chose randomly when to stop the process. After each run was finished, the consistency

Table 3. The area of each hypothesis according to the starting price. The simulation results indicate the relative size of the subset predicted by each of the four hypotheses (the area of the hypotheses)

Starting price	*The area of Hyp.1*	*The area of Hyp.2*	*The area of Hyp.3*	*The area of Hyp.4*
3	.52	.22	.84	.56
5	.50	.25	.77	.52
7	.53	.31	.70	.52

Table 4. The measure of predictive success of the four hypotheses according to starting price

Starting price	*Predictive success of Hyp.1*	*Predictive success of Hyp.2*	*Predictive success of Hyp.3*	*Predictive success of Hyp.4*
3	–.19	–.22	.16	.36
5	–.08	–.25	.19	.36
7	–.06	–.26	.09	.27

with hypotheses 1 to 4 was checked. For each starting level 10,000 repetitions of the simulation were made. The results of the simulations are presented in Table 3.[9]

From Table 3 we learn that hypothesis 2 is the most precise one, and that hypothesis 3 is the least precise one. Hypothesis 1 and hypothesis 4 are of similar precision.

The measure of predictive success developed by Selten and Krischker (1983), and later investigated in a deeper way by Selten (1991), can be described as follows:

$$m = r - a \qquad (6.1)$$

where m = measure of predictive success, r = hit rate (the relative frequency of correct predictions in the experiment), and a = the area (the relative size of the predicted subset as found in the simulation, compared with the set of all possible outcomes).[10] The measures of predictive success in the experiment are presented in Table 4.

As can be seen in Table 4, for hypotheses 1 and 2, the measures of predictive success implies that expected utility theory is not a good descriptive theory for the current experiment. However, this is not the case for hypotheses 3 and 4. Moreover, according to the measure of success used, hypothesis 4 performed better than hypothesis 3. The measure of success shows that by using the prediction of the disposition effect with the peak of the process as a reference level, we get a more precise theory than when assuming that the initial price is the reference level.

This suggests that using hypothesis 4 clearly increases the predictive power of the model relative to the SEU theory, and even relatively to the prediction of the disposition effect with the initial price as the reference level.

7. CONCLUDING REMARKS

The purpose of the experiment described was to investigate the influence of prior gains and losses on the risk attitude of people. Unlike the case of real market data, the stylized experimental setup allows us to have some insight into the decision-making process of individuals. Furthermore, using a stylized decision problem makes the benchmark prediction very clear and testable.

It turns out that prior gains and losses do influence the risk attitude, and in a different way from that predicted by the rational theory (expected utility). The disposition effect prediction that people will be reluctant to sell losing assets found strong empirical support with the traditional assumption that the reference level is the initial purchase price of the stock. This finding supports the empirical research done on real market data. The use of a stylized process also allows for more refined tests about the way reference levels are formed. In particular, it is possible to learn about how it depends on the history of gains and losses. This is important because, for example, prospect theory is useless as a descriptive theory without a "good" assumption about the reference levels. It was found that when the peak of the process was used as a reference level, the descriptive power of the theory increased dramatically.

Of course, the use of stylized experiments has many drawbacks. Real market decisions are taken over a longer horizon; the money involved is much more substantial; markets can correct many individual irrationalities, etc. This leads to the challenge of the findings in this paper: testing the disposition effect using real market data and the refined reference level assumption. For example, in an intriguing empirical work, Odean (1998) tested the disposition effect by analyzing trading records for 10,000 accounts at a large discount brokerage house. Odean found support to the disposition hypothesis using the buying price as a reference level. Given the results presented in the current research, the empirical work of Odean might have found even a stronger effect by using the pick of the process as a reference level. I believe that this is an example of the importance of combining empirical work from real market data with stylized laboratory experiments.

Finally, the findings in this paper can explain the following story which was told to me by a friend: "My father invested some money in the stock market when the index was 190. A few months later the index jumped to 240, and then, after some time, fell to 200. Knowing my father I urged him to sell his position, telling him that the market was too risky at the moment. My father replied: 'To sell now, after I *lost* 40 points? No way!'"

ACKNOWLEDGMENT

I thank Eric van Damme for many helpful comments and suggestions as well as for providing the access to the necessary funds, and Martin Dufwenberg, Werner Guth, and Jan Potters for their comments. E-mail: ugneezy@gsb.uchicago.edu

NOTES

1. Kahneman and Tversky (1979). See Kahneman, Knetsch, and Thaler (1991) for a survey of the literature, and the discussion below.
2. As a benchmark theory they used the tax-loss-selling hypothesis of Constantinides (1983, 1984).
3. A similar process, known as "the gambler's ruin problem," is a classical problem in the random walk literature. For an elaborated discussion of this process and early solutions by Bernoulli and De Moivre see Thatcher (1957). For a detailed analysis see Ross (1989).
4. Though the random walk approach is controversial nowadays (e.g., Fama, 1991, De Bondt and Thaler, 1994), it is still accepted as a first approximation.
5. The reduction axiom is not an extra assumption to the SEU assumptions; it is implied by them. See Karni and Schmeidler (1991) and Volij (1994).
6. An equivalent statement is the Arrow-Pratt measure of risk aversion; for a twice differentiable $u(x_t)$, NIARA is equivalent to the statement that

$$\lambda(x_t) = -u''(x_t)/u'(x_t)$$

does not increase with x_t. See Pratt (1964) and Arrow (1965).
7. For the analytical way of getting the results presented in table 1, see Gneezy (1996) which also presents evidence that people systematically underestimate the probability of success. It is argued there that people "anchor" to p and, since with $p > .5$ it holds that $p < p_x$, they systematically underestimate p_x.
8. The money used was New Israel Shekels (NIS), with N = NIS 100 and each step was of NIS 10. At the time of the experiment, NIS 100 = \$30. We keep reporting the results in dollar units and say that $N = 10$ for convenience.
9. Note that the choice of the probability at which the process stops at each period may influence the results of the simulations. For example, the shorter the process (the larger the probability that the process will stop at each period), the higher is the chance that the stationary hypothesis will be consistent with the decision to sell. The .1 probability was chosen because the average length of the process with this parameter is similar to what was observed in the real experiment. Moreover, it turns out that the consistency with the hypotheses is not very sensitive to this probability. In particular, varying the probabilities between .05 and .15 does not change the results qualitatively.
10. Other measures of success which have been used in the literature, such as $m = (r - a)/(1 - a)$, are found by Selten (1991) to be inferior to (6.1).
11. Note that (A3) does not claim that $[x(u) + \Pi]$ is the best strategy at $x(u)$ (in fact it is not). It just shows that *Continue* is better than *Sell*.

REFERENCES

Arrow, K. J. (1965). *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrjo Hahnsson Foundation.

Arrow, K. J. (1971). *Essays in the Theory of Risk-Bearing*, Amsterdam: North-Holland.

Binswanger, H. P. (1981). "Attitudes toward Risk: Theoretical Implications of an Experiment in Rural India." *Economic Journal*, 91, 867–90.

Brealey, R. A., and S. C. Myers (1988). *Principles of Corporate Finance*, New York: McGraw-Hill.

Constantinides, G. M. (1983). "Capital market equilibrium with personal tax." *Econometrica*, 51, 611–636.

Constantinides, G. M. (1984). "Optimal stock trading with personal taxes: Implications for prices and the abnormal January effect." *Journal of Financial Economics*, 13, 65–69.

De Bondt, W. F. M., and R. H. Thaler (1994). "Financial decision making in markets and firms: A behavioral perspective." *NBER* Working Paper 4777.

Dufwenberg, M. & A. Muren (2002). "Discrimination by gender and social distance", Working Paper 2002:2, Department of Economics, Stockholm University.

Edwards, W. (1992). "Toward the demise of economic man and woman: Bottom lines from Santa Cruz." in *Utility theories: Measurements and applications*. Ed Ward. Edwards. Kluwer Academic Publishers, 253–70.

Fama, E. F. (1991). "Efficient capital markets: II." *Journal of Finance*, 46, 5, 1575–1617.

Ferris, S. P., R. A. Haugen, and A. K. Makhija (1988). "Predicting contemporary volume with historic volume at differential price levels: Evidence supporting the disposition effect." *Journal of Finance*, 43, 677–699.

Fishburn, P. C., and G. A. Kochenberger (1979). "Two-piece von Neumann-Morgenstern utility functions." *Decision Science*, 10, 503–18.

Gneezy, U. (1996). "Probability judgment in multi-stage problems: Experimental evidence of systematic biases." *Acta Psychologica*, 93, 59–68.

Hammond, P. J. (1988). "Consequentialism and the independence axiom," in *Risk, Decision and Rationality*. Ed. Bertrand Munier. Dordrecht, Holland: D. Reidel Publishing Co., 503–15.

Heath, C., S. Huddart, and M. Lang (1999). "Psychological factors and stock option exercise." *Quarterly Journal of Economics*, 114, 2, 601–627.

Hershey, J. C., H. C. Kunreuther, and P. J. H. Schoemaker (1982). "Sources of bias in assessment procedures for utility functions." *Management Science*, 28, 936–954.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1991). "Anomalies: The endowment effect, loss aversion, and status quo bias." *Journal of Economic Perspectives*, 5, 193–206.

Kahneman, D., and A. Tversky (1979). "Prospect theory: An analysis of decision under risk." *Econometrica*, 47, 263–91.

