Foundations and Trends® in
Technology, Information and Operations Management
Vol. 5, No. 1 (2011) 1–86

DOI: 10.1561/0200000022

Using Laboratory Experiments to Build Better Operations Management Models

By Elena Katok

Contents

Foundations and Trends® in
Technology, Information and Operations Management
Vol. 5, No. 1 (2011) 1–86

DOI: 10.1561/0200000022

Using Laboratory Experiments to Build Better Operations Management Models

Elena Katok

The Pennsylvania State University, Smeal College of Business, University Park, PA 16870, USA, ekatok@psu.edu

Abstract

Laboratory experiments give researchers a great deal of control, making them useful for testing analytical models. In this monograph I discuss methodological issues in designing and conducting laboratory experiments. I also summarize some of the recent advances in using laboratory experiments in Operations Management.

1

Introduction

Much of the work in Behavioral Operations Management (BOM) lives at the boundary of analytical and behavioral disciplines — work that has a substantial tradition. In the next section I will briefly summarize the history of the uses of laboratory experiments in economics, and how the field of BOM can learn from this tradition.

Laboratory experiments are a major method we use in BOM. Similar methods have been employed in a number of other social science fields, including economics (auctions), psychology and sociology (social networks), law (jury behavior), political science (coalition formation), and anthropology and biology (reciprocity).

There are three major purposes that laboratory experiments serve [105]. (1) To test and refine existing theory. Much of the BOM work so far has been on this topic. For example, experiments testing behavior in the newsvendor model [10, 114] test how well people are able to optimize under uncertainty. (2) To characterize new phenomena leading to new theory. An excellent example is the literature on social preferences. For example, Loch and Wu [85] found in a lab experiment that concerns with status and relationship have an effect on the performance of the wholesale price contract. Cui et al. [28] develop a fairness model and apply it to the setting of a wholesale price contract, to formally characterize conditions that may lead to channel coordination with the wholesale pricing. Özer et al. [97] develop a model of trust and trustworthiness that explains some of the regularities in their lab experiment. (3) To test new institutional designs. This type of work has not yet made its way in the operations literature, but there are several notable examples in economics, such as designing the Federal Communications Commission (FCC) auctions for radio spectrum [49] or designing the market for medical interns [103].

Laboratory studies complement other methods by bridging the gap between analytical models and real business problems. Analytical models are built to be parsimonious and general, and are primarily normative in nature. They use assumptions to make the mathematics tractable. These models can be tested using a variety of empirical methods, including surveys, field studies, field experiments, or laboratory experiments. Empirical methods, are by their nature, descriptive. All empirical methods involve a trade-off between the internal and the external validity. Surveys and field studies that use secondary data have high external validity (they are close to the real settings being studied), but may be low on internal validity (the ability to establish the cause and effect relationship based on the data) because they often suffer from being confounded, or not having all the data that would ideally be required. This is because researchers cannot directly manipulate the factors or levels in the study — they have to accept data that is available to them.

The relative advantage of experiments is control. Experiments can take place in the field or in the laboratory, and field and lab experiments also differ in their level of control and in their level of external validity (field experiments have higher external validity, but usually allow for less control). Laboratory experiments can be designed to fully manipulate all factors at all desired levels, and to match the assumptions of the analytical model being tested. So laboratory experiments are high on the internal validity, but because the environment is often more artificial, they are lower on the external validity.

A good experiment is one that controls the most plausible alternative hypotheses that might explain the data. It also allows the researcher to cleanly distinguish among possible explanations. For example, the Schweitzer and Cachon [114] study looks at the behavior in the newsvendor problem. In the setting in which the critical fractile is above 0.5 (called the high profit condition) the authors find that average orders are below the optimal order and above the mean demand. At this point a potential plausible explanation is risk aversion — risk averse newsvendor should order less than the risk neutral newsvendor. But the Schweitzer and Cachon [114] design cleverly includes a low profit condition, with the critical fractile below 0.5. In that treatment risk aversion still implies that orders should be below optimal, but the authors find that orders are above optimal. Thus, the design can clearly rule out risk aversion as the (only) explanation.

Three factors make experimental work rigorous. The first one is theoretical guidance. To interpret the results of an experiment, researchers need to be able to compare the data to theoretical benchmarks. Systematic deviations from theory can provide insights into factors missing from the analytical model, and guidance into how the model can be improved.

The second factor is induced valuation. In his seminal paper, Smith [116] explains how a reward medium (for example money) can be used to control the objectives of the laboratory participants. When participants are rewarded based on their performance in the experiment, researchers have a cleaner test of how people pursue their goals. This test is not confounded by not knowing what those goals are.

The third factor is careful control of institutional structure. Strategic options and information available to participants should match with those assumed by the theoretical model. For example, real bargaining is typically done face-to-face and is often unstructured, making modeling bargaining extremely challenging. But some assumptions can be imposed on the bargaining process to make a model tractable, while still capturing some essential features of real bargaining. For example, bargainers may assume to exchange alternating offers, and to capture the fact that no bargaining process can go on forever we may assume that the pie they are bargaining over is discounted at each iteration. These two assumptions allow for a tractable model [108] that provides useful insights and clear empirical predictions. A model can be further streamlined by assuming that the bargaining process is finite. It turns out that what the model predicts about how the pie will be split depends on length of the bargaining process, and the relative discount rates of the two players. These predictions cannot be tested in the field because real bargaining processes are substantially different from the model, but the model can be tested in the laboratory. For example, Ochs and Roth [93] found that in a finite version of this bargaining game, players in the second period often make offers that are less in absolute terms than the original first period offers they received. These "disadvantageous counteroffers" however, are better in relative terms. Bolton [13] showed, among other things, that these fairness concerns are significantly reduced when players are paid based on a tournament structure.The results of these, and many other tests, provided seminal insights that formed the basis for the theory of social preferences [11, 39].

One of the questions that are often asked about laboratory experiments is about whether their results can be carried over into the real world. Smith [117] addresses this question with the concept of *parallelism*. He writes: "Propositions about the behavior of individuals and the performance of institutions that have been tested in laboratory micro economies apply also to non-laboratory micro economies where similar *ceteris paribus* conditions hold."(p. 936). In other words, behavioral regularities persist as long as relevant underlying conditions are substantially unchanged.

The art of designing good experiments (as well as the art building good analytical models) is in creating simple environments that capture the essence of the real problem while abstracting away all unnecessary details. Thus, the first step in doing experimental work is to start with an interesting theory. What makes a theory interesting is that (1) it has empirical implications, and (2) these implications are worth testing, meaning that they capture a phenomenon that is sufficiently real and interesting so that learning about it adds to our knowledge of the real world.

This monograph focuses on controlled laboratory experiments used to test existing, and develop new, theory in Operations Management. Much of the methodology I discuss is in line with economics rather than psychology, which also provide a valid and useful, but different, paradigm. The rest of this monograph is organized as follows: in Section 2 I will present a (very) short history of experimental economics, focusing specifically on some fundamental games that proved to be important in economics as well as in BOM. These games will come up again in subsequent sections. In Section 3 I will discuss some basics of experimental design as well as "best practices" for conducting laboratory experiments. In that section I will touch on issues related to providing a context, the effect of subject pool, the effect of incentives, and the uses of deception. The goal of Sections 4, 5, and 6 is to outline how experiments have been used to shed light on behavioral factors within three different operational contexts that have been the focus of my research: the behavior in the Newsvendor problem (Section 4), supply chain contracts (Section 5), and procurement auctions (Section 6). I conclude this monograph in Section 7 with a discussion of my view of future trends and promising directions for future research.

2

A Short History of Laboratory Experiments in Economics and Some Prototypical Games

2.1 Individual Decisions

The desire to test whether people behave consistently with mathematical models is perhaps as old as the desire to analytically model human behavior. The well-known *St. Petersburg Paradox* [7] was the first to illustrate the problem with modeling people as maximizing their expected profits. It goes as follows: A fair coin is tossed until a heads comes up. You get \$1 when it lands on heads the first time, \$2 when it lands on heads the second time, \$4 when it takes three tosses, \$8 when it takes four tosses. Name the greatest certain amount that you would pay to play this game once. The expected value of this bet is $\sum_{n=1}^{\infty} \frac{n}{2^n}$, and does not converge. Yet most people would value this lottery at about \$20. Bernoulli proposed a "utility function" with diminishing marginal utility so that the sums converge.

There were early experiments on individual choice testing ordinal utility theory, starting as early as Thurstone [120], who estimated individual's indifference curves through a large sequence of hypothetical questions. Almost immediately, and as a reaction to this work, Wallis and Friedman [124] criticized it for basing the analysis on hypothetical choices and encouraged future experiments in which subjects are confronted with real, rather than hypothetical, choices.

After the publication of von Neumann and Morgenstern's *Theory of Games and Economic Behavior* [123] various aspects of expected utility theory were tested. The most famous of those tests is known as the *Allais Paradox* [1]. Allais presented his subjects with two hypothetical choices. The first between alternatives A and B:

A: 100 million francs with certainty
B: 10% chance of 500 million francs
89% chance of 100 million francs
1% chance of 0

The second was between alternative C and D:

C: 11% chance of 100 million francs
89% chance of 0
D: 10% chance of 500 million francs
90% chance of 0

An expected-utility maximizer who prefers A to B should also prefer C to D, but a common pattern observed was to prefer A to B and D to C. This experiment has been subsequently replicated using (much smaller) real stakes.

The Allais Paradox is only one of many violations of the expected utility theory, and identifying numerous other violations and modifying or extending the model to account for these violations produced an enormous amount of literature at the intersection of economics and cognitive psychology. See Machina [88] for an overview and Camerer [19] for a detailed literature survey of individual decision-making.

In spite of numerous documented violations, the expected utility theory continues to be the predominant paradigm in economics. One reason for this is that, although numerous alternatives have been proposed, none are as elegant or analytically tractable as the original model. Thus, in Operations Management, in spite of Bernoulli's early demonstration in 1728, the majority of models assume expected profit maximization, and even allowing for risk aversion is a fairly new phenomenon.

		Column Player	
		Defect	Cooperate
Row Player	Cooperate	Row Earns -1 Column Earns 2	Row Earns ½ Column Earns 1
	Defect	Row Earns 0 Column Earns ½	Row Earns 1 Column Earns -1

Fig. 2.1 Payoffs in the Prisoner's Dilemma Game [44].

2.2 Simple Strategic Games

Following von Neumann and Morgenstern [123], economists also became interested in testing models of strategic interactions. One of the first strategic games studied in the laboratory is known as the *Prisoner's Dilemma* [44]. In this game two players (labeled Row and Column) must simultaneously choose one of two options (that for transparency we will label Cooperate and Defect, but that carried neutral labels "1" and "2" in the experiments). The payoffs are displayed in Figure 2.1.

Both players in the Prisoner's Dilemma game have the *dominant strategy*. A player has a dominant strategy when her preferred option does not depend on the choice of the other player. Observe that the Column Player earns more from Defecting than from Cooperating regardless of what the Row player does (2 vs. 1 if Row Cooperates, and $1/2$ vs. -1 if Row Defects). Similarly, the Row player earns more from Defecting than from Cooperating regardless of what the Column player does (1 vs. $1/2$ if Column Cooperates, and 0 vs. -1 if Column Defects). Thus the unique equilibrium in the Prisoner's Dilemma game is for both players to defect, Row earning 0 and Column earning $1/2$. This outcome is inefficient, because both players can be better off from cooperation.