Karni, E., and D. Schmeidler (1991). "Atemporal dynamic consistency and expected utility theory." *Journal of Economic Theory*, 54, 401–408.

Kreps, D. M. (1988). *Notes on the Theory of Choice*, Boulder, Co: Westview Press.

Lakonishok, J., and S. Smidt (1986). "Capital gain taxation and volume of trading." *Journal of Finance*, 41, 951–974.

Machina, M. J. (1989). "Dynamic consistency and non-expected utility model of choice under uncertainty." *Journal of Economic Literature*, 1622–1668.

Markowitz, H. (1952). "The utility of wealth." *Journal of Political Economy*, 60, 151–8

Payne, J. W., D. J. Laughhunn, and R. Crum (1980). "Translation of gambles and aspiration level effects in risky choice behavior." *Management Science*, 26, 1039–1060.

Odean, T. (1998). "Are investors reluctant to realize rgeir losses?" *Journal of Finance*, 53, 1775–1798.

Pratt, J. W. (1964). "Risk aversion in the small and in the large." *Econometrica*, 32, 122–136.

Rapoport, A., R. Zwick, and S. G. Funk (1988). "Selection of portfolios with risky and riskless assets: Experimental tests of two expected utility models." *Journal of Economic Psychology*, 9, 169–194.

Rapoport, A. (1984). "Effects of wealth on portfolios under various investment conditions." *Acta Psychologica*, 55, 31–51.

Ross, S. M. (1989). *Introduction to Probability Models*, London: Academic Press.

Savage, L. J. (1954). *The Foundation of Statistics*, New York: Wiley.

Selten, R. (1991). "Properties of a measure of predictive success." *Mathematical Social Science*, 21, 153–167.

Selten, R., and S. Krischker (1983). "Comparison of two theories for characteristic function experiments," in *Aspiration Levels in Bargaining and Economic Decision Making*. Ed. R. Tietz, Springer. Berlin and Heidelberg, 259–264.

Shefrin, H. M., and M. Statman (1985). "The disposition to sell winners too early and ride losers too long." *Journal of Finance*, 40, 777–790.

Thaler, R. H., and E. J. Johnson (1990). "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice." *Management Science*, 36, 643–60.

Thatcher, A. R. (1957). "Studies in the history of probability and statistics VI. A note on the early solutions of the duration of the play." *Biometrika* 44, 515–518.

Tversky, A., and D. Kahneman (1991). "Loss aversion in riskless choice: A reference-dependent model." *Quarterly Journal of Economics*, 106, 1039–61.

Tversky, A., and D. Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty*, 5, 297–323.

Volij, O. (1994). "Dynamic consistency, consequentialism and reduction of compound lotteries." *Economics Letters*, 46, 121–129.

Weber, M., and C. F. Camerer (1998). "The disposition effect in securities trading: An experimental analysis." *Journal of Economic Behavior and Organization*, 33, 167–184.

APPENDIX

Proof of Claim 1

The following two properties will be used to show that an SEU maximizer will use a stationary strategy in the above process.

(a) The "laws of motion" of the process are a time-invariant set of transition probabilities (see Gneezy, 1996). Formally, let $x_t \in \{1, \ldots, N-1\}$ be the state of the process at time $t = 0, 1, 2 \ldots, T$ where T is the time at which the process ends, and may be infinite. Denote by H_t the history of the process up to time t, that is, $H_t = \{x_0, x_1, \ldots, x_t\}$. At every t, a decision-maker has to choose an action $a_t \in \{Sell, Continue\}$. Note that if $a_t = Sell$ then $t = T$ and the process ends. The time invariant property states that the probability of moving from $x_t = i$ to $x_{t+1} = j$, $i, j \in \{1, \ldots, N-1\}$ depends only on x_t and a_t, and not on $x_{t-1}, x_{t-2} \ldots, x_0$. i.e., $p_{ij} = \Pr\{x_{t+1} = j \mid x_t = i, a_t\}$.

(b) Choosing a strategy, starting at x_t, leads to a well defined stochastic process on the set of final states x_T. This is true also for the "always continue" strategy, because of the convergence of the process (see Gneezy, 1996).

From (a) and (b) we get that a given stochastic process at time t with a price x_t yields the same probability distribution over x_T as it does at time s with a price x_s if $x_t = x_s$ even if $t \neq s$. From consequentialism we know that an SEU maximizer is only interested in the probability distribution over x_T. Since any general utility function may be used by the decision-maker, the chance of indifference is assumed to be zero. So, any preferred stochastic process (i.e., preferred strategy) at time t must also be preferred at time s for $x_t = x_s$.

Proof of proposition 1

Since σ^* is a stationary strategy it creates a compound lottery with at most two possible prices. Denote the lower price $x(l)$ and the upper price $x(u)$. That is, σ^* creates a compound lottery which yields $x(l)$ with probability α and $x(u)$ with probability $1 - \alpha$, or $[x(l), \alpha; x(u), 1 - \alpha]$.

Case 1: If the decision-maker chooses *Sell* at x_t then the proposition holds, and in particular $x(u) = N$ may hold.

Case 2: If she chooses *Continue* at x_t, then $x_t < x(u)$. In particular, at $x(u) - 1$ the decision-maker chooses *Continue*, that is

$$[x(l), \beta; x(u), 1 - \beta] \succ x(u) - 1$$

(this is equivalent to the statement "*Continue*" at $x(u) - 1$). (A1)

In other words, when having $x(u) - 1$ the decision-maker chooses to participate in the compound lottery that gives her $x(l)$ with probability β and $x(u)$ with probability $1 - \beta$. Denote this lottery by $\Pi = [x(l) - x(u) + 1, \beta; 1, 1 - \beta]$, or equivalently

$$[x(u) - 1 + \Pi] \succ x(u) - 1. \tag{A2}$$

If $x(u) < N$ then one of the strategies that the decision-maker can use at $x(u)$ is "*Sell* after winning \$1 or losing $x(l) - x(u) + 1$, and *Continue* otherwise," namely take the lottery Π, since if this strategy is available at $x(u) - 1$ it is also available at $x(u)$ for $x(u) < N$. From (A2) and the NIARA assumption we get

$$[x(u) + \Pi] \succ x(u). \tag{A3}$$

That is, a decision-maker who chose *Continue* at $x(u) - 1$ will also choose *Continue* at $x(u)$ if $x(u) < N$;[11] hence, $x(u) = N$ must hold.

If needed, we can repeat this step showing that $x(u) = N$. □

The duration of the process

To find the probability distribution over prizes after T periods for $N = 10$, let $i, j \in \{0, 1, \ldots, 10\}$. E_i is the event of the decision-maker having \$$i$. $E_i(t)$ denotes that at period t, the event E_i occurs. $p_{i,j}$ is the probability of $E_j(t + 1)$, conditional on $E_i(t)$. $p_{i,j} \geq 0$ and $\sum_j p_{i,j} = 1$. It is convenient to arrange the set of $p_{i,j}$ in the form of a transition matrix P, such that i refers to the row and j to the column, so

$$P = (p_{i,j}) \tag{A4}$$

The transition matrix is,

$$p = \begin{matrix} 1 & 0 & \cdot & \cdot & \cdot & & & & 0 \\ q & 0 & p & & & & & & \\ & q & 0 & p & & & & & \cdot \\ \cdot & & & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & & & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & & & & & \cdot & \cdot & \cdot & \\ & & & & & & q & 0 & p \\ 0 & & & & & \cdot & \cdot & \cdot & 1 \end{matrix}$$

and

$$p_j(t + 1) = \Sigma_i p_{i,j} p_x(t) \tag{A5}$$

The set of all such equations can be written in the matrix form

$$p(t + 1) = p(t)P \tag{A6}$$

where $p(t)$ is a row vector, whose elements are $p_0(t)$, $p_1(t)$, . . . , $p_{10}(t)$. By applying (A6) repeatedly,

$$p(t) = p(0)P^t \tag{A7}$$

where $p(0)$ is the row vector with 1 at $E_i(0)$ and zero elsewhere. By adding the probability of being at 0 to the probability of being at N at $t = T$ we find the probability that the process ends within T periods.

Chapter 13

SUPPLY CHAIN MANAGEMENT: A TEACHING EXPERIMENT

Rachel Croson
University of Pennsylvania

Karen Donohue
University of Minnesota

Elena Katok
Pennsylvania State University

John Sterman
Massachusetts Institute of Technology

Abstract

How firms choose and manage their inventory is a question of interest for academics and practitioners in many fields, including Operations Management, Marketing, and Information Technology. Much recent attention has focused on the possibilities of information-sharing systems to aid in this setting, including sharing inventory information among firms (SAP) and sharing point-of-sale data (EDI). This classroom exercise illustrates the existence and implications of bounded rationality on the part of inventory managers, and shows how systems like these can help in inventory decision-making.

1. INTRODUCTION

This chapter describes an experiment that illustrates the challenges of supply chain management. Supply chain management involves the management of orders and shipments of goods through a supply chain; for example, shipping beer from the manufacturer to the distributor to the wholesaler and then to the retailer for sale to customers, and transmitting the orders for beer back up the supply line.