Players in the Flood [44] study played 100 times, and average earnings were 0.4 for Row and 0.65 for Column — far from the equilibrium prediction but also far from perfect cooperation. The author interpreted his results as evidence against the equilibrium solution, but also included in his paper a comment by John Nash, who pointed out that in a game repeated 100 times, while Defect continues to be the unique equilibrium, other strategies are also nearly in equilibrium,[1] so the experiment to test the theory should be conducted with random matching of the players. The game of Prisoner's Dilemma continued to fascinate social scientists for decades, and still does, because of its broad applications. It has been "... used as a metaphor for problems from arms races to the provision of public goods." [p. 10].

Another topic deeply rooted in experimental economics that has important implications for Operations Management is bargaining. Güth et al. [51] were the first to conduct an experiment on the *Ultimatum Game*, that has since became the standard vehicle for modeling the negotiation process. The game involves two players. The Proposer received $10 and has to suggest a way to distribute this amount between himself, and the other player, the Recipient. The Recipient, upon observing the Proposer's split, can either accept it, in which case both players earn their respective amounts, or reject it, in which case both players earn 0. The Ultimatum Game has the unique subgame perfect equilibrium that can be computed using *backwards induction*. Looking at the responder's decision first, and assuming the responder would prefer any positive amount of money to 0, it follows that the responder should be willing to accept the smallest allowable amount (1 cent). Knowing this, the responder should offer 1 cent to the responder and take $9.99 for himself. In fact Proposers offer a split that is closer to 60% for themselves and 40% for the responder, and moreover, responders tend to reject small offers.

Since the Güth et al. [51] experiments were conducted, hundreds of ultimatum experiments have been reported. Roth et al. [107] conducted a large-scale study in four countries: US, Yugoslavia, Japan, and Israel. In each country they compared the Ultimatum Game (one proposer and one responder, canned buyer and seller) and the Market game (one seller and nine buyers). In the Market game the buyers submit sealed bids and the seller can accept or reject the highest offer. They found that in all four countries, the Market game quickly converged to the equilibrium prediction, in which the seller receives nearly the entire pie, while the results of the Ultimatum Game showed no signs

[1] For example, in the "tit-for-tat" strategy, players start by cooperating, and then mimic the behavior of the other player in the previous round [3].

of converging to this equilibrium. There were some differences reported in the Ultimatum Game among the four countries.

Ochs and Roth [93] report on a series of two-stage bargaining experiments in which player 1 makes an offer, player 2 can accept or reject, and if player 2 rejects, the pie is discounted (multiplied by $\delta < 1$), and player 2 can make an offer to player 1. Player 1 can then accept or reject, and if player 1 rejects, both players earn 0. We can work out the equilibrium again using backwards induction. Starting with stage 2, player 2 should be able to earn the entire discounted pie, which is δ. Knowing this, player 1 should offer player 2 δ in the first stage, and player 2 should accept it.

Ochs and Roth [93] report two important regularities: (1) disadvantageous counteroffers: player 2 in the second stage makes an offer that gives himself (player 2) less than player 1's offer in stage 1, and (2) the deadline effect: most agreements happen in the last second. In regards to the disadvantageous counteroffers, Ochs and Roth [93] conclude: "We do not conclude that players 'try to be fair.' It is enough to suppose that they try to estimate the utilities of the player they are bargaining with, and [. . .] at least some agents incorporate distributional considerations in their utility functions." (p. 379).

Forsythe et al. [45] specifically explore the question of what motivates proposers in the Ultimatum Game. To do this, they conducted the *Dictator game.* The Dictator game is almost the same as the Ultimatum Game, but the responder does not have the right to veto an offer. This means that there are no strategic reasons to yield any ground. Contributions reflect "pure" preferences. I will discuss the Forsythe et al. [45] paper in more detail in the following section. I refer the reader to Roth [104] for a review of bargaining experiment prior to 1995. This literature also gave rise to both, analytical and behavioral literature on other-regarding preferences (that is, incorporating concerns for others' earnings directly into the utility function). I refer the reader to Cooper and Kagel [24] for a review.

2.3 Games Involving Competition: Markets and Auctions

A central pillar of economic theory is the principle that prices clear markets. Competitive Equilibrium (CE) prices are determined at a point at which supply meets demand, but how exactly prices arrive at this level is (still) not well understood. Adam Smith famously termed this the "Invisible Hand." Some practical questions that economists disagreed on regarding the requirements for CE prices to come about included the number of buyers and sellers and the amount of information.

Chamberlin [22] set out to gain initial insights into this question with a laboratory experiment that involved a large number of students in the roles of buyers and sellers. Each buyer had a privately known value, each seller had a privately known cost, and they interacted through a series of unstructured bilateral negotiations. So this market had a large number of traders, but no centralized information. Chamberlin [22] reported that prices were quite dispersed and showed no tendency of quickly converging to equilibrium, and as a result there was substantial inefficiency.

Smith [115] conducted a famous experiment in which he essentially repeated Chamberlin's experiment, but added a *double auction* institution that allowed buyers and sellers to make and accept public bids and asks.[2] Additionally, Smith [115] repeated the market several times, allowing buyers and sellers to keep their costs and valuations for several rounds. The price converged to the equilibrium level reliably and quickly (but not in the first round). Smith's early work on the double auction institution is foundational and generated a long and fertile literature (see [58]).

The behavior of two-sided markets (multiple buyers and multiple sellers) is more complicated than behavior of one-sided markets. Markets with a single seller and multiple buyers are called *forward auctions*, and markets with a single buyer and multiple sellers are called *reverse auctions*.

The field of auction theory is extensive (see [80] for a comprehensive review), and laboratory experiments have been used to test many of these models. I refer the readers to Kagel [68] for a comprehensive review of work done prior to 1995 and to and Kagel and Levin [71] for work done since 1995.

[2] The story is that Vernon Smith initially became interested in this question after he was a subject in Chamberlin's experiment at Harvard [46].

There are two streams of auction research that are particularly relevant to BOM research because much of the work in procurement auctions builds on these ideas. I will briefly summarize some of the results here.

The first deals with testing the model in which bidders are assumed to have valuations that are independent, drawn from the same distribution (symmetric), and privately known (the independently known private value (IPV) model).

Much of the early laboratory experiments on auctions dealt with testing the implications of the revenue equivalence theory (for forward auctions) in the IPV setting. Vickrey [122] showed that if bidders are risk neutral, the expected seller revenues in forward auctions are the same in the four basic auction formats:

- *The sealed-bid first price*: bidders submit sealed bids and the object is awarded to the bidder who submitted the best bid, and this bidder pays his bid.
- *The sealed-bid second price*: bidders submit sealed bids and the object is awarded to the bidder who submitted the best bid, but he pays the amount of the second best bid.
- *The open-bid ascending* (*English*): bidders place bids dynamically during a live event. At the end of the auction the object is awarded to the bidder who submitted the best bid, and he pays the amount of his bid.
- *Clock descending* (*Dutch*): the price starts high and decreases at a regular pre-determined rate (the clock). The first bidder to stop the clock wins the object and pays the price on the clock.

If bidders are not risk neutral, the equivalence does not generally hold. If they are risk averse the equivalence holds between the sealed-bid first price and Dutch, and the sealed-bid second price and English.

Virtually all laboratory work to date that deals with revenue equivalence in auctions deals with forward auctions (see Ref. [68] for a review). Generally, laboratory tests reject all versions of revenue equivalence. Sealed-bid first price revenues were reported to be higher than Dutch revenues [25], but later Lucking-Reiley [86] reported the opposite effect in a field experiment. Katok and Kwasnica [73] found that prices in the Dutch auction critically depend on the speed of the clock, and thus can be either above or below the sealed-bid first price prices. Several models of bidder impatience have been offered [20, 73].

Similarly, there is no support for the revenue equivalence between the sealed-bid second price and English auctions because although both formats have the same dominant bidding strategy, bidders in English auctions tend to follow it, while bidders in sealed-bid second price auctions tend to place bid above their valuations [69] and are extremely slow to learn to not do that.

There is an important literature stream that examines bidding behavior in sealed-bid first price auctions and compare it to the equilibrium bidding behavior. Cox et al. [26] reported that bidding in sealed-bid first price auctions is more aggressive than it should be in equilibrium, and thus the revenue is higher when the sealed-bid first price auction is used than when the English auction is used. Cox et al. [26] show that qualitatively, this difference is consistent with risk aversion. Equivalently, in a procurement setting, Holt [57] shows that when bidders are risk averse, the expected procurement cost in equilibrium is lower in sealed-bid first price auctions than in their open-descending counterparts. But as Kagel [68] points out, in sealed-bid first-price auctions, "... risk aversion is one element, but far from the only element, generating bidding above the [risk- neutral Nash equilibrium]" (p. 525).

There are a number of studies that show that risk aversion does not organize the data well in many auction-like settings [21, 60, 69, 70]. There are also a number of more recent studies that propose other explanations, such as aversion to regret [37, 38, 41], learning [92, 94], and simply reacting to errors [50]. While the precise explanation for overly-aggressive bidding in sealed-bid first price auctions appears to be elusive, the fact that bidders tend to bid more competitively in sealed bid than in open-bid auctions appears to be quite robust and general. In Section 6 I will show that this regularity applies to a wider set of auctions than just sealed-bid first price; the "sealed-bid effect" applies also to dynamic auctions in which bidders are not certain whether they are winning or losing the auction.

3

Established Good Practices for Conducting BOM Laboratory Experiments

In this section I discuss several methodological topics related to good practices in designing and conducting laboratory experiments.

3.1 Effective Experimental Design

In laboratory experiments, researchers generate their own data, and this allows for much better control than in studies that rely on data that occurs naturally. The topic of experimental design is one that deserves a significantly more comprehensive treatment than what I can provide in a short review article. I refer the readers to List et al. [84] for a brief review, and to Atkinson and Donev [2] for a more detailed treatment, while Fisher [43] provides a very early textbook on the subject.

When we design an experiment we are specifically interested in the effect of certain variables, called *focus variables*, but not in the effect of some other variables, called *nuisance variable*. For example, if we are interested in testing a new auction mechanism, we may be specifically interested in the effect of the number of bidders, or the amount and type of feedback — those are focus variables. We may not be specifically interested in the effect of the bidder's experience, or gender, or major — these are nuisance variables. Focus variables should be systematically manipulated between treatments. For example, we may run some treatments with 2 bidders, and some treatments with 4 bidders, to establish the effect of the number of bidders. We call this varying the focus variables at several number of *levels*. In contrast, nuisance variables should be held *constant* across treatments, so that any treatment effects cannot be attributed to the nuisance variables, or to the *interaction effect* between the focus and the nuisance variables. For example, it would be a very poor design to have 2-bidder auctions to include only females and all 4-bidder auctions to include all males, because not holding gender constant introduces a confounding interaction effect between the gender and the number of bidders.

The simplest way to avoid inadvertently confounding the experimental design with nuisance variables is to randomly assign participants to treatments from a set of participants recruited from the same subject pool. Thus, it is not advisable, for example, to recruit participants from classes, because doing this may inadvertently assign all subjects from the same class to a single treatment. Similarly, it is not advisable to recruit subjects directly through student organizations, clubs, or fraternities. The idea is to avoid any systematic composition of subjects in a specific treatment.

A good experiment requires at least two *treatments*, one being the *baseline* treatment and the second being a comparison treatment. An experiment with only one treatment is not so much an experiment, as it is a demonstration. Sometimes demonstrations can be quite influential and informative (for example, Sterman [118] is a one-treatment experiment, that is a demonstration of the "bullwhip" effect).

The most straightforward way to construct treatments in an experiment is to simply vary each focus variable at some number of levels and conduct a separate treatment for each combination. This is known as a *full factorial design*. An example of a full factorial design in an experiment with focal variables being the number of bidders and the auction format, may be to vary the number of bidders at $n = 2$ or 4, and the auction format at sealed-bid or open bid. So the resulting 2×2 full factorial design is shown in Figure 3.1.

The advantage of the full factorial design is that it provides the cleanest evidence for the effect of each variable, as well as all possible

		Number of Bidders	
		n=2	n=4
Auction Format	Open-Bid	OB-2	OB-4
	Sealed-Bid	SB-2	SB-4

Fig. 3.1 An example of a 2×2 full factorial design.

interaction effect. But the disadvantage is that in an experiment with a large number of focal variables, a full factorial design can become prohibitively expensive because of the number of subjects required.

A practical way to deal with budget constraints is to use a fractional factorial design instead of full. For example, suppose you have three focal variables and you would like to vary each at two levels, which we denote as $+$ and $-$. This yields a $2 \times 2 \times 2$ full factorial design with the following eight treatments:

$$+++ \quad ++- \quad +-+ \quad +-- \quad -++ \quad -+- \quad --+ \quad ---$$

Suppose you can only afford to run four treatments. The question is, which four to run? Imposing a constraint that the third factor is the product of the first two results in a balanced design (this example can be found in Friedman and Sunder [46]).

$$+++ \quad +-- \quad -+- \quad --+$$

Another way to construct an experiment when a full factorial design is not feasible is to design treatments in a way that allows you to make a direct comparison with the baseline. This is advisable when you are primarily interested in the effect of individual focal variables, rather than in the interaction effects. For example, the experiment in Katok and Siemsen [75] uses this design because the experiment contains four focal variables (so the full factorial design would have required 16 treatments, if each was to be varied at two levels). Instead, the authors conducted five treatments:

$$++++ \quad -+++ \quad +-++ \quad ++-+ \quad +++-$$

That investigated the effect of each of the four variables, and compares them to the baseline $(++++)$ one at a time.

Some nuisance variables cannot be directly controlled (for example, subject's alertness). If you have reason to suspect that there may be some nuisance variable present, you can try to eliminate its effect by randomizing. For example, if you believe that subjects who arrive to the lab earlier are better organized and are likely to be more alert, you may try to randomize roles as subject arrive.

A *random block* design holds one or more nuisance variables constant across treatments. An example is a *within-subjects design* that has the same subject participate in more than one treatment. In theory it controls all possible individual differences among subjects since each subject is exposed to each treatment. In practice, however, within subjects design introduces potential *order effect*: the order in which treatments are presented to subjects may matter. One method to deal with the order effect is to randomize the order and then statistically test for the order effect. This may not be ideal, however, if the number of treatments is large because failure to detect order effects does not provide a convincing evidence that they are not there, but only that the design does not have sufficient power to detect them.

A very clever way to use within subjects design but avoid the order effect is called the *dual trial design*. Kagel and Levin [67] used this design when they investigated the effect of the number of bidders in a group on bidding behavior in sealed-bid common-value auctions. Each decision involved an individual, who, upon seeing his private signal, placed two bids, one for the small group, one for the large group. Both decisions were made on the same screen, so order effects were not an issue. At the same time, the design controlled for all individual differences, so differences in behavior could be fully attributed to the number of bidders.

3.2 Context

I will begin with some thoughts on the pros and cons of providing context in experiments. In experimental economics, researchers often describe the experimental tasks to participants using an abstract frame. An abstract frame uses neutral labels for roles and actions. For example, rather than being called "Supplier" and "Buyer" players might be labeled "Mover 1" and "Mover 2", while possible choice might be described in terms of selecting from a set of options, rather than making business decisions, such as selecting prices and quantities.

There are two reasons for using an abstract frame. One reason is to avoid leading the participants by unintentionally (or intentionally) biasing decisions. For example, in an experiment that deals with trust, a participant may have to decide whether to reveal some information truthfully or not. Labeling these actions using loaded language, such as "Tell the Truth" or "Deceive", is likely to result in different behavior than labeling the actions "Option A" and "Option B." While the above example is quite stark, often what might be considered leading is in the eye of the beholder. One researcher may think that the language is neutral, while another researcher (or a referee) may think it is leading. For this reason, using abstract and neutral language is a good practice.

The second reason has to do with a perception that abstract and neutral language somehow makes the experiment more general. If participants are given a specific "cover story," the results are more related to this specific context than to a different context the same basic setting may represent just as well. So one school of thought is that since an abstract frame is equally applicable to different settings, the abstract frame is better.

An alternative way to view an abstract frame, however, is that it is not related to *any* real setting. So rather than being more general, it may be less general, because it applies only to a strange and abstract game, and not to any business situation to which participants can relate. This point brings us to the main downside of using an abstract frame — it makes the experiment more difficult to explain to participants and may result in more confusion, slower learning, and potentially noisier data.

Unfortunately, there is no simple rule of thumb about context, because one thing is certain: context matters a great deal. More generally, there is a great deal of evidence that *framing* (how the problem is described to participants) can have a large effect on behavior [121]. In BOM we tend to have a cover story that is related to the application we are investigating. This is often reasonable because it may increase the external validity of the experiment and link it closer to the real operations setting under investigation. Researchers should take great care, however, in balancing the need for context with unintentional framing and leading.

3.3 Subject Pool

Perhaps one of the first questions people ask about laboratory experiments has to do with the subject pool effect. After all, managers solve business problems; so how valid are results of experiments that use students (mostly undergraduates) as subjects? The first point that is important to emphasize is that laboratory experiments can be conducted with any subject pool. Using students is convenient, but it is not an inherent part of the laboratory methodology. The second point to emphasize is that to the extent that there is any systematic evidence that managers perform any better (or any worse, for that matter) than do students, the differences tend to be observed for very specialized set of tasks, and these are typically not the tasks that participants are asked to perform in controlled laboratory experiments.

There are some obvious practical reasons for using undergraduate students in experiments. Students are readily available on college campuses, so they can be easily recruited to participate in studies. The cost of providing students with sufficient financial incentives to take the study seriously and pay attention is relatively low (for planning purposes I use a figure of $20 per hr.). It is convenient to invite students to physically come to the lab and participate in a study. This procedure makes it easier to make sure that participants do not communicate, and it is also easier, in this setting, to insure that all participants have common information.

In my opinion, the main downside of using managers in experiments is that it is impractical to incentivize them with money. So either the cost of the experiment rises dramatically, or managers are not directly incentivized with money. Depending on the study, having monetary incentives may or may not be critical — I will discuss the importance of incentives in the next section — but the decrease in control that comes from not having incentive compatibility (having the earnings of the participants be directly related to their actions) should be weighted against the possible benefits of having a non-student subject pool.

Does subject pool make a difference? It is quite clear at this point that there is no evidence that managers perform systematically better or worse than students. There are not many studies that systematically considered the subject pool effect; most studies that deal with subject pool do so opportunistically. For example, Katok et al. [76] conducted a set of experiments that examine the effect of time horizons on the performance of service level agreements. They replicated two of the most important treatments in their study with managers (students in an executive education class) who were not incentivized with money, but simply were asked to play the game in order to help the researchers with their study. They report that the only difference between the students' and the managers' behavior is that there is more variability in the manager data than there is in the student data.

Moritz et al. [90] investigate the correlation between cognitive reflection test (CRT) scores and the quality of decisions in the newsvendor problem. They have data for students and managers for one of the treatments in their study, and for that treatment the two subject pools perform qualitatively the same. There are also a few other studies that report no difference between the performance of students and professionals in laboratory experiments [4, 100].

One study that does systematically look at the differences between students and managers is Bolton et al. [12]. In the context of the newsvendor game, the authors compare performance of three subject pools: undergraduate students (called Juniors), masters-level students (called Seniors), and managers in an executive education class (called Managers). In the experiment, subjects made a sequence of newsvendor decisions, and additional information was revealed to them sequentially. Everyone started knowing the price and cost information that they need in order to compute the critical ratio, and were given historical demand information. After 40 rounds (called Phase 1), participants were told that the demand distribution was uniform from 1 to 100. After another 40 rounds (called Phase 2) participants received a tutorial on how to compute the optimal solution, and made the last 20 decisions (called Phase 3).

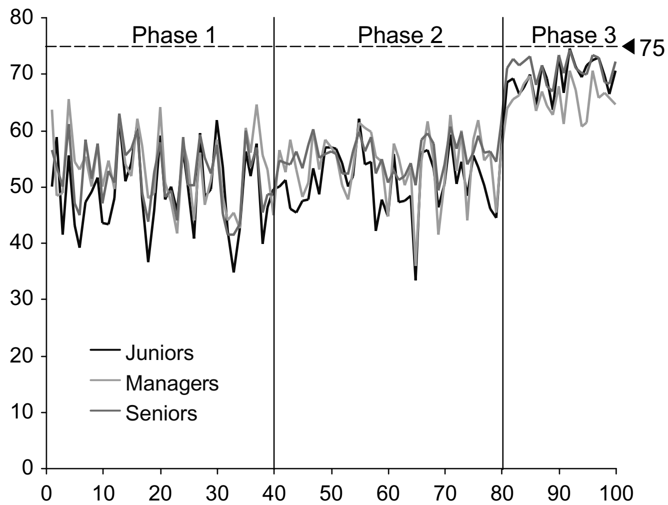


Fig. 3.2 Mean order quantities in the Bolton et al. [12] experiment.

Figure 3.2 summarizes mean order quantities in the Bolton et al. [12] study. All three groups exhibit the pull-to-center effect, and do not exhibit any learning within each phase, and all three groups perform better after the tutorial on computing the optimal order quantity (Phase 3). There is no evidence that Managers perform any better than the other two groups, and in fact, managers perform slightly worse in Phase 3 than do masters students. This study is notable because the experiment is extremely carefully done. The subject pool is the only difference between the treatments — everything else, including the user interface, the instructions, the incentives, was kept identical.

Managers in the study were procurement professionals, and the analysis in the paper controls for their position in the organization, their education, and their years of experience. While there are some intriguing findings related to this demographic information (higher level executives tend to do better, for example) the overall result is that managers do not perform better than students do. This result is a typical one related to the subject pool effect. There is no systematic evidence that student subject pool yields different results than professionals. Therefore, using student participants in laboratory experiments is a good procedure, and is a reasonable first step, unless there are some very specific reasons to believe that professionals are likely to behave significantly and systematically different.

3.4 Incentives

In this subsection I will discuss the role of incentives. Economists use real monetary incentives in their experiments. Smith [116] introduced the idea of *induced-value theory* that explains that using monetary incentives provides a way to gain control over economically relevant characteristics of the laboratory participants. In other words, paying subjects based on their performance in the game causes them to wish to perform better because better performance results in making more money. If the amounts of money subjects earn are significant to them, and if they were recruited using earning money as the incentive (as opposed, for example, to giving course credit for participating), then the participants' innate characteristics become less relevant, and researchers can be more confident that their participants are truly trying to play the game in a way that was meant.