Recent research and business practice highlights the importance of managing one's supply chain. Better supply chain management can coordinate decision-making across firms. Anticipated benefits include decreased inventory costs, reduced flow times, and better matching of supply and demand. Companies as diverse as

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 285–296.
© 2005 *Springer. Printed in the Netherlands.*

Boeing, Dayton Hudson, and Eastman Chemical are investing in initiatives to better coordinate production and order decisions between supply chain partners (Stein and Sweat 1998). In academia, 50% of the presentations sponsored by the Manufacturing and Service Operations Management Society (MSOM) at INFORMS 1998 and 1999 meetings were dedicated to supply chain related topics.

However, managing a supply chain isn't easy. Demand from consumers is often both highly variable and unpredictable, including not only random events and seasonality but also growth and fluctuations induced by competitive, industry, technological, and macroeconomic processes. Demand is typically nonstationary and may not have constant variability. There is often a lag between when an order is placed and when the product arrives, either because of just-in-time production (the production process begins when the order arrives), or because it takes time to transport the product physically. Stockouts and shortages occur at various levels in the supply chain, adding further delays. And often one manufacturer, distributor or wholesaler serves multiple retailers, while each retailer receives inventory from numerous other actors.

A large body of research investigates these issues theoretically (see Croson and Donohue 2004 for a discussion). In addition to these operational challenges, there are also cognitive limitations that managers face which prevent them from optimally managing their supply chains. This chapter describes an in-class experimental game that can be used to illustrate a number of these challenges, operational and cognitive, that managers face in supply chain management.

The experiment is well-suited for undergraduate, MBA or executive teaching, and has been used in all those forums. Exactly which treatments you choose, and how deep the debriefing should be, will depend on the sophistication of your audience as well as the manner in which you choose to implement the experiment (physical or computer). Section II below describes the game, section III details possible implementation options and debriefing strategies, and section IV provides some final thoughts. Finally, an appendix lists resources (electronic and physical) for implementation.

2. THE BEER GAME

The experiment is based on a management simulation called the Beer Game. The game was developed by Sloan's System Dynamics Group in the early 1960s as part of Jay Forrester's research on industrial dynamics (Forrester 1958, 1961; Jarmain 1963). It has been played all over the world by thousands of people ranging from high school students to chief executive officers and government officials.[1]

The game's structure mimics the ordering and production decisions of a four-level serial supply chain (a retailer, wholesaler, distributor and manufacturer, who buy from and sell to only each other). Participants play the game over several dozen hypothetical weeks. Each week, players decide how many cases of beer to order from their immediate suppliers to maintain enough inventory to fill orders from their immediate customers while maximizing supply chain profit. This task is complicated by delays in order processing, production, and shipping.

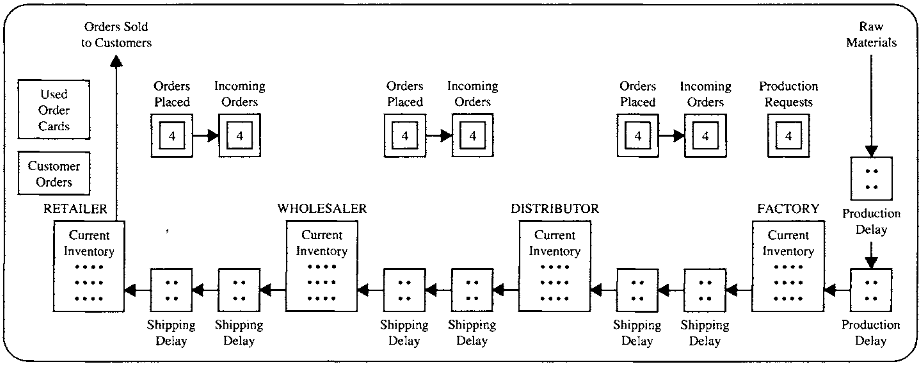


Figure 1. The Beer Game Board.

There are four roles in the game–retailer, wholesaler, distributor, and manufacturer (also called factory). Each position begins with an inventory of beer (usually, 12 cases), receives orders from, and ships beer to the player downstream. Each position also orders beer from the player upstream. A schematic of this game is depicted in Figure 1.

Each week, customers purchase beer from the retailer, who provides the beer requested out of inventory. The retailer in turn places orders for more beer with the wholesaler, who ships the beer requested out of their own inventory. The wholesaler orders and receives beer from the distributor, who in turn orders and receives beer from the factory, where the beer is brewed. At each stage there are shipping delays and order processing delays.

The boxes along the top of the figure represent the delays in order processing. When a player places an order, it takes two weeks for the order to be received by their upstream supplier. The boxes along the bottom of the figure represent the delays in shipping beer. When a player sends off a shipment, it takes two weeks for the shipment to be received by their downstream customer. Note that for the manufacturer, it takes three weeks to manufacture beer. Manipulations on these delays are discussed below.

The players' objective is to minimize total team costs. It costs $.50 for each case of beer that each player holds in inventory each week. If your customer has ordered beer and you have none in inventory, you incur a backlog cost of $1.00/case/week, which captures both the lost revenue and the ill will a stockout causes among customers. Furthermore, backlogged orders carry over; if you haven't served a customer in week 3, you can sell to them from beer that arrives at your site in week 4. Costs are assessed at each link of the distribution chain.

There is no communication (other than the transmittal of orders and shipments) allowed between positions within the team. This restriction means that the players can't explicitly coordinate their decisions. Manipulations on this limited information will be discussed below.

The numbers in the boxes show typical initial conditions. The simulation begins in an equilibrium where each player has an inventory of 12 cases of beer, and orders, shipments, and beer in the pipeline of order processing and shipping delays all reflect a steady throughput of four cases per week. In the classic game consumer demand begins at 4 cases of beer per week (for the first four weeks), then jumps to 8 cases of beer per week and remains there for the remainder of the game. Alternative demand patterns will be discussed below.

The objective of players in the game is to maximize the earnings of their supply chain (experimental manipulations to highlight incentive problems within the chain will be discussed below). When consumer demand is stationary and/or known (note, it is not in the demand function described above), the cost-minimizing strategy in this game is an order-up-to policy where each individual places an order to lift their inventory to some level S in each period, where S is a function of the demand distribution, the length of the lags, and other institutional factors (Clark and Scarf 1960, Chen 1999). However, like many equilibria, it is difficult to calculate and has been the subject of much theoretical research (e.g., Caplin 1985, Graves 1999, Chen 1999, Chen and Samroengraja 1999, Cachon 1999, Cachon and Lariviere 1999, Cachon and Fisher 2000, Chen et al. 2000).

The game can be run using a tournament payoff structure or payoffs based on the performance of each team. In a classroom tournament, each person puts up a nominal wager such as $1, and the pot goes to the team with the lowest total costs (or highest profit), winner take all (in very large groups, prizes can go to the top few teams). For research or other purposes, participants can be paid based on their supply chain's absolute or relative earnings. The tournament is exciting and works well for teaching; payoffs proportional to each team's outcomes provide a straightforward incentive to maximize performance.

2.1. Expected results

One of the most interesting and reliable results that the beer game illustrates is the *bullwhip effect.* The effect itself is described by two regularities; oscillations of orders at each level of the supply chain and amplification of these oscillations as one moves farther up the chain. Both oscillation and amplification are costly to supply chains, as factories gear up (and down) to meet ever-changing demand. Oscillation and amplification also cause firms to incur inventory and/or stockout costs, as they either have too much beer in their warehouse or not enough to sell to their customers. Oscillation and amplification in supply chains has been documented since at least the pioneering work on business cycles of Wesley Mitchell in the 1930s, and formal models of supply chains that explain how it arises date from at least Forrester (1958).

Proctor and Gamble first coined the term bullwhip effect to describe the systematic ordering behavior witnessed between customers and suppliers of Pampers diapers (Lee, Padmanabhan, Whang 1997). While customers use diapers at a fairly constant rate, Proctor and Gamble found that wholesale orders fluctuated considerably over

Figure 2. Amplification in the Macroeconomy.
Source: Federal Reserve Industrial Production Data, series B51000 Consumer Goods, B54000 Intermediate Products, B53010 Materials, each shown as the ratio to the best-fit exponential growth trend.

time. The firm also found that the orders it placed for raw materials with its suppliers fluctuated even more than these wholesale orders. Other companies have observed a similar tendency in their internal supply chains (Baljko 1999a, 1999b). Baganha and Cohen (1998) provide empirical evidence of these problems in industries with high order variation, while Kahn (1987) shows this pattern exists in the macroeconomy as well. Figure 2, depicts this pattern using detrended US Industrial Production from 1945–2002. Sterman (2000) presents many other examples, in industries from aircraft to zinc.

From its original conception and first implementation, the Beer Game has been updated in a number of ways. The original game was run on a physical game board, but now there are a number of computer interfaces that can be used if the facilitator wishes. The interfaces are often easier to use, but some argue that the physical implementation creates excitement and energy that the "colder" computer versions miss. The original game uses a particular (nonstationary) demand distribution, which is effective at making the point but doesn't have nice theoretical properties. Newer versions use stationary and known demand distributions (e.g., customers order x cases of beer each week when x is drawn from a uniform distribution, 0 to 8), which are more difficult to explain but enable the equilibrium to be calculated. These and other comparisons will be discussed in the next section on implementation.