Financial incentives are most convenient, but in principle other types of incentives can be used. The main factor is that the amount of the reward medium earned should be proportional to how well participants perform (as opposed to being given simply for participating). So for example, in theory course credit could be used, as long as the amount of course credit is proportional to the amount of profit made in the game. In practice it is difficult to make course credit given in this way sufficiently salient, though.

There are a number of valid variations in incentive-compatible ways to reward participants. The *binary lottery* procedure involves awarding participants virtual lottery tickets based on their performance — each lottery ticket increases the probability of winning a prize. This procedure has a theoretical advantage of controlling for risk aversion (because regardless of risk preferences, everyone should prefer more lottery tickets to fewer (see Ref. [105])), but a practical disadvantage of being less straightforward than simply paying money.

Another variation is to pay for one or several randomly chosen rounds instead of the average for all rounds. Neither method can be said to be clearly better, so it is a matter of preference which payment method is used.

A more important practical question is to what extent using real incentives matters. Much of important and influential work has been based on experiments based on hypothetical choices [72], and experiments that use hypothetical choices are accepted in many branches of social science, such as psychology, marketing, and management. Sometimes behavior in hypothetical situations does not differ from behavior in real situations, but sometimes it does differ. I will discuss two studies that directly address this issue.

Forsythe et al. [45] investigate the reasons for more equitable distribution in the Ultimatum game [51] than the subgame perfect equilibrium prediction. The authors consider two alternative hypotheses for equitable distributions: (1) proposers are trying to be fair to responders, or (2) proposers make large offers because they realize that responders are likely to reject offers that are too small. In order to be able to distinguish between the two hypotheses, the authors conducted some treatments with a modification of the Ultimatum game, called the Dictator game; the only difference being that in the Dictator game responders cannot reject offers — they have to simply accept whatever (if any) offer the proposer chooses. If equitable distribution is driven primarily by the proposers' desire to treat responders fairly, the offers in the Ultimatum and the Dictator games should not differ. But if it is the fear of being rejected that drives equitable offers, then offers in the Dictator game should be significantly lower.

The authors conducted their two games (Ultimatum and Dictator) under two different payment conditions: real and hypothetical. Figure 3.3 displays histograms of offers in the four treatments on the Forsyth et al. [45] study. Each treatment included two separate sessions (April and September) and within each treatment the distributions for April and September do not differ.

The striking point is that the distributions of offers without pay are not different for the Ultimatum and the Dictator games (compare Figure 3.3(c) and (d)), while with pay they are strikingly different (compare Figure 3.3(a) and (b)). In other words, proposers are quite generous with hypothetical money, but not with real money. Had this

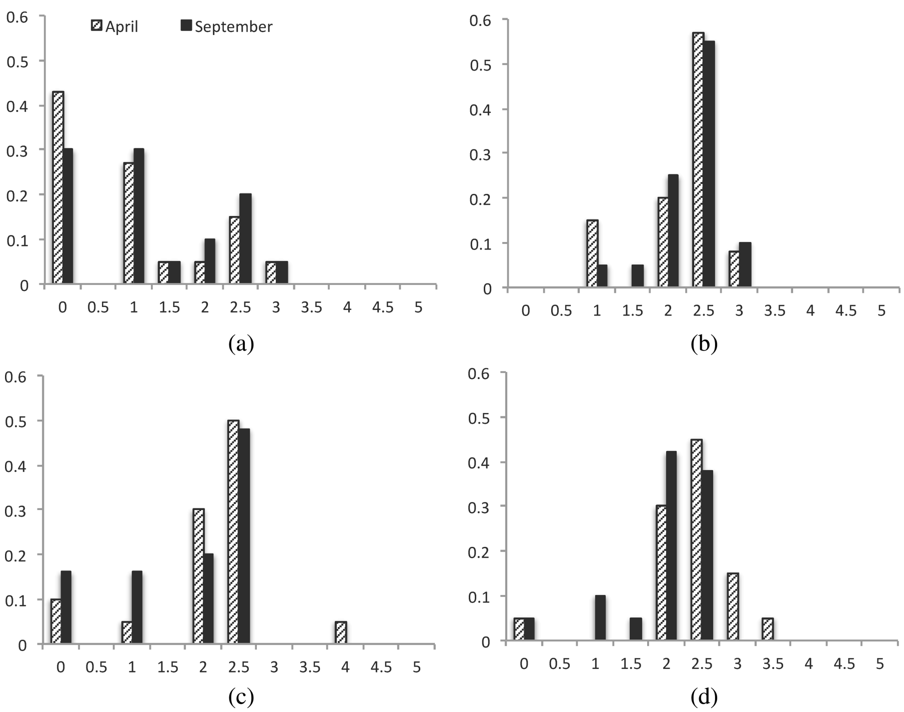


Fig. 3.3 Distribution of offers in the Forsythe et al. [45] study. (a) Dictator Game with Pay, (b) Ultimatum Game with Pay, (c) Dictator Game without Pay and (d) Ultimatum Game without Pay.

study been conducted without real incentives, the researchers would have drawn incorrect conclusions about the underlying causes for equitable distributions in the Ultimatum game.

Another well-known study that directly compares real and hypothetical choices is by Holt and Laury [59]. The authors study the effect of the magnitude and real vs. hypothetical incentives on risk preferences. The instrument they use to elicit risk preferences is presented in Table 3.1.

Participants are asked to make a choice between the Option A and Option B lottery in each row. The Option A lottery is safe, while the Option B lottery is risky. But as we move down the rows, the probability of a high payoff in the Option B lottery increases (and becomes certain in the 10th row). A risk neutral subject should switch from Option A

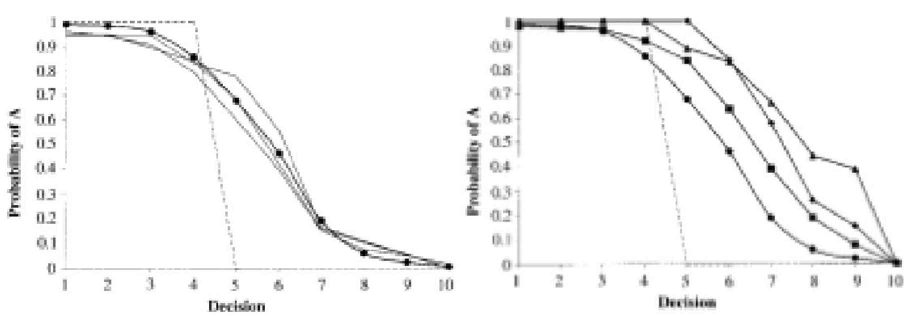
Table 3.1. The instrument to elicit risk preferences in Holt and Laury [59].

Option A	Option B	Expected payoff difference
1/10 of \$2, 9/10 of \$1.60	1/10 of \$3.85, 9/10 of \$0.10	\$1.17
2/10 of \$2, 8/10 of \$1.60	2/10 of \$3.85, 8/10 of \$0.10	\$0.83
3/10 of \$2, 7/10 of \$1.60	3/10 of \$3.85, 7/10 of \$0.10	\$0.50
4/10 of \$2, 6/10 of \$1.60	4/10 of \$3.85, 6/10 of \$0.10	\$0.16
5/10 of \$2, 5/10 of \$1.60	5/10 of \$3.85, 5/10 of \$0.10	−\$0.18
6/10 of \$2, 4/10 of \$1.60	6/10 of \$3.85, 4/10 of \$0.10	−\$0.51
7/10 of \$2, 3/10 of \$1.60	7/10 of \$3.85, 3/10 of \$0.10	−\$0.85
8/10 of \$2, 2/10 of \$1.60	8/10 of \$3.85, 2/10 of \$0.10	−\$1.18
9/10 of \$2, 1/10 of \$1.60	9/10 of \$3.85, 1/10 of \$0.10	−\$1.52
10/10 of \$2, 0/10 of \$1.60	10/10 of \$3.85, 0/10 of \$0.10	−\$1.85

in row 4 to Option B in row 5, but the more risk averse participants may switch later. Eventually every participant should prefer Option B in the 10th row.

Holt and Laury [59] vary the magnitude of the stakes by conducting treatments with payoffs in Table 3.1 multiplied by the factors of 20, 50, and 90. They also conduct each treatment with real as well as hypothetical stakes.

Figure 3.4 shows the summary of the proportion of participants choosing Option A in each treatment. More risk averse individuals should choose more Option A's. The key finding is that behavior looks



(a) (b)

Fig. 3.4 Summary of Holt and Laury [59] data. (a) Low real payoffs [solid line with dots] compared with hypothetical payoffs [thin lines] and risk-neutral benchmark [dashed line] and (b) Real payoffs: low [solid line with dots], 20× [squares], 50× [diamonds], 90× [triangles], risk neutral [dashed].

very similar for small stakes real choices and for hypothetical choices, and the size of the stakes does not seem to matter much with hypothetical choices (Figure 3.4(a)). However, risk aversion levels increase as real stakes increase (Figure 3.4(b)).

There are other types of decisions, however, for which there is no evidence that real vs. hypothetical payments make a difference. For example, in the newsvendor experiments, the Schweitzer and Cachon [114] study had virtually no real incentives, but when Bolton and Katok [10] replicated the study with carefully controlled incentives they found no discernable difference in behavior.

Another issue that should be seriously considered is that without real incentives participants may pay less attention and the resulting behavior may well be noisier. So if a study is being conducted without real incentives, it is particularly important that there is some other mechanism to insure that participants take their decisions seriously. For example, participants might be asked for their help with research and be given an option to opt out. Those are, admittedly, weak manipulations — providing real incentives is, as a rule, better.

In summary, we can conclude that providing real incentives can matter a great deal. Decisions that involve social preferences or risk are definitely affected by real incentives. Decisions that involve more straightforward optimization tasks seem to be less affected by real incentives. It is not always clear a priori whether real incentives will matter or not. Therefore, initial experiments should be conducted with real incentives, until it has been systematically shown that behavior is not affected by hypothetical incentives.

3.5 Deception

The last methodological topic I will discuss is the use of deception.What constitutes deception ranges from deliberately providing subjects with false information, to not specifically stating some information, allowing subjects to draw their own, perhaps incorrect, conclusions. An example of the former is telling participants that they are interacting with another human participant, while in fact they are interacting with a computer. An example of the latter might be inviting 32 participants into the lab, matching them repeatedly in four groups of eight, but only telling them that each period they are randomly matched with another person in the room (a technically true, but slightly misleading statement). While both are examples of deception, the former is definitely considered unacceptable by experimental economists, while the latter is not.

Davis and Holt [31] cite the loss of experimental control, as the primary reason deception is considered unacceptable.

> Most economists are very concerned about developing and maintaining a reputation among the student population for honesty in order to ensure that subject actions are motivated by the induced monetary rewards rather than by psychological reactions to suspected manipulation. (pp. 23–24).

There are two ways experimental control might suffer due to the use of deception: indirect and direct. Participants who have experienced deception in previous experiments may not trust the experimenters in future, unrelated experiments. Thus, the use of deception by a few researchers might (indirectly) contaminate the entire subject pool. There is also potentially a direct loss of control, because when subjects are being deceived in a study, they may (correctly) suspect that they are being deceived. After all, a reason to deceive subjects in the first place is to investigate phenomena that may not naturally occur without deception. It is difficult to assess the direct effect of deception, but generally speaking, since deception diminishes control, it is better to try to design experiments without using deception.