3. HOW TO IMPLEMENT

3.1. Board game

The board game and appropriate materials can be ordered from the System Dynamics Society (the web site can be found in the resources list at the end of this chapter). These materials include the physical board, depicted in Figure 1 along with a deck of cards representing customer orders and other materials and instructions. At the beginning of each week, the retailer draws a new card representing customer demand. There are also game chips (small coins can be used), representing cases of beer. Each individual begins with 12 chips in inventory, and 4 in each of the shipping positions above. As beer moves through the system, the chips physically move as well. So when the retailer sells beer to the customers to fill orders, the chips are removed from the board. When the shipment arrives from the wholesaler to the retailer, the chips are moved from the closest square into the retailers' inventory,

The materials also include order slips on which players write the number of cases of beer that they wish to order. These slips then move along the upper track in Figure 1 to the upstream supplier who fills the orders as they come in. Instructions and a videotape for debriefing are also included.

The physical game creates a lot of excitement, and the tangibility and visibility of orders and inventory help players understand how the supply chain works. Yet the board game also poses some implementation challenges. There is a particular "order of events" in which play occurs, and players must manually record their inventory positions and orders. Players sometimes make mistakes (as occur in real supply chains as well); it is helpful to have facilitators to help run the game smoothly. An experienced person can run the game without help for up to 50 people, but for larger groups more facilitators are needed. For example, the game is the capstone event of the annual orientation for incoming MBA students at the MIT Sloan School of Management. To run the game for more than 360 people in a single session (about 50 teams) requires the assistance of about 40–50 facilitators. These facilitators are typically 2nd year students who are trained by the session leader just prior to the event (training requires a few hours).

It takes three or four weeks in the game for the players to learn and feel comfortable with the procedure. There is also (often) a delay around week 10 when players first stock out and begin to accumulate backlogs of unfilled orders; it is helpful to explain the meaning of backlogs at this point, particularly for players with limited business experience. The game runs for 36 or 48 weeks, but participants are usually told the game will last a simulated year (50 weeks) to avoid endgame effects.

To implement on a board game you'll want to have the materials, and also some facilitators to watch the game, answer questions and help with calculation. For larger groups you'll need a loud voice or wireless microphone to call out the steps for each week.

The manual game takes about 2 hours to play, including time for instructions. At the end, the instructor collects the record sheets, and graphs results (orders and net inventories) for discussion either after a break or in the next class. Additional support for physical implementation, forms and other material can be found online in the resource section at the end of this chapter.

3.2. Computer game

The game is sufficiently popular that a number of individuals have developed web-based and/or server-based versions of the game. Some websites where the game can be played online can be found in the resources section at the end of this chapter.

Play is typically faster in computerized versions. There is never an issue of the order of moves; the computer automatically enforces them. You avoid accounting and other math errors participants often make. It is also easier to control the communication between participants; in the board game the other members of the supply chain are sitting next to each other and in the computer game they most likely are not. Finally, computer implementations allow you to save the data directly for charts and graphs, rather than having to type it in or make graphs by hand.

However, the computer implementation has disadvantages as well. In some of these games the parameters are fixed, thus doing the manipulations described below is not feasible without original programming. Finally (and most importantly for classroom use), the students simply don't get as excited or emotional about the game when it's played via computer as they do when it's played face-to-face. The game is not merely about the dynamics of supply chains. Participants in the board version typically experience considerable frustration and often blame their teammates for the poor performance and instability they experiences. These emotional dynamics are at least as important as the dynamics of orders and inventory. The game illustrates important lessons about the dynamics of teams, group and individual decision making, and the tendency of people to form unfounded and dysfunctional attributions about others that hinder individual and organizational learning (see Senge 1990; Sterman 2000). Indeed, for these reasons the game is widely used in teambuilding and organizational development workshops, and by companies in the service, financial, and other industries where there are no physical inventories at all. The board version is superior to the computer version in addressing these issues.

3.3. Debriefing

So now you've played the game and collected the students' outcomes. Now what? The first step is to put the outcomes into a form that will be clear and understandable to the students. Typically these involve graphs. These graphs can depict the orders each individual placed, their inventory/backlog position, and/or profits/costs earned in the game.

Figure 3 shows an example of the orders placed and inventories/backlogs from supply chains consisting of graduate students and business executives, originally reported in Sterman (1989). Each column shows the results of a single supply chain, with the retailer on the bottom and the manufacturer on the top. Orders placed are

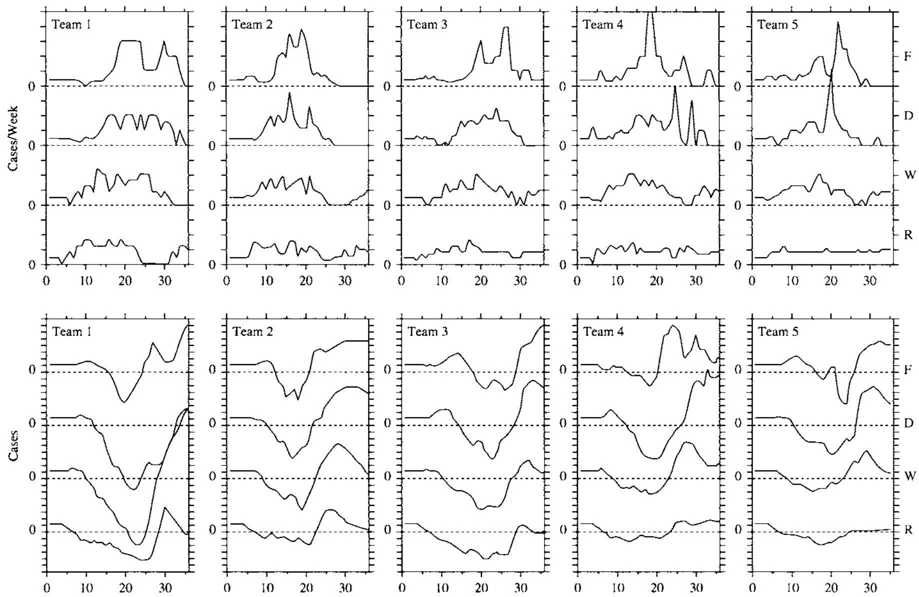


Figure 3. Orders Placed and Net Inventory Positions of Typical Teams.

in the top panel, and net inventory in the bottom (backlogs are shown as negative inventory).

As is typical, we observe oscillation and amplification. The peak order placed by the manufacturer is on average more than double the peak order of the retailer. Graphs that you make from your course will look remarkably similar to these. Though each player is free to make their own decisions, the same patterns of behavior emerge in all the supply chains, demonstrating the role of institutions in shaping our behavior.

In addition to (and possibly before) showing the outcomes, you may want to have a discussion about the process. Participants in the game (especially the board game) often report feeling frustrated and helpless. Many blame others – their teammates, the customers, or even the professor – for their failings.

One interesting exercise to do is the following. Immediately after the game, ask the players (except retailers, who directly experienced it) to graph their best estimate of customer demand. Most draw a pattern with huge swings (see Figure 4 for an example). Blaming the customer for the observed cycle is attractive (and often observed in real organizations) but, it turns out, incorrect. Customer demand is, in fact, not cyclic as depicted below.

Research reported in Sterman (1989) shows how this occurs. Most people have difficulty appreciating the multiple feedback loops, time delays and nonlinearities in the system, and use instead a very simple heuristic to place orders. In particular, many players ignore the supply line of beer they have ordered but not yet received. Players faced with a backlog will typically order enough to eliminate it. The delays

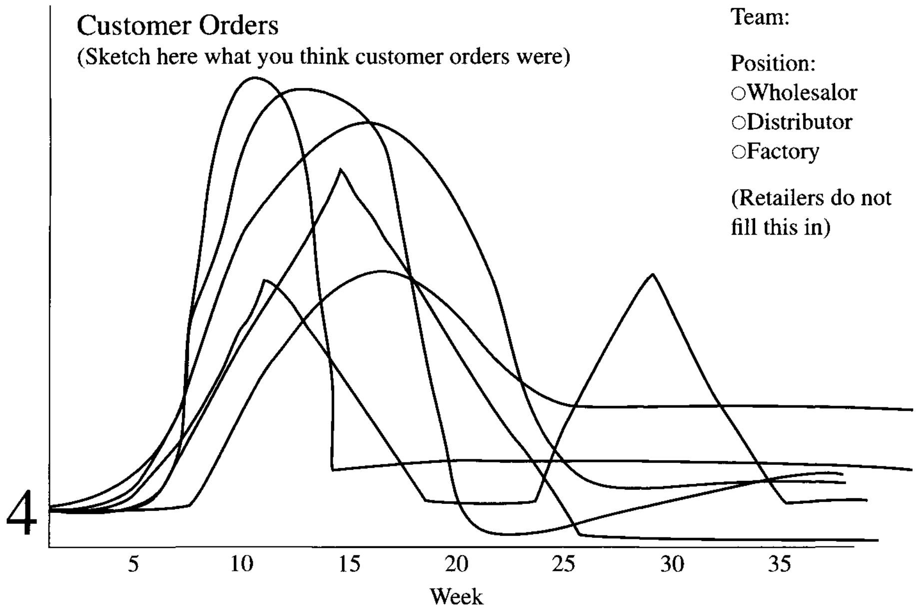


Figure 4. Ensemble of typical player estimates of customer demand.
In fact, orders in the classic game never exceed 8 cases/week and do not fluctuate.

in order fulfillment mean this beer is not immediately received. However, knowing they have ordered enough to clear the backlog, players should reduce their orders and wait patiently for the beer they've already ordered to arrive. Most, however, do not, but continue to order until they actually take delivery and clear the backlog. Consequently, they over-order, often by large amounts, and accumulate large excess inventories. The tendency to ignore the supply line contributes to the amplification and distortion of customer demand as it is passed up the supply chain (see Sterman 1989, Kleinmuntz 1985 for analysis and discussion; Sterman 2000 provides examples from diverse real systems and case studies of industries where firms and investors ignore time delays including real estate, commodities, shipbuilding, and high technology). Game leaders can also estimate the decision rules of the players using the procedure described in e.g., Sterman 1989 and Croson et al. 2003, which can also provide some useful material for discussion in the debrief.