The use of deception is common in psychology. In their review article, Ortman and Hertwig [95] report that more than 1/3 of studies published in psychology use deception. Even more importantly, studies that use deception are routinely studied in undergraduate psychology courses. Since the subject pool for psychology studies typically comes from the population of undergraduate psychology majors, these participants are generally aware that they are likely to be deceived, and they tend to expect this. Moreover, the type of deception psychology studies use often includes directly deceiving subjects about the purpose of the experiment, or using confederates, and investigating resulting behavior. Jamison et al. [63] provide the following typical example of deception in a psychology study: "... subjects were given two poems by Robert Frost, were told that one was by Frost and one by a high school English teacher, and were then asked to rate the merits of the two poems. After the experiment they were debriefed and told that both poems were actually by Frost and that the experiment was looking at how beliefs regarding authorship affected the rating of the poems." (p. 478).

In contrast, experimental economists almost never use deception, so participants in economic experiments do not generally expect to be deceived. In a few studies that used deception, tend to use it for convenience, or to study reactions to behavior that is unlikely to occur naturally. This effect is usually achieved by telling subjects that they are matched with a human participant, while they are in fact matched with a computer agent programmed to behave in some specific way [8, 110, 111, 126, 127].

There are some studies that have investigated indirect effects of deception. Jamison et al. [63] is perhaps the most direct study. The authors conducted an experiment that consisted of two parts. During the first part, participants played the trust game.[1] Half of the participants were not deceived, and the other half were deceived in that they were told that they were matched with a human participant, while in fact they were matched with a computer programmed to imitate the behavior of human participants in earlier studies. The deceived participants were debriefed at the end of the study and told that they were matched with computerized partners.

Three weeks later the authors conducted the second phase of the study, for which they invited the same group of participants to an experiment that looked unrelated. This second experiment involved a Dictator game, a risk version assessment task similar to Holt and Laury [59], and a prisoner dilemma game. Jamison et al. [63] analyzed

[1] In the trust game the first mover must decide on the fraction x of her initial endowment to pass to player 2. This fraction triples, and player 2 decides on the fraction y of the amount to return to player 1.

the effect of having been previously deceived on participation rates in the second study and on the behavior in the second study.

Jamison et al. [63] report that deception does have an effect on participation as well as behavior. Females who have been deceived are significantly less likely to return than the females who have not been. Also, participants who were unlucky and have been deceived are less likely to return than the participants who have been unlucky but have not been deceived. In terms of behavior, participants who have been deceived behave more erratically (less consistently) in answering the risk aversion questions, indicating that they may not be taking the study as seriously. The only other difference between deceived and not deceived participants is that females or inexperienced subjects who have been deceived, and who had the role of the first moved in the trust game, tend to give less in the Dictator game.

One may argue that the evidence we have so far indicates that the indirect effects of deception in terms of damaging the subject pool seem to be fairly minor. It may be that the true costs are actually higher, because the participants in the Jamison et al. [63] study came from the economics subject pool, so they were students who were not previously deceived. A single deception incident may not have significantly changed their behavior, but it may be that repeatedly deceiving participants will alter the characteristics of the subject pool in more serious and permanent ways (see [106] for a related argument).

3.6 Infrastructure and Logistics

Some of the infrastructure and logistic requirements needed to conduct laboratory experiments include funding to pay the participants, and efficient way to recruit those participants, the approval for the use of human subjects, that is required by US universities, the software to implement the game, and a computer lab to run the experiment.

Laboratory experiments tend to be relatively inexpensive compared, for example, to experiments conducted in natural or physical sciences. Many research-oriented universities provide small grants for data collection that is often sufficient for a study with a reasonable sample size.

Subject recruitment is most efficiently done through the internet, and several recruitment systems have been developed and are freely available for academic use (e.g., ORSEE software, developed by Ben Greiner, can be accessed from this URL: http://www.orsee.org/).

Human subject approval typically requires providing information about your study and your recruitment process to an office on campus (this is usually the same office that reviews medical studies to insure that human subjects are not subjected to major risks). Studies involving laboratory experiments in social studies also have to be reviewed by the Institutional Review Board (IRB) or a similar regulatory body. Check your university rules about obtaining human subject approval for studies before you start your work.

The majority, although not all, of these experiments are conducted using a computer interface. Computer interface is a convenient and efficient way to collect data, but the downside is that implementing even a simple game may require a significant amount of work. Fortunately, Urs Fischbacher developed a platform called z-Tree (Zurich Toolbox for Readymade Economic Experiments) for implementing laboratory experiments (http://www.iew.uzh.ch/ztree/index.php). This software is freely available to academic researchers. It has a fairly intuitive structure and syntax that is easy to learn even for a person with modest programming skills. And it has a good tutorial, a wiki, and an active user listserv. z-Tree is designed to be used in a computer lab on computers networked using a LAN, so it is not ideal for the use over the internet, but this is perhaps its only limitation. It is flexible enough to create experimental software quickly, and even includes some advanced GUI features, such as graphs and chat boxes (see [42]).

z-Tree can be easily installed in any computer lab, so a dedicated lab, although convenient and useful to have, is not essential. If you are fortunate enough to be given access to a dedicated lab, some useful features to have are privacy partitions for subject computers and an overhead projector. Larger labs are more convenient because they facilitate larger sessions, making data collection more efficient.

Using computer interface to conduct laboratory experiments is a common practice, and the ease and availability of the z-Tree software contributed to this trend. Computer interface usually increases control and greatly simplifies the logistics of collecting data. Occasionally, though, designs require experiments to be conducted by hand. One example of this need to run an experiment without a computer is a paper [14], that studies the effect of participant–experimenter anonymity. The authors needed a transparent way to insure not only true anonymity, but also the perception of anonymity in a type of a simplified Ultimatum game. They achieved this effect by conducting the experiment by hand, that involved participants passing marked boxes back and forth with the aid of several research assistants.

4

Inventory Ordering with Stochastic Demand: The Newsvendor Problem

4.1 The "Newsvendor" Problem in the Theory

The Newsvendor problem is a fundamental building block for most of the inventory theory [101]. It is a normative model, created not to describe how people behave, but how they *should* behave. It is a part of most models dealing with the effectiveness of supply chain incentive systems [18]. It is also considered a starting point for dealing with more analytically challenging environments, such as concurrently setting prices and ordering inventory [99], or ordering inventory in competitive settings [83]. A common assumption in all these models is that the newsvendor will act optimally to maximize her expected profit. The missing link in the analytical modeling literature is the question of whether decision-makers do order optimally, and if not, then how to induce the optimal ordering behavior.

The newsvendor problem, as the name implies, is a single-period model in which the decision-maker places an order Q before knowing the actual demand for the period D. Each unit is sold at a price of p and costs c. If the amount ordered, Q, exceeds D, then exactly D units are sold, and $Q - D$ units are discarded. If D exceeds Q then Q units are sold and potential profit for $D - Q$ units is foregone.

If D is a random variable with distribution function F and density function f, the profit when Q is ordered and the demand is D can be written as:

$$\pi(Q, D) = p \min(Q, D) - cQ \tag{4.1}$$

and expected profit is

$$E[\pi(Q, D)] = (1 - F(Q))(p - c)Q + \int_0^Q f(x)\pi(Q, x)dx. \tag{4.2}$$

It is well-known that the order quantity Q^* that maximizes the average profit must satisfy

$$F(Q^*) = \frac{p - c}{p}, \tag{4.3}$$

the ratio $(p - c)/p$, known as the *critical fractile*. Intuitively, at optimality, the decision-maker is just indifferent in terms of expected profit between being one unit short and one unit over. Solving the newsvendor problem is not an easy task for an uninformed subject. It requires some understanding of what a demand distribution is, and the notion of independent draws; it requires one to understand the marginal analysis of overage and underage costs, etc. Thus, the newsvendor problem is a target-rich environment for behavioral research.

4.2 The "Newsvendor" Problem in the Laboratory

The recent literature on the newsvendor behavior in the laboratory is aimed at testing this assumption and identifying explanations for deviations from optimal behavior. Some papers also consider how to improve performance through additional education or manipulating feedback. An important practical question is how to design better supply chain contracts that considers the better model of behavior.

The seminal paper in BOM investigating human behavior in solving the Newsvendor problem is Schweitzer and Cachon [114]. They examined a condition in which the optimum inventory order was above average demand (critical fractile of 0.75) and one in which the optimum order was below average demand (critical ratio of 0.25). The game was repeated and subjects were provided feedback on realized demand and

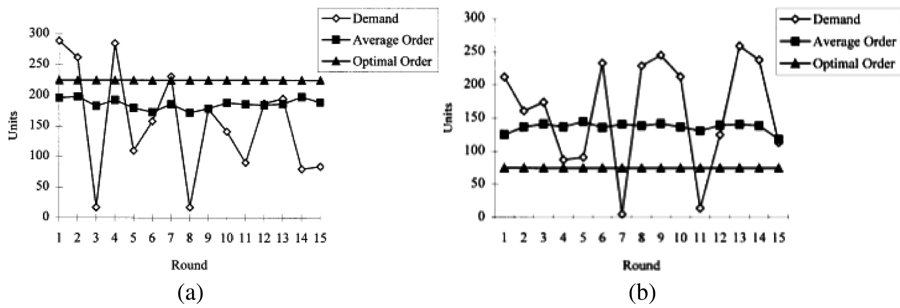


Fig. 4.1 Summary of the average orders and comparison with optimal order and average demand in the Schweitzer and Cachon [114] study. (a) High Profit Condition and (b) Low Profit Condition.

profitability at the end of each period. The subjects also had access to a spreadsheet that could be used to calculate profit distributions for different orders.

Figure 4.1 summarizes the main results in Schweitzer and Cachon [114]. The key observation is that in the setting with the critical fractile of 0.75, average orders are below optimal, but in the setting with the critical ratio of 0.25, average orders are above optimal. So average orders are between the optimal level and the average demand. This pattern has been termed the "pull-to-center" effect.

Schweitzer and Cachon [114] proceeded to check whether the pull-to-center effect is consistent with some of the commonly-used preferences. They found that the pattern is not consistent with risk aversion, loss aversion, prospect theory, underestimating opportunity cost, waste aversion, or stockout aversion. The pattern is consistent with minimizing ex-post inventory error, anchoring on mean demand and insufficiently adjusting toward the optimal order level, as well as on chasing prior demand (anchoring on prior order and adjusting in the direction of the prior demand).

A number of studies demonstrated that the pull-to-center effect persists in the Newsvendor problem with a variety of parameters and different feedback information. Recall that we discussed the study by Bolton et al. [12], who showed that the behavior does not change when decision-makers are professional procurement managers.

The newsvendor problem is complicated for a decision-maker not trained in analytical modeling, so the fact that participants are not solving the problem correctly is not surprising in and of itself. There are two features of learning identified from the behavioral decision literature: (1) people have limited information processing capacity, and (2) people are adaptive (see [56] and references therein). A major consequence of limited information processing capacity, critical for the pattern of adaptation, is that *information tends to be absorbed selectively and sequentially.* Bolton and Katok [10] systematically enhance feedback and experience factors known to be important in other decision-making contexts, in an attempt to isolate the features of the decision support that would facilitate the ability of participants to solve the newsvendor problem correctly.

They systematically improve information and feedback provided to participants through (1) limiting the number of ordering options, (2) providing information on the performance of foregone decisions, and (3) constraining decision-makers to placing standing orders.

There are two notable results in the Bolton and Katok [10] study. First, decreasing the number of ordering options to as few as three does not improve performance relative to the standard setting with 100 ordering options. This remarkable lack of improvement persists in both, high and low profit conditions, as well as in a setting in which participants are given feedback on how the options they did not choose performed, in addition to how the option they did take performed. Second, the intervention that does help improve performance is constraining decision-makers to placing ten-period standing orders. This intervention works significantly better than providing information about the expected profitability of each of the three options to participants upfront.

The authors conclude that learning-by-doing is an effective way to induce optimal ordering because this intervention forces decision-makers to take the longer view rather than being focused on the short-term demand fluctuations. An implication of this finding is that effective procedures and incentives for inventory decision-makers should focus on long-term trends, restraining responses to short-term fluctuations. Knowledge gained from hands-on experience is more readily utilized than knowledge gained from a third party source. It follows that sensible restriction of the options before the decision-maker enables more targeted feedback and hence more effective learning-by-doing.

4.3 Can Errors Explain it All?

Su [119] proposed an elegant explanation of the pull-to-center effect. The basic idea that he terms *bounded rationality* is that faced with a number of decision options, a boundedly rational decision-maker chooses option i with probability

$$\psi_i = \frac{e^{u_i/\tau}}{\sum_i e^{u_i/\tau}}, \tag{4.4}$$

where u_i is the utility of option i, and τ is a measure of rationality, called the coefficient of certitude. Note that as β increases, ψ_i approaches uniform random choice, while when $\tau = 0$, the optimal (highest utility) option is chosen with certainty. When the demand in the Newsvendor problem is Uniform from a to b, the behavioral optimal newsvendor solution is

$$Q \sim N\left(p - \frac{c}{p}(b-a), \tau\frac{b-a}{p}\right). \tag{4.5}$$

For example, if we use the Bolton and Katok [10] experimental parameters with $D \sim U(0,100), p = 12, c = 3$ (high profit condition) or $D \sim U(50,150), p = 12, c = 9$ (low profit condition), the resulting optimal order would be normally-distributed, with the mean of 75 and the standard deviation of 8.33β.

If τ is not very small, then the probability that Q is either below a or above b is quite high. In the laboratory, we do not generally observe orders that are either below the minimum or above the maximum demand, so an additional assumption that is required is that the distribution of Q is truncated at a and b.

Figure 4.2 illustrates the plot of the mean order implied by Equation (4.5) and truncated at a and b, as a function of τ, and we can see that indeed, for example, when $\tau = 5$, mean orders match the average orders observed in the laboratory.

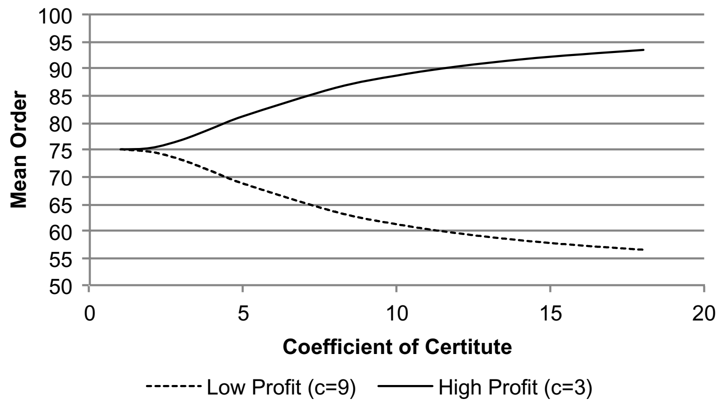


Fig. 4.2 Mean order as a function τ implied by the Su [119] model.

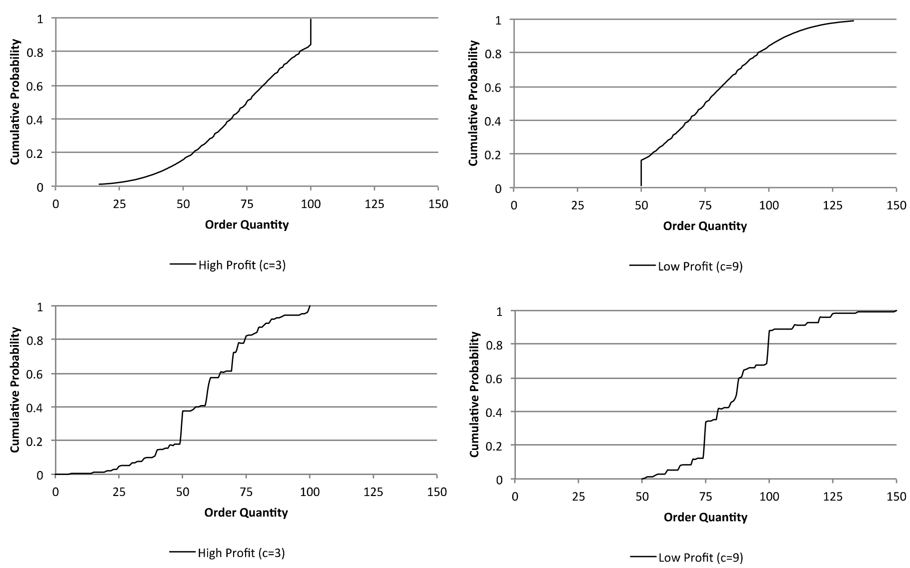


Fig. 4.3 Cumulative distributions of orders implied by the Su [119] model for $\beta = 5$ (top panel) and the actual cumulative distributions of orders in Bolton and Katok [10] (bottom panel).

But if we plot the cumulative distribution of order quantities, which is illustrated in the top panel of Figure 4.3, we see that this pull-to-center effect is due entirely to truncation.

The bottom panel of Figure 4.3 shows the actual cumulative distributions of order quantities in Bolton and Katok [10]. It is apparent from the figure that actual cumulative distributions do not exhibit a mass at either end point; errors do not fully explain the behavior in newsvendor experiments.

Kremer et al. [79] also investigate the question of "do random errors explain newsvendor behavior?" but do so with a direct laboratory test. They restrict the Newsvendor problem to seven demand realizations, and the seven corresponding order quantities, so that the resulting payoff table that includes profits for each possible order and demand realization in the low and high profit conditions are shown in Table 4.1.

Kremer et al. [79] conducted their experiments in the Operations frame and the Neutral frame.Operations frame is the standard newsvendor problem frame that describes the problem in terms of placing the order, and the profit depending on the demand realization, as well as

Table 4.1. Low and High profit conditions in the Kremer et al. [79] study. The optimal order is in **bold**.

(a) Low Profit

Order/Alternative	Demand/State 500	550	600	650	700	750	800
500	7.8	7.8	7.8	7.8	7.8	7.8	7.8
550	**4.9**	**8.6**	**8.6**	**8.6**	**8.6**	**8.6**	**8.6**
600	2.1	5.7	9.4	9.4	9.4	9.4	9.4
650	−0.8	2.9	6.5	10.1	10.1	10.1	10.1
700	−3.6	0.0	3.6	7.3	10.9	10.9	10.9
750	−6.5	−2.9	0.8	4.4	8.1	11.7	11.7
800	−9.4	−5.7	−2.1	1.6	5.2	8.8	12.5

(b) High Profit

Order/Alternative	Demand/State 300	400	500	600	700	800	900
300	4.2	4.2	4.2	4.2	4.2	4.2	4.2
400	3.8	5.6	5.6	5.6	5.6	5.6	5.6
500	3.4	7.2	7.0	7.0	7.0	7.0	7.0
600	3.0	4.8	6.6	8.4	8.4	8.4	8.4
700	2.7	4.4	6.2	8.0	9.8	9.8	9.8
800	**2.3**	**4.1**	**5.8**	**7.6**	**9.4**	**11.2**	**11.2**
900	1.9	3.7	5.5	7.2	9.0	10.8	12.6

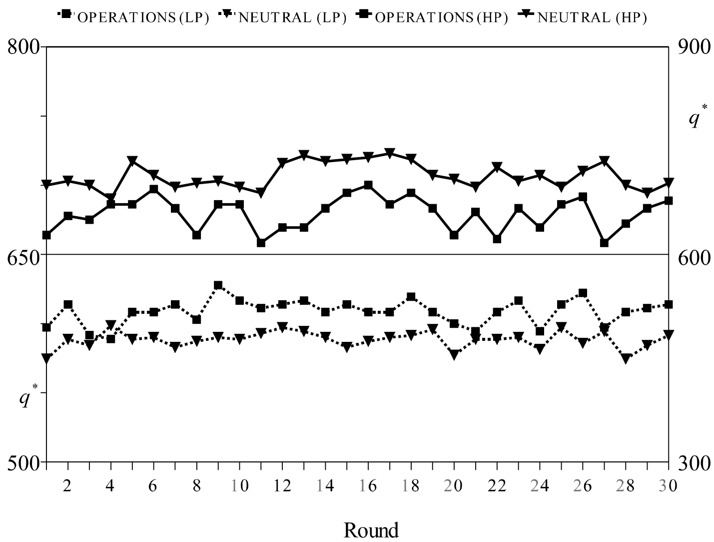


Fig. 4.4 Average orders over time in the Kremer et al. [79] study.

on the production cost and the selling price. In the Neutral frame the game was simply presented in terms of seven states of the world and seven alternative actions (so the rows in the payoff table were labeled "Alternative" instead of "Order").

Average orders over time in the Kremer et al. [79] study are shown in Figure 4.4.

Average orders in the Neutral frame are closer to optimal in both, high and low profit conditions; this result cannot be accounted for by the Su [119] errors model, since the errors explanation is invariant to the frame.

A full explanation for the Newsvendor behavior is proving to be elusive. It is likely that there is no single explanation. Bostian et al. [15] estimate an experience-weighted attraction (EWA) learning model, that seems to capture some of the dynamics. Moritz et al. [90] attempt to find an explanation grounded in cognitive psychology. They report that individuals who score higher on the cognitive reflection test (CRT) also perform better in the high profit condition, and they chase demand less. The correlation is not significant in the low profit condition. Rudi and Drake [109] report that participants react to being able to observe the missed sales by placing higher orders than when this information is unavailable (as it is likely to be in practice). Croson et al. [27] propose the overconfidence bias as the explanation. Some of the other Newsvendor studies include Feng et al. [40] who report on the experiments conducted in China; Ho et al. (2010), who report on the multi-location version of the problem, Gavirneni and Xia [48], who consider different kinds of anchors that decision-makers use; Chen et al. [23], who look at the effect of payment timing; Brown and Tang [16], Benzion et al. [6]; Schultz et al. [113], who look at the positive and negative frame; Lurie and Swaminathan [87], who look at the effect of information frequency; and Gavirneni and Isen [47], who examine the verbal protocol. Deng and Shen [33] propose a model to explain asymmetries in the pull-to-center effect. Sebastian et al. [112] frame the problem in terms of costs and in terms of profits, and find that average orders are always higher with the profit frame. They conclude that subjects perceive penalty costs as being higher than opportunity costs.

4.4 Closing the Loop

While explanations for the behavior seem to be elusive, Operations Management is a practical discipline, so what we learned in the laboratory can be applied in practice to make better decisions and design better contracts. The simplest contracting setting involves a supplier/retailer channel in which the supplier proposes a contract to the retailer. The supplier needs to know how the retailer will react to any given contract in order to design a contract that's best (for example most profitable). The research reviewed so far shows that human subjects who take on the role of retailers do not make decisions consistent with profit maximization, but do behave in a predictable way at an aggregate level. It follows that the supplier should be able to do better with the contract that considers the actual ordering behavior than with a contract that assumes that the retailer orders based on the critical ratio.