Game leaders should allow about as much time for debriefing and discussion of the results as it takes to play. One of the profound lessons of the game is that well intentioned, intelligent people can create an outcome no one expected and no one wants. Second, individual decision-making is often biased, and those biases have large and important consequences. Finally, faced with poor performance, most people blame external events (fluctuations in customer demand) or other people. These attributions are not correct – customer demand does not fluctuate, and essentially every person generates the same patterns, indicating it is not particular people that cause the dysfunctional behavior but the way in which the structure of the system molds our behavior so that diverse people generate the same patterns. Focusing on external events or the people in the system diverts attention from opportunities for improvement – the design of the system – and thwarts learning (Sterman 2000, Senge 1990).

3.4. Manipulations

In addition to the baseline game described here, a number of scholars have investigated the impact of changing the structure or parameters of the game. Forrester (1961), and later, Kaminsky and Simchi-Levi (1998), show that shortening the time-delays improves performance. Croson and Donohue (2003) and Gupta et al., (2001) investigate the impact of sharing customer demand (point-of-sale data) with the entire supply chain. Croson and Donohue (2004) show that performance improves when inventory information is shared across the supply chain. Croson et al. (2003) investigate changing the demand function and using automated players as counterparts.

Other manipulations include alternative demand distributions (including constant, uniform[0, 8] and s-shaped), changing the initial conditions (the amount of inventory and orders everyone starts with), changing the ratio of costs of inventory/backlog, altering the objective function (playing for yourself versus for the team), allowing communication among teammates before or during the game, and the introduction of simulated (usually, optimal) players in the other roles. Some of these manipulations are easier to do in the board version and others in the computer version. If you have a large class, or two classes which will run at separate times, implementing one of these manipulations and comparing the two treatments adds interest to your discussion.

4. FINAL THOUGHTS

There is by now a large and growing literature on the benefits of experiential learning generally and classroom experiments in particular. This chapter has introduced a classroom experiment on supply chain management. The experiment is particularly useful in classes on operations management, but is also of interest in other domains like decision-making, organizational behavior, strategy, or macroeconomics. The game is fun to play, engages students and provides enduring lessons for the management of complex dynamic systems.

4.1. Resources

System Dynamics Society: ordering form for board game
http://www.albany.edu/cpr/sds/Beer.htm
MacNeil-Lehrer Report, (1989) *Risky Business – Business Cycles*, Video, Public Broadcasting System, aired 23 October 1989. (videotape available for debriefing)

Craig Kirkwood's System Dynamics Resource page includes materials and instructions for running the board game version.
http://www.public.asu.edu/~kirkwood/sysdyn/SDRes.htm.

Electronic versions can be found at the following sites.
http://beergame.mit.edu/guide.htm (done by Li and Simchi-Levi)
http://www.beergame.lim.ethz.ch/ (done by Nienhaus)
http://www.bwl.tu-darmstadt.de/bwl1 (done by Dickerhof and Schilz, supervised by Prof. Dr. Hartmut Stadtler)

NOTE

[1] The reader should note that there is no beer in the beer game, and the game does not promote drinking. Originally the "production-distribution game", most students are more excited about producing beer than widgets or toasters. When played in, say, high schools, it easily becomes the apple juice game. Many organizations customize it to their context, for example, BP (British Petroleum) recast it as "The Oil Biz Game" for use in their refinery and distribution operations.

REFERENCES

Baganha, M. and M. Cohen (1998). "The Stabilizing Effect of Inventory in Supply Chains." *Operations Research*, 46, S72–S83.

Baljko, J. (1999a). "Expert Warns of 'Bullwhip Effect'." *Electronic Buyers' News*, July 26.

Baljko, J. (1999b). "Creativity Turns EFTC Supply Chain to Gold." *Electronic Buyers' News*, June 7.

Cachon, G. (1999). "Managing supply chain demand variability with scheduled ordering policies." *Management Science*, 45, 843–856.

Cachon, G. and M. Fisher (2000). "Supply Chain Inventory Management and the Value of Shared Information." *Management Science*, 46, 1032–1048.

Cachon, G. and M. Lariviere (1999). "Capacity Choice and Allocation: Strategic Behavior and Supply Chain Performance." *Management Science*, 45, 1091–1108.

Caplin, A. (1985). "The Variability of Aggregate Demand with (*S, s*) Inventory Policies." *Econometrica*, 53, 1395–1409.

Chen, F. (1999). "Decentralized Supply Chains Subject to Information Delays." *Management Science*, 45, 1016–1090.

Chen, F., Z. Drezner, J. Ryan, and D. Simchi-Levi (2000). "Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Leadtimes and Information." *Management Science*, 46, 463–493.

Chen, F. and R. Samroengraja (1999). "The stationary beer game." *Production and Operations Management*, 9, 19–30.

Clark, A. and H. Scarf (1960). "Optimal policies for a multi-echelon inventory problem." *Management Science*, 6, 475–490.

Croson, R. and K. Donohue (2002). "Experimental Economics and Supply-Chain Management." *INTERFACES*, 32, 74–82.

Croson, R. and K Donohue (2003). "The Impact of POS Data Sharing on Supply Chain Management: An Experimental Study." *Production and Operations Management*, 12, 1–11.

Croson, R. and K. Donohue (2004). "Behavioral Causes of the Bullwhip Effect and the Observed Value of Inventory Information." Working Paper, OPIM, University of Pennsylvania.

Croson, R., K. Donohue, E. Katok and J. Sterman (2003). "Safety and Coordination Stock: The Bullwhip Effect Revisited." MIT Working Paper.

Forrester, J. (1958). "Industrial Dynamics: A Major Breakthrough for Decision Makers." *Harvard Business Review*, 36, 37–66.

Forrester, J. (1961). *Industrial Dynamics*. Cambridge, MA: MIT Press.

Graves, S. (1999). "A Single-Item Inventory Model for a Nonstationary Demand Process." *Manufacturing and Service Operations Management*, 1, 50–61.

Gupta, S., J. Steckel and A. Banerji (2002). "Dynamic Decision Making in Marketing Channels: An Experimental Study of Cycle Time, Shared Information and Customer Demand Patterns," in Zwick, Rami and Amnon Rapoport (eds.), *Experimental Business Research*. Kluwer Academic Publishers: Norwell, MA and Dordrecht, The Netherlands.

Jarmain, W. (ed.) (1963). *Problems in Industrial Dynamics*. Cambridge, MA: MIT Press.

Kahn, J. (1987). "Inventories and the Volatility of Production." *American Economic Review*, 77, 667–679.

Kaminsky, P. and D. Simchi-Levi (1998). "A new computerized beer game: A tool for teaching the value of integrated supply chain management." in *Global Supply Chain and Technology Management*, POMS Series in Technology and Operations Management (eds.), H. Lee and S. M. Ng, Vol. 1, pp. 216–225.

Kleinmuntz, D. (1985). "Cognitive Heuristics and Feedback in a Dynamic Decision Environment." *Management Science*, 31, 680–702.

Lee, H., P. Padmanabhan and S. Whang (1997). "Information Distortion in a Supply Chain: The Bullwhip Effect." *Management Science*, 43, 546–558.

Mosekilde, E., E. Larsen and J. Sterman (1991). Coping with complexity: Deterministic Choas in human decision making bahavior. In J. L. Casti & A. Karlqvist (eds.), *Beyond Belief: Randomness, Prediction, and Explanation in Science*, 199–229. Boston: CRC Press.

Roberts, E. (ed.) (1978). *Managerial Applications of System Dynamics*. Cambridge, MA: Productivity Press.

Senge, P. (1990). *The Fifth Discipline*. New York: Doubleday.

Stein, T. and J. Sweat (1998). "Killer supply chains." *Information Week*, 708, Nov. 9, 36–46.

Sterman, J. (1984). "Instructions for Running the Beer Distribution Game," D-3679, System Dynamics Group, MIT, E60-383, Cambridge, MA 02139.

Sterman, J. (1988). "Deterministic Chaos in Models of Human Behavior: Methodological Issues and Experimental Results." *System Dynamics Review*, 4, 148–178.

Sterman, J. (1989). "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment." *Management Science*, 35, 321–339.

Sterman, J. (1992). "Teaching takes off: Flight simulators for management education." *OR/MS Today*, 19, 40–44.

Sterman, J. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*, Irwin/McGraw-Hill: New York.