A study by Becker-Peth et al. [5] is an attempt to demonstrate how behavioral contracts might be designed. They consider the buyback contract, in which the retailer pays the supplier a wholesale cost w

for the Q units ordered, and the supplier returns to the retailer a rebate of b for all unsold units. The analytical solution involves the critical ratio that takes into account the fact that the cost of overage is decreased by b.

The authors begin with three preliminary experiments that highlight three regularities of newsvendor behavior: (1) average orders depend on contract parameters, and do not necessarily fall between mean demand and optimal order; (2) in the buyback contract, subjects do not value revenue from sales and revenue from the rebate the same way; and (3) most subjects reframe the problem and base their decision on the profit margin and the cost of unsold products — components of the critical fractile — the cost of unsold units is perceived as a loss. Based on these three observations, they proceed to formulate a behavioral model:

$$Q = (1-\alpha)F^{-1}\left(\frac{p-w}{(p-w)+\beta(w-\gamma b)}\right) + \alpha\mu. \qquad (4.6)$$

The model has three behavioral parameters: α controls how far the person biases the order toward mean demand, γ controls how different the person values income from returns than income from sales, and β is the loss aversion parameter.

In the experiment each participant faced 28 different one-shot newsvendor problems with different levels of w and b. Figure 4.5 shows the 28 combinations and the optimal orders that correspond to each (left panel), as well as a comparison of the optimal order to the actual average order (right panel).

Figure 4.5(b) clearly shows that average orders depend not only on the critical ratio, but also on the actual values of w and b. The Maximum Likelihood Estimation shows that the aggregate values for the behavioral parameters are $\alpha = 0.27, \beta = 2.2$, and $\gamma = 1.03$.

Figure 4.6 shows that the distribution of orders also depends on the values of w and b, and there is a fair amount of individual heterogeneity in orders, so the behavioral model can also be estimated for each individual.

Becker-Peth et al. [5] proceed to show with an out-of-sample test that contracts designed based on Equation (4.6) that perform

Newsvend or orders							
	w						
b	5	20	35	50	65	80	95
0	95	80	65	50	35	20	5
15	.	94	76	59	41	24	6
30	.	.	93	71	50	29	7
45	.	.	.	91	64	36	9
60	.	.	.	.	88	50	13
75	.	.	.	.	.	80	20
90	.	.	.	.	.	.	50

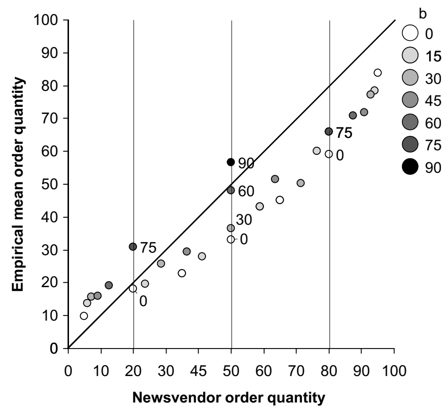


Fig. 4.5 Optimal and average orders in the Becker-Peth et al. [5] experiment. (a) Optimal Orders and (b) Comparing Average Orders to Optimal.

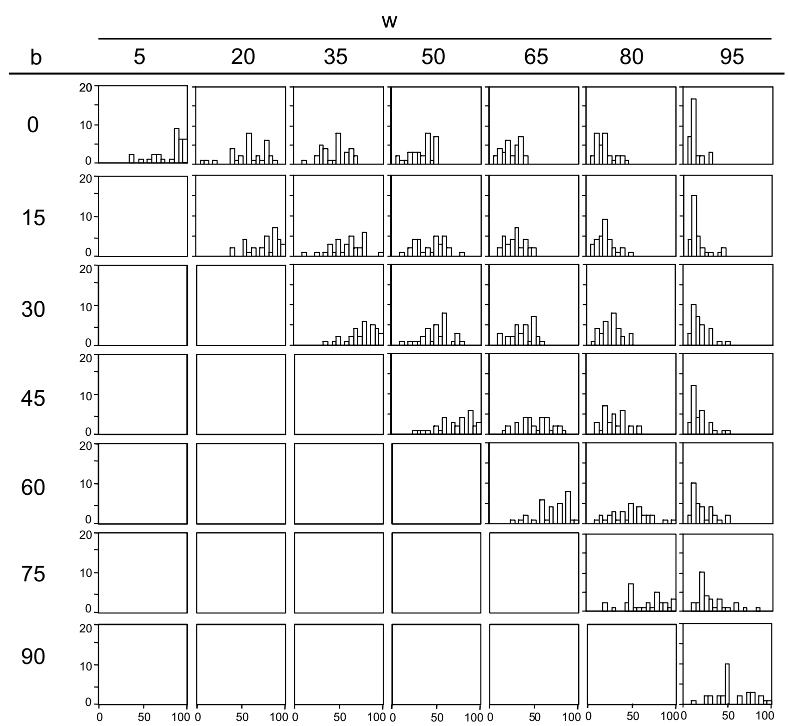


Fig. 4.6 The distribution of orders for each combination of w and b in the Becker-Peth et al. [5] study.

significantly better than the contracts designed based on the profit maximization assumption.

While the Becker-Peth et al. [5] study may have some limitations (such as the use of student subjects in the experiments) it is one of the first studies that attempts to close the loop between analytical modeling work and the practical problem that motivated it.

5

Supply Chain Coordination

5.1 Laboratory Tests of Channel Coordination with Random Demand

The newsvendor problem is often a building block in the channel coordination problem, because the firm that faces random demand is often modeled as a newsvendor. Katok and Wu [78] report on the first laboratory tests of these models. They consider a setting in which the retailer is the newsvendor and the supplier proposes a contract (this is also called a "push" contract because the supplier "pushes" the product to the retailer). Since the wholesale price contract results in the double marginalization problem, it is worthwhile to implement more complicated contracts that can, in theory, coordinate the channel by inducing the retailer to place the first best order. One way to do this is for the supplier to assume some of the risk associated with the random demand [17]. Katok and Wu [78] investigate two such risk sharing contracts, the buyback and the revenue sharing, and compare their performance to the wholesale price contract.

I already introduced the buyback contract in Section 4.3 as part of the discussion of the Becker-Peth et al. [5] study. Under the revenue sharing contract, the retailer pays the supplier w per unit ordered, and an additional amount r for each unit sold. The buyback (BB) and the revenue sharing (RS) contracts can be thought of in terms of the per-unit cost to the retailer for each unit sold and each unit unsold. These contracts can be designed to be mathematically equivalent when their per-unit costs of sold and unsold units are the same, as follows:

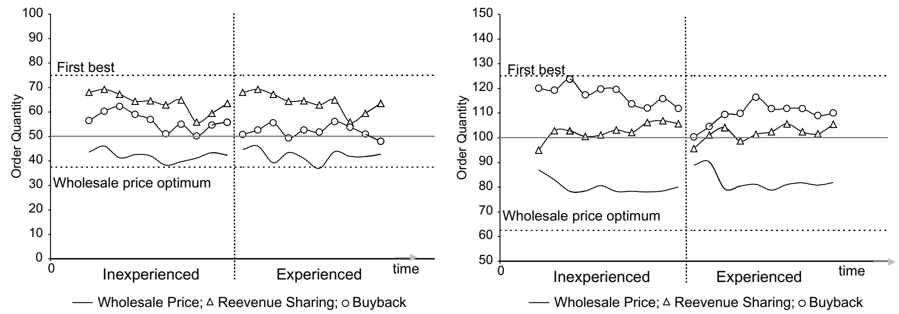
$$\begin{aligned} &\text{Per-unit cost of sold units: } w_{BB} = w_{RS} + r \\ &\text{Per-unit cost of unsold units: } w_{BB} - b = w_{RS}. \end{aligned} \tag{5.1}$$

Katok and Wu [78] compare these contracts to one another and to the wholesale price contract (for which the cost of sold and unsold units is simply w) from the point of view of the retailer (called the Retailer game) responding to a fixed contract from an automated supplier, and the supplier (called the Supplier game) who faces an automated retailer programmed to order optimally. This study does not consider two human individuals because the stated purpose of the study is to observe how decision-makers set contract parameters and respond to contract parameters, without confounding their actions with concerns about social preferences, such as fairness. The next section discusses the role of social preferences in contracting.

Katok and Wu [78] look at the two demand conditions (high and low) in order to be able to investigate potential loss aversion. They also look at the role of experience and framing. The behavior in the Retailer game is largely consistent with earlier work in that there is the pull-to-center effect in the average orders for all three contracts. This makes the two coordinating contracts less efficient than they should be. Figure 5.1 shows how average orders evolve over time. An interesting observation is that for the two coordinating contracts, average orders sometimes move toward the optimal order (high demand inexperienced subjects revenue-sharing contract, and experienced subjects revenue sharing and buyback contracts) and sometimes they move away from the optimal order (low demand, and high demand inexperienced subjects, buyback contract).

Orders under the buyback and the revenue sharing contract also appear different, but the differences largely disappear with experience.

Figure 5.2 shows the average retailer orders induced in the Supplier game. The main results from the Supplier game are (1) suppliers



(a) (b)

Fig. 5.1 Average retailer orders over time in Katok and Wu [78]. (a) Demand $\sim$ U(0, 100) and (b) Demand $\sim$ U(50, 150).

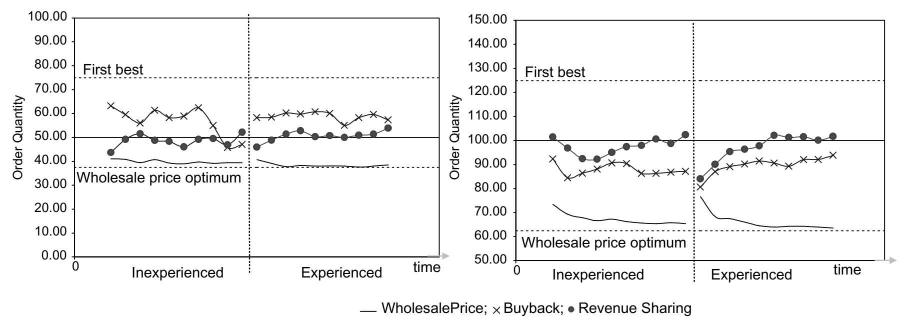


Fig. 5.2 Average retailer orders induced in the Supplier game in Katok and Wu [78]. (a) Demand$\sim$ U(0, 100) and (b) Demand$\sim$ U(50, 150).

set nearly-optimal wholesale prices in the simple wholesale price contract, and (2) supplier do not take nearly enough risk under the two coordinating contracts (meaning that they set the rebate parameter in the buyback contract too low, and the wholesale price in the revenue-sharing contract too high) and as a result do not come close to being able to coordinate the channels. The finding that coordinating contracts do not perform well, even when retailers are programmed to act optimally, is an important one because it suggests that there may be significant potential gains from better decision support for designing contracts.

Davis [29] reports on a study that takes a similar approach to Katok and Wu [78] but looks at "pull" contracts. Under a pull contract, the supplier decides on the production quantity and ships products to the retailers as demand is realized. So the supplier is the newsvendor in this setting, while the retailer designs the contract to induce the supplier to produce enough (see Cachon [18] for an analytical treatment).