Chapter 14

EXPERIMENT-BASED EXAMS AND THE DIFFERENCE BETWEEN THE BEHAVIORAL AND THE NATURAL SCIENCES

Ido Erev and Re'ut Livne-Tarandach
Technion, Haifa, Israel

Abstract

An analysis of exams used to evaluate college students highlights an important difference between the natural and the behavioral sciences. Most questions in the natural sciences ask the examinee to predict the results of particular experiments. On the other hand, nearly all questions in the behavioral sciences deal with abstract terms. The current analysis clarifies this difference, and proposes two related steps that can lessen the gap. The first steps involve the development of questions that focus on experiments that have been run. A field study suggests that the discrimination power of questions of this type is not lower than the discrimination power of theory-based questions. The second step requires some changes in the information collected by researchers and presented to students.

1. INTRODUCTION

One of the clearest indications of the gap between the behavioral and the natural sciences is provided by the exams used to evaluate students. Typical questions in natural science exams ask examinees to predict the outcome of particular experiments. For example, a question might show a simple circuit and ask, "If the switch is closed at time zero, which of the following curves shows the current through the resistor as a function of time?" Typical questions in the behavioral sciences, on the other hand, ask examinees to state the meanings of particular terms, or to associate them with particular theories.

The current paper proposes and evaluates a method that can facilitate the development of natural science-like prediction questions for the behavioral sciences. We believe that such a change can benefit behavioral scientists in two ways. First, our experience teaching engineering and business students with good backgrounds in the natural sciences suggests that these students tend to find concept-focused multiple-choice exams inappropriate to evaluate their achievements. These students do not understand why they must memorize the structure of abstract theories from

R. Zwick and A. Rapoport (eds.), Experimental Business Research, Vol. III, 297–308.
© 2005 *Springer. Printed in the Netherlands.*

which no clear predictions may be drawn. Including in exams questions that focus on predictions would increase the face validity of these exams, improve students' attitudes toward the exams and the course material on which they are based, and so make the students more effective learners. Second, it is also possible that a new focus on predictions in exams, courses, and textbooks will affect mainstream research and will help close the gap between the behavioral and the natural sciences.

The paper is organized as follows. Section 1 evaluates and quantifies the assertion that there is a qualitative difference between exam questions in the natural and behavioral sciences. In this section, we compare the GRE subject exams in Physics and Psychology (exams used by leading universities to evaluate candidates for graduate school; see http://www.gre.org/edindex.html). The results show substantial differences between the two. Whereas nearly all the questions in the Physics exam focus on concrete situations, in the Psychology exam nearly all questions focus on abstract terms.

We believe that the focus of behavioral science exams on abstract terms and theories is driven by the fact that an understanding of leading behavioral theories does not ensure accurate predictions of behavior (see related discussion in Heiner, 1983). Thus, exam developers cannot know with certainty the correct answer to a question that requires prediction. At best, they can know what the theory predicts. Section 2 presents one solution to this problem, based on the fact that it is possible to ask questions about specific experiments (laboratory or natural) that have been run. For example, a question might describe a specific experiment and ask the examinee to choose among several possible results. Notice that this focus on concrete situations does not mean that theories are not important. Understanding useful models of human behavior should help students remember the results of important experiments studied in class, and predict behavior in experiments that were not included in the class material.

Section 3 summarizes a pilot (case) study that evaluates the validity (discriminative power) of experiment-based questions. The results show almost no difference between experiment-based and concept-based questions in their ability to discriminate between strong and weak examinees.

Section 4 highlights four problematic properties of experiment-based questions. Here, we suggest that modifying the information students receive (and researchers collect) concerning the value of descriptive models can reduce the negative effect of these properties, and increase the discriminative power of experiment-based questions. Two procedures that can facilitate this goal are discussed.

2. COMPARISON OF GRE SUBJECT EXAMS

In order to evaluate the difference between typical exams in the natural and behavioral sciences, the current section compares the GRE subject exams in Psychology and Physics. We chose to focus on GRE exams because these exams have been carefully developed to evaluate the knowledge taught in undergraduate programs, and they are used to evaluate candidates for top graduate schools.

To help students prepare for the GRE exams (which are developed by the Educational Testing Service, or ETS), the GRE web site (http://www.gre.org/edindex.html) presents a practice book in each subject. The practice books in Psychology and Physics include 214 and 99 questions, respectively. In our analysis, we have divided these questions into three categories: "abstract," "experiment-based," and "mixed." A question was classified as "abstract" if the correct answer is a property of a theory. A question was classified as "experiment-based" if the correct answer is the result (or the likely result) of a particular experiment. A question was classified as "mixed" if it explicitly asks about the relationship between a particular theory and a particular experimental result. Table 1 presents examples of the three types of questions from the GRE practice books in Psychology and Physics.

The right-hand column of Table 1 presents the distribution of questions over the three categories in the two exams. The results reveal a large difference between the two exams. The experiment-based category accounts for 63% of the questions in Physics, and only 10% in Psychology. The abstract category accounts for 84% of the questions in Psychology, and only 9% in Physics.

3. EXPERIMENT-BASED QUESTIONS

Examination of the (few) experiment-based questions in the Psychology practice book reveals that these questions tend to focus on a small set of robust experimental results. These deal with taste aversion, extinction, serial effects in memory, physiological mechanisms of the senses, and perceptual biases. Only one of the 21 questions in social and/or organizational psychology focuses on a specific experiment. This question focuses on the mirror effect.

We believe that the relatively small number of experiment-based questions in Psychology is an indication of the fact that the number of robust experimental phenomena is small. Thus, exam writers seeking to develop experiment-based questions in psychology must struggle with two related problems. First, in many cases the correct answer (the likely experimental outcome) is unknown. Second, even when the exam developers know the likely outcome, the examinee may not be able to derive it; even the best students are expected to err in some cases. This fact is expected to reduce the discrimination power of experiment-based questions.

The method proposed here to facilitate the development of experiment-based questions in Psychology is based on a technical solution of the first problem. We propose to focus questions on specific experiments that have been run. With this focus, the identification of the "correct answer" is easy: it is the obtained (robust) experimental result.

To clarify this technical solution it is constructive to present an example. Question 3 in Table 1 focuses on the relationship of an abstract term to an experimental result (it is a "mixed" question). The following question is an experiment-based modification of this question:

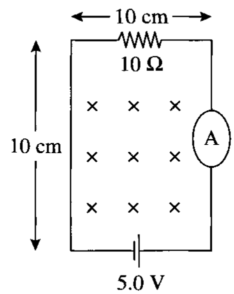
Question 7

Brown and McNeill (1966) read the definitions of uncommon words to subjects and then asked them to supply those words. When asked for words they thought they knew but could not recall, subjects often responded with words that:

Table 1. Examples of GRE questions in Psychology and Physics that were classified to the different categories. The right hand column presents the proportion of questions in each category.

	Question type	*Example*	*Proportion in GRE test*
Psychology			
	Abstract	Question 1: According to Piaget, the major cognitive attainment of the sensorimotor period is (A) speech perception (B) shape constancy (C) mental representation (D) nonegocentric thought (E) recognition memory	**0.84**
	Experimental based	Question 2: Subjects are presented with a randomly arranged list of animals, fruits, and tools, and then asked to recall the list in any order they wish. Their recall protocols are most likely to show which of the following? (A) The items with the same initial letters occur close together. (B) The items that rhyme occur close together. (C) The items that belong to the same conceptual category occur close together. (D) The items occur in an order highly similar to that used for presentation. (E) The items from only one of the conceptual categories are recalled.	**0.1**
	Mixed	Question 3: Brown and McNeill (1966) read the definitions of uncommon words to subjects and then asked them to supply those words. When asked questions about any words they thought they knew but could not recall, subjects often responded with words that were phonologically similar to the target words. The phenomenon investigated in this experiment is called (A) eidetic imagery (B) proactive inhibition (C) the complexity-of-expression phenomenon (D) the tip-of-the-tongue phenomenon (E) the template model	**0.06**

Table 1. (cont'd)

	Question type	*Example*	*Proportion in GRE test*
Physics	Abstract	Question 4: According to the Standard Model of elementary particles, which of the following is NOT a composite object? (A) Muon (B) Pi-meson (C) Neutron (D) Deuteron (E) Alpha particle	**0.09**
	Experimental based	Question 5:  2. The circuit shown above is in a uniform magnetic field that is into the page and is decreasing in magnitude at the rate of 150 tesla/second. The ammeter reads (A) 0.15A (B) 0.35A (C) 0.50A (D) 0.65A (E) 0.80A	**0.63**
	Mixed	Question 6: When the beta-decay of ^{60}Co nuclei is observed at low temperatures in a magnetic field that aligns the spins of the nuclei, it is found that the electrons are emitted preferentially in a direction opposite to the ^{60}Co spin direction. Which of the following invariances is violated by this decay? (A) Gauge invariance (B) Time invariance (C) Translation invariance (D) Reflection invariance (E) Rotation invariance	**0.28**

A. Were phonological similar to the target word.
B. Were semantically similar to the target word.
C. Had the exact opposite meaning of the target word.
D. Were names of animals or natural phenomena.

In the original question (Question 3 in Table 1) the correct answer is an abstract term. In the modified question the correct answer is the experimental result (item A) that is summarized by this term.

It is also easy to develop experiment-based questions that involve quantitative predictions. Here is one example:

Question 8
In an experiment conducted by Gneezy, Haruvy and Yafe (2003), groups of six participants were invited to eat lunch in an inexpensive Haifa restaurant. Three conditions were compared. In all three conditions the participants received 70 Israeli shekels participation fee, and were asked to personally state their order on a form. In Condition Individual, each participant paid her own bill. In Condition Split, the bill was split between the six group members. In Condition 1/6, each participant paid only 1/6 of her own order. The participants were informed of the form of payment (individual, split, or 1/6) prior to their decision about the order. The average order was 37 shekels in Condition Individual, and 57 shekels in Condition 1/6. What was the average order in Condition Split?

A. 37 B. 43 C. 51 D. 57

Notice that in the last question, the derivation of the correct answer (C) requires a good understanding of more than one principle. Good students should know that: People tend to free-ride (so the average order is likely to be larger than 37 shekels), but to a lesser extent than the rational model would predict (that would imply an average order of around 57 shekels). A student who understands the importance of individual differences and the regression effect should conclude that 51 is a more reasonable answer than 43.

The focus on experiments that have been run does not solve the second problem listed above. That is, it is not clear that the information taught in psychology courses is sufficient to ensure that good students will be able to derive the correct answers to experiment-based questions. The next section evaluates the magnitude of this problem.

4. A PILOT (CASE) STUDY

In an attempt to explore the discriminative power of experiment-based questions in Psychology exams, we incorporated questions of this type in five multiple-choice exams in courses that we have taught at the Technion. Two of the exams determined the grades in the course "Introduction to Experimental Psychology," a core course in the undergraduate program in Industrial Engineering that was taught in the fall semester of 2002. The first exam focused on cognitive psychology, the second on social and organizational psychology. Two additional exams focused on the same course and material in the 2003 spring semester. The final exam was the mid term in the elective "Thinking and Decision Making," which we taught in the fall of 2002.

The students in these courses were informed that their grades would be determined based on a new (and experimental) exam format. They were told that most questions in the exams would ask them to predict (or guess) the results of specific experiments. The students were also informed that some of these experiments would be covered in the course material, but that others would not. It was emphasized that in all experiment-based questions, the correct answer is the obtained experimental result; thus, in the case of experiments not covered in the course, it would be impossible to know the correct answer with certainty. Nevertheless, a good understanding of the course material should ensure good educated guesses and high final grades.

In order to evaluate the discriminative power of the different question formats, we first classified each question in one of three categories: "Abstract or mixed" (as defined in Section 2); "Covered experiment-based" (questions dealing with experiments that were covered in the course); and "New experiment-based" (questions dealing with experiments not covered in the course). A discrimination score was computed for each question, based on the difference between the mean Grade Point Average (GPA) of those students who answered the question correctly and the mean GPA of those who did not. The courses we consider are taken in the fifth or later semester, whereas most courses taken in earlier semesters (which, therefore, determined the relevant GPA) are in math, the natural sciences, and engineering. The standard deviation of the GPA scores over the three courses was 4.82.

Table 2 presents the mean discrimination score in each category for each of the five exams. The results reveal a surprisingly small and inconsistent difference between the three categories. Over the five exams, the differences between the three types of questions are insignificant.

In addition to this analysis, we examined the discriminative power of questions asked in an Introduction to Psychology exam given by a different teacher during the spring semester of 2001. This course was taught in a traditional way, with an emphasis on psychological theories rather than experimental findings. We analyzed the discriminative power of all multiple-choice questions (17 in number) used in the final exam for this course. According to the classification used above, all the questions in this test were abstract ones. The discrimination analysis shows an average discrimination score of 1.12, and a standard deviation of 1.39. These findings are similar to the average discrimination scores of the abstract questions used in our exam (see Table 2 bottom row).

Another interesting statistic involves the class evaluation (the students' assessment of the quality of the course). In the first semester in which the new method was used, the class evaluations dropped slightly relative to previous semesters. However,

Table 2. Discrimination scores of the different type of questions in the five exams.

Question type	*Discrimination scores (standard deviation), and number of questions*				
	Cognitive psychology		*Social psychology*		*Decision making*
	Fall 2002	*Spring 2003*	*Fall 2002*	*Spring 2003*	*Fall 2002*
Old experimental based (material covered in the course)	**0.8909** (1.6) N = 29	**1.98** (3.01) N = 18	**.929** (1.27) N = 18	**0.991** (2.21) N = 22	**2.46** (2.717) N = 12
New experimental based (not covered in the course)	**0.185** (1.29) N = 6	**1.86** (2.15) N = 6	**1.456** (2.37) N = 16	**0.335** (2.8) N = 10	**1.52** (3.74) N = 11
Abstract	**1.05** (0.39) N = 3	**0.988** (2.46) N = 10	**.202** (0.706) N = 5	**0.748** (1.39) N = 7	**1.55** (3.68) N = 10

in the second semester, the class evaluations rose above the mean of the evaluations in classes using traditional methods.

5. POTENTIAL IMPROVEMENTS

The insignificant difference in discriminative power between our experiment-based and abstract questions is most naturally explained by means of opposing effects that cancel each other out. It seems that some of the unique properties of experiment-based questions have a positive effect on discrimination, while other properties have a negative effect. If this is indeed so, then modifying the problematic properties of experiment-based questions should increase their value. The present section offers a first step in this direction. Here, we identify four problematic properties of experiment-based questions, and discuss ways to reduce the negative effect of these properties.

Biased samples.

The most important problem involves the criteria for publishing experimental results. One of the first things editors look for is "surprising findings" – that is, findings that violate popular models of the type presented in textbooks. As a result, the papers published in top psychology journals – the papers we used to develop the new experiment-based exam questions – represent a biased sample of experiments. In some cases, a good understanding of the textbook models simply would not help students predict the results of these experiments.

Vague boundaries.
A second problem is that most descriptive models focus on relatively narrow sets of empirical results and/or stylized facts. In most cases, researchers (and/or textbooks) pay relatively little attention to defining the set of situations that can be addressed by the proposed model.

Introspection and intuition.
A third shortcoming of experiment-based questions involves the possibility that in certain cases, intuition, introspection and/or personal experience can be used to derive more accurate predictions than the descriptive models taught in the course. For an example, consider the following question:

Question 9:
A study on the use of ear protectors in large factories in Israel (Zohar, Cohen & Azar (1980)) shows that typical workers:

A. Use ear protectors less often than they should according to the safety rules.
B. Use ear protectors only when instructed.
C. Use ear protectors more often than they should according to the safety rules.

Most students would guess that the correct answer is A. However, students who take Maslow's (1970) motivation pyramid (one of the models taught in the course) too seriously are likely to err: this model implies that physiological needs and personal safety are always satisfied before addressing other needs.

Creative experimental paradigms.
A fourth shortcoming involves the substantial differences among the various experimental paradigms used to demonstrate different phenomena. Typical paradigms in the behavioral sciences involve many details that are not manipulated during the experiments, including incentives offered the participants; the cover story; general instructions given; the use of deception; the subjects' demographic makeup; and the possibility of "clarification" questions. As noted by Hertwig and Ortmann (2002), these details tend to vary from study to study. Hence, basing an exam question on any experiment not taught in class requires simplifying the description of the paradigm. In our questions we tried to keep the important details, but the distinction between important and minor details is not necessarily clear-cut, and it may be that our simplifications left out important details and so impaired the predictive ability of the good students.

5.1 Standardized descriptive models

One approach that may increase the discriminative power of experiment-based questions is provided in Erev, Roth, Slonim & Barron (2004). This paper suggests a procedure for standardizing descriptive models that can be used to reduce the first three problematic properties discussed above.

The suggested standardization process is similar to the standardization of psychological tests (see Anastasi, 1996). It includes two major steps. The first step starts with a translation of the relevant theory (or model) to theory-based point prediction rules. The term "point prediction rules" refers to an equation and/or computer program that uses precise input (the parameters of the situations and the parameters of the model) to make precise predictions of future behavior. The parameters of this precise version of the model are estimated by running experiments. To ensure robust estimates, the experimental conditions are randomly drawn from a well-defined universe of tasks to which the model is assumed to apply. Thus, the first part of the standardization process reduces the need to rely on biased samples of questions and models with vague boundaries. Instead, it implies a random selection of experimental conditions, and a clear definition of the boundaries of the model.

The second stage of the standardization procedure involves estimating the optimal weighting of the point prediction and "new data." The "new data" consist of a few observations of individuals' behavior in an experiment identical to the experiment to be predicted. To clarify this concept and its relationship to the current context, it is convenient to consider a concrete example. Consider the following question:

Question 10.
In the experiment conducted by Erev et al. (2004), participants were asked to select among 100 hypothetical gambles with one non-zero outcome. One of the problems presented the following pair of gambles:

> Gamble 1: Earn \$60 with $p = 0.80$; earn 0 otherwise
> Gamble 2: Earn \$74 with $p = 0.75$; earn 0 otherwise

What was the proportion of subjects preferring Gamble 1?

A. 0.09 B. 0.39 C. 0.69 D. 0.99

The correct answer is B (0.39). An example of new data in this prediction task is the observation of the examinee's own (introspective) preferences. For example, an examinee may know that prospect theory with the parameters estimated by Tversky and Kahneman (1992) predict a choice of Gamble 1, but that her own tendency is to prefer Gamble 2. The standardization proposed by Erev et al. (2004) allows an estimation of the optimal (least squared) weighting of the two predictors (the model and the new data). The optimal weight to be given to the model is summarized with one statistic: the model's Equivalent Number of Observation (ENO). When combining the model with k new observations, the model weight is (ENO)/(ENO+k), and the data weight is (k)/(ENO+k). Thus, the availability of the ENO statistic addresses the third problem listed above: It provides guidance where predictions based on a model conflict with introspection/intuition. Notice that students do not have to learn ENO values by heart. Deep understanding of the robust behavioral principles should allow the derivation of accurate estimates of the relevant ENO's.

5.2 Standardized experiments

Hertwig and Ortmann (2002) have argued that the large set of experimental conventions used by psychologists impairs our ability to compare findings and draw general conclusions. To address this problem, they suggest that experimenters should be encouraged to replicate their results in standardized settings. We believe that this idea can help address the final problem discussed above. Replicating the important experimental results in a particular field according to one basic standardized paradigm would facilitate the development of experiment-based questions. Following such standardization, exam developers would not have to begin each question with a long description of the experiment's unique paradigm. It would be reasonable to expect students to know the basic paradigms.

6. SUMMARY

An analysis of GRE exams highlights an important difference between the natural and the behavioral sciences. Most questions in Physics ask the examinee to predict the results of particular experiments. On the other hand, nearly all questions in Psychology deal with abstract terms. The current analysis clarifies this difference, and proposes two related steps that can lessen the gap.

The first step addresses the difficulty of developing experiment-based questions in the behavioral sciences. We assert that the main stumbling block, from the developer's point of view, lies in identifying questions with unambiguous correct answers. The solution proposed here is technical. It requires focusing each question on a particular experiment that has been run. With this focus the correct answer is crystal clear: It is the observed experimental result. Our analysis suggests that the discriminative power of experiment-based questions based on this technical solution is at par with the discriminative power of more typical abstract questions.

The second step requires some changes in the information collected by researchers and presented to students. We assert that the discriminative power of experiment-based questions can be improved through the standardization of descriptive models and experimental procedures. The standardization of descriptive models as suggested by Erev et al. (2004) is expected to have three benefits: It would allow unbiased selection of experimental tasks; it would clarify the boundaries of descriptive models; and it would provide guidance where models conflict with intuition, introspection and or personal experience. The standardization of experimental procedures (see Hertwig and Ortmann, 2002) is expected to be beneficial in that it would facilitate clear and parsimonious presentations of experiment-based questions.

We believe that the use of experiment-based questions to evaluate students in behavioral science courses is likely to have many attractive outcomes. In addition to making behavioral science exams more similar to those in the natural sciences, this effort will advance the behavioral sciences in substantial ways. A focus on predictions in exams is likely to have a similar effect on courses, on textbooks, and on mainstream research.

ACKNOWLEDGMENT

Please address all correspondence to Ido Erev erev@tx.technion.ac.il. We thank Yael Sagi, Miriam Erez, Ella Miron, and Greg Barron for help in data collection, and Meira Ben Gad for editorial assistance. This research uses ideas developed in conversations with Alvin E. Roth.

REFERENCES

Anastasi, A. (1996). *Psychological testing* (7th ed.). New York: Macmillian.

Brown, R. W. & McNeill, D. (1966). "The tip-of-the-tungue phenomenon." *Journal of Verbal Learning and Verbal Behavior*, 5, 325–337.

Erev, I., Roth, A. E., Slonim, R. L. & Barron, G. (2004). Descriptive models as prior beliefs with known equivalent number of observations (ENO). Working paper.

Gneezy, U., Haruvy, E. & Yafe, H. (2003). The inefficiency of splitting the bill. Working paper.

Hertwig, R., & Ortmann, A. (2002). "Experimental practices in economics: A methodological challenge for psychologists?" *Behavioral & Brian Sciences*, 24.

Maslow. A. (1970). *Motivation and personality* (2nd Ed.). New York: Harper and Row.

Tversky, A., & Kahneman, D. (1992). "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty*, 9, 195–230.

Zohar, D., Cohen, A. & Azar, N., (1980). "Promoting increased use of ear protectors in noise through information feedback." *Human Factors*, 22, 69–79.

Author Index

Subject Index

The Authors

The Editors

Rami Zwick
Department of Marketing
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
mkzwick@ust.hk

Amnon Rapoport
Department of Management and Policy
University of Arizona
Tucson, AZ 85721
and
Department of Economics
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
amnon@u.arizona.edu

Chapter Contributors

Wilfred Amaldoss
Department of Marketing
Fuqua School of Business
Duke University
Durham, NC 27708
wilfred.amaldoss@duke.edu

David Bardolet
UCLA Anderson School
110 Westwood Plaza
Los Angeles, CA 90095-1481
dbardole@anderson.ucla.edu

Andreas Blume
Department of Economics
University of Pittsburgh
Pittsburgh, PA 15260
ablume@pitt.edu

David Bodoff
ISMT Department
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
dbodoff@ust.hk

David V. Budescu
Department of Psychology
University of Illinois
603 E. Daniel Street
Champaign, IL 61820
dbudescu@uiuc.edu

Colin F. Camerer
Rea A. and Lela G. Axline Professor of Business Economics
Division of the Humanities and Social Sciences
Mail Code 228-77
California Institute of Technology
Pasadena, California 91125
camerer@hss.caltech.edu

Juin-Kuan Chong
SKK Graduate School of Business,
Sungkyunkwan University
53 Myungryun-dong 3-ga, Jongro-gu
Seoul, Korea 110-745
bizcjk@nus.edu.sg

Rachel Croson
Department of Operations and Information Management
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104-6340
crosonr@wharton.upenn.edu

Douglas V. DeJong
Tippie College of Business
University of Iowa
Iowa City, IA 52242
douglas-dejong@uiowa.edu

Karen Donohue
Department of Operations and Management Science
The Carlson School
University of Minnesota
Minneapolis, MN 55455-9940
kdonohue@csom.umn.edu

Martin Dufwenberg
Department of Economics
University of Arizona
Tucson, AZ 85721-0108
martind@eller.arizona.edu

Ido Erev
Faculty of Industrial Engineering and Management
Technion
Haifa 32000
Israel
erev@tx.technion.ac.il

Craig R. Fox
UCLA Anderson School
110 Westwood Plaza #D511
Los Angeles, CA 90095-1481
cfox@anderson.ucla.edu

Uri Gneezy
The University of Chicago Graduate School of Business
5807 South Woodlawn Avenue
Chicago IL 60637
ugneezy@gsb.uchicago.edu

Jin K. Han
Singapore Management University
469 Bukit Timah Rd.
Singapore 259756
jkhan@smu.edu.sg

Ralph Hertwig
University of Basel
Department of Psychology
Applied cognitive science
Missionsstrasse 60/62
4055 Basel, Switzerland
ralph.hertwig@unibas.ch

Teck-Hua Ho
William Halford Jr. Family Professor of Marketing
Co-Director, Berkeley Social Sciences Laboratory (Xlab)
Haas School of Business
545 Student Services #1900
University of California at Berkeley
Berkeley, CA 94720-1900
hoteck@haas.berkeley.edu

Elena Katok
Department of Supply Chain and Information Systems
Smeal College of Business
Penn State University
University Park, PA 16802-1913
ekatok@psu.edu

Chung K. Kim
College of Business Administration
Ewha Womans University
Seoul, Korea
acckim@ewhat.ac.kr

Karen M. Kramer
Office of Research
Department of Preventive Medicine and Public Health
University of Kansas School of Medicine- Wichita
1010 N. Kansas
Wichita, KS 67214
kkramer@kumc.edu

Hugues Levecq
Reims Management School
59 Rue Pierre Taittinger
BP 302
51061 Reims Cedex
France
hugues.levecq@reims-ms.fr

Haim Levy
Department of Finance
School of Business Administration
The Hebrew University
Jerusalem 91905
Israel
mshlevy@mscc.huji.ac.il

Moshe Levy
Department of Finance
School of Business Administration
The Hebrew University
Jerusalem 91905
Israel
mslm@mscc.huji.ac.il

Daniel Lieb
Fuqua School of Business
Duke University Box 90120
Durham, NC 27708-0120
dsl7@duke.edu

Michael Maier
Tippie College of Business
University of Iowa
Iowa City, IA 52242
michael-maier@uiowa.edu

Robert Meyer
Department of Marketing
Wharton School of Business
University of Pennsylvania
Philadelphia, PA 19104
meyerr@wharton.upenn.edu

Andreas Ortmann
Charles University
Economics Institute, Academy of Sciences of the Czech Republic
Politickych veznu 7, 111 21 Prague 1, Czech Republic
andreas.ortmann@cerge-ei.cz, aortmann@yahoo.com

John Sterman
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA, 02142
jsterman@mit.edu

Re'ut Livne-Tarandach
Faculty of Industrial Engineering and Management
Technion
Haifa 32000
Israel
livnere@tx.technion.ac.il

William S. Waller
Eller College of Management
University of Arizona
Tucson, Arizona 85721
wwaller@eller.arizona.edu

Hongtao Zhang
ISMT Department
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong
imhzhang@ust.hk

Shengui Zhao
Department of Marketing
Wharton School of Business
University of Pennsylvania
Philadelphia, PA 19104
shengui@wharton.upenn.edu