Davis [29] compares the "pull" wholesale price contract to a contract he terms "overstock allowance", as well as a service level agreement (SLA). Under the overstock allowance contract (which is analogous to the buyback contract, but in the pull setting), the retailer pays the supplier for each unit sold, as well as a separate premium for each unit that is not sold (when units exceed demand). Under the SLA (which is one of the most common pull contracts used in practice), the retailer pays the supplier a bonus whenever the production quantity does not fall short of realized demand.

In his experiment, Davis [29] looks at retailers matched with suppliers programmed to produce optimally. He finds that the pull wholesale price contracts performs close to optimal (this is consistent with the Katok and Wu [78] results about the push wholesale contract in the Supplier game). But just as suppliers in the Katok and Wu [78] study do not take enough risk, retailers in the Davis [29] study do not take enough risk either. Davis proposes a simple behavioral model based on regret with respect to the coordinating parameter (the premium in the overstock allowance contract, and the bonus in the SLA) and shows using Maximum Likelihood Estimation techniques that this model organizes the data better than either the standard theory or a model based on risk aversion.

5.2 Channel Coordination with Deterministic Demand and Social Preferences

Several studies investigate channel coordination in a bilateral monopoly setting. This setting is simpler than the newsvendor one because it does not include a random component, so it lends itself well to clean laboratory studies. In a bilateral monopoly setting a supplier has a constant production cost c per unit and a retailer faces the linear demand function $d(p) = A - Bp$, where $d(p)$ is the amount of product sold, p is the market price, and A and B are market constants. The supplier moves first and makes a take-it-or-leave-it offer to the retailer that includes contract parameters. For example, under the wholesale price contract, the supplier offers the retailer a per-unit wholesale price w. The retailer either accepts the contract by placing an order $Q = d(p)$ or rejects it, in which case both players earn 0.

The wholesale price contract under bilateral monopoly results in the double marginalization problem, just like it does in the Newsvendor setting. Cui et al. [28] developed a behavioral model with inequality averse supplier and retailer. If the retailer is primarily concerned with avoiding disadvantageous profit distribution, inequality aversion can worsen the inefficiency, because the retailer will punish the supplier for setting the wholesale price too high, and set an excessively high market price. When the retailer is sufficiently averse to the advantageous inequality though (the retailer suffers disutility from earning more than the supplier), a supplier can coordinate the channel with a simple wholesale price contract by setting the wholesale price above the production cost, yet still sufficiently low to induce the inequality-averse retailer to set the market price below the expected profit maximizing level. In essence, the retailer rewards the supplier's generosity by foregoing some of its own profits to split the channel profit equally.

Some evidence of this positive reciprocity emerges from laboratory settings. Loch and Wu [85] show that in an experimental game in which a supplier and a retailer play repeatedly and are primed for a positive relationship, retailers tend to set market prices below the expected profit maximizing best reply; that is, they reward suppliers for low wholesale prices. But even though channel efficiency increases as a result of the positive relationship priming, it does not increase enough to coordinate the channel. Either retailers are not sufficiently averse to the advantageous inequality, or suppliers are unable to recognize that they can improve their earnings through lower wholesale prices.

This failure of the wholesale price contract to coordinate the channel, even in the exceedingly advantageous setting of Loch and Wu [85] experiment, brings us back to the question of how channels might be coordinated. Standard theory offers contracting arrangements that eliminate double marginalization by aligning the economic incentives of supply chain members. In a bilateral monopoly setting, many of these arrangements are mathematically equivalent to the two-part tariff mechanism in which the supplier sells to the retailer at cost, but charges a fixed fee. The coordination occurs because when the supplier sells at cost, the retailer's marginal cost becomes the same as the supplier's, inducing channel-optimal order quantity, and the supplier earns positive profits through a fixed transfer fee.

Ho and Zhang [55] test the two-part tariff mechanism, as well as a mathematically equivalent quantity discount mechanism, and find that neither arrangement improves channel efficiency relative to the wholesale price contract, though the quantity discount contract performs better than the two-part tariff. They also report that loss aversion, in conjunction with bounded rationality conceptualized as the quantal response equilibrium (QRE) [89], can explain the difference in performance between two mathematically equivalent mechanisms.

Lim and Ho [82] compare the wholesale price contract to two other contracts that in theory should result in the same outcome — two-block and three-block contracts — and find that more blocks increases efficiency, though not close to 100%. According to Lim and Ho [82], a model that includes "... counterfactual (or foregone) profits [the retailer] would have earned if the lower marginal prices were actually applied... to those blocks with higher marginal prices" (p. 321), in conjunction with QRE, fits the data well. They conclude that retailers experience disutility from paying radically different marginal prices in different blocks, and because contracts with more blocks include a finer gradation of marginal prices, retailers find them more palatable.

Whereas the research focus of Lim and Ho [82] and Ho and Zhang [55] primarily centers on developing and estimating behavioral models to explain why coordinating contracts perform very differently in the laboratory than in theory, they do not directly address another major cause of efficiency losses in the laboratory: retailer rejections. Coordinating contracts deliver less than 100% efficiency because many of them get rejected. For Lim and Ho [82], for example, 11% of the two-block contracts and 15% of the three-block contracts are rejected, and Ho and Zhang [55] find that 25.76% of the two-part-tariff contracts and 17.77% of quantity discount contracts are rejected. Conditional on being accepted, contracts in those studies are nearly 100% efficient.

One explanation for rejections is related to bounded rationality [119]. The bounded rationality framework has a great deal of intuitive appeal as a potential explanation for why retailers sometimes reject contracts that allocate most of the channel profits to suppliers. Recall that basic bounded rationality idea is that when people are faced with several options, they do not select the option with the highest utility with certainty; rather they select it only with some probability that depends on the relative utility of this option and the coefficient of certitude. The higher the utility and the precision parameter, the greater is the chance of choosing the option with the highest utility. Therefore, contracts that allocate less profit to the retailer are more likely to be rejected by that retailer.

But an alternative explanation for retailer rejections is fairness. The retailer may demand fairness for herself, and derive sufficient disutility from a contract that allocates most of the profits to the supplier, to make a rejection preferable, because a rejection results in a fair split of 0–0. An early model of bargaining that incorporated fairness concerns by Bolton [13], proposes a utility function with an asymmetric loss component that includes only a disutility from receiving less than an equal share. Fehr and Schmidt [39] extended the model to include disutility from the advantageous inequality, and Bolton and Ockenfels [11] to include more general utility functions and incomplete information. In a meta-analysis, DeBruyn and Bolton [32] estimate the asymmetric loss component utility function from Bolton [13] and find not only that a specification incorporating both fairness and bounded rationality fits many different data sets from bargaining experiments, but also that it has significant predictive power out-of-sample.

Katok and Pavlov [74] develop a new analytical model for coordinating a channel in which both parties are interested in fairness. Let π_R and π_S denote the retailer's and the supplier's profit, respectively, resulting from the retailer's acceptance or rejection. The retailer's utility is:

$$U(\pi_R, \pi_S|\alpha, \beta) = U_R = \pi_R - \alpha[\max(\pi_S - \pi_R, 0)] - \beta[\max(\pi_R - \pi_S, 0)] \tag{5.2}$$

where $\alpha \geq 0$ measures the retailer's disutility of earning less than the supplier (disadvantageous inequality), and $\beta \geq 0$ measures the retailer's disutility of earning more than the supplier (advantageous inequality). The supplier's utility is analogous to Equation (5.2).

If the supplier has full information about the retailer's preferences for fairness (he knows the values of α and β), and the retailer is fully rational (i.e. the retailer always chooses the option that delivers the highest utility; $\tau = 0$) the supplier can coordinate the channel by offering the retailer just enough to make accepting the contract preferable to rejecting it. However, when the retailer's preferences for fairness are his own private information, the optimal contract generally includes the possibility that even a fully rational retailer will reject it, so, a self-interested supplier will not coordinate the channel. If the retailer is boundedly rational, coordination will generally not be achieved even under full information. With incomplete information, bounded rationality generally increases rejections and further decreases efficiency.

Katok and Pavlov [74] proceed to test the model in the laboratory, in a setting with linear demand $d(p) = 100 - p$, and the production cost of $c = 20$. Most of their design centers around the minimum order quantity (MOQ) contract in which the supplier offers the wholesale price w and a minimum order quantity $q_{\min}$ to the retailer, and the retailer either orders $q \geq q_{\min}$ or rejects the contract by ordering $q = 0$.

The goal of their design is three-fold: (1) to validate the central assumption of the model that players are motivated by fairness concerns; (2) to measure the extent to which using a flexible contract improves efficiency and supplier profitability, relative to the wholesale price contract; and (3) to test the main conclusion of the Katok and Pavlov [74] analytical model, which is that rejections and the resulting inefficiency are caused by incomplete information about the fairness parameters.

The test of the fairness assumption is done by comparing behavior in two MOQ treatments. In both treatments the supplier is a human subject who proposes a contract to the retailer. In the baseline treatment, called simply MOQ, the retailer is also human, and the consequence of the rejection is that both players earn 0. The hypothesis that bounded rationality can explain retailer's rejections is tested in a treatment called MOQ-D (D stands for the Dictator game) in which the retailer is human, and can accept and reject offers, but the consequence of a rejection is that the retailer earns 0, but the supplier still earns the profit he would have earned had the contract been accepted. Bounded rationality implies the same rejections in MOQ and MOQ-D, while fairness implies lower rejections in MOQ-D. Rejections turn out to be 23.06% in the MOQ treatment, but only 0.56% in the MOQ-D treatment; clearly retailers reject unfair offers intentionally to punish suppliers; bounded rationality does not explain rejections. Figure 5.3(a) compares channel efficiency in the MOQ and MOQ-D treatments over the course of the session. We can see that while in the MOQ-D treatment efficiency jumps to 100% after the first few initial rounds, in the MOQ treatment efficiency stays well below 100%. Figure 5.3(b) compares the average $q_{\min}$ in the two treatments. To coordinate the channel the supplier has to set $q_{\min} = 40$. We can see that suppliers in the MOQ treatment start by setting $q_{\min}$ too low, but learn over time to set it close to the optimal level. However, the efficiency in the MOQ treatment does not come close to 100% even by the end of the session, so clearly, retailer rejections continue to be a major cause of channel inefficiency.

To measure the extent to which using a flexible contract improves efficiency and supplier profitability, relative to the wholesale price contract, we compare the performance of the MOQ contract to a wholesale price contract (WP), in which the supplier only sets w, but not the $q_{\min}$. Channel efficiency of the WP contract is 69%, which is not significantly

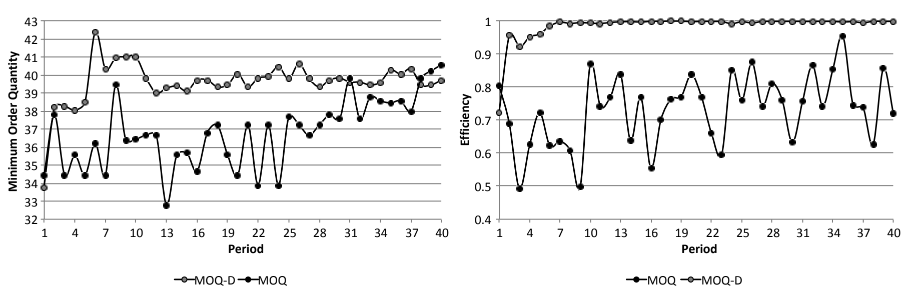


Fig. 5.3 Efficiency over time in the MOQ and MOQ-D treatments. (a) Efficiency and (b) MOQ.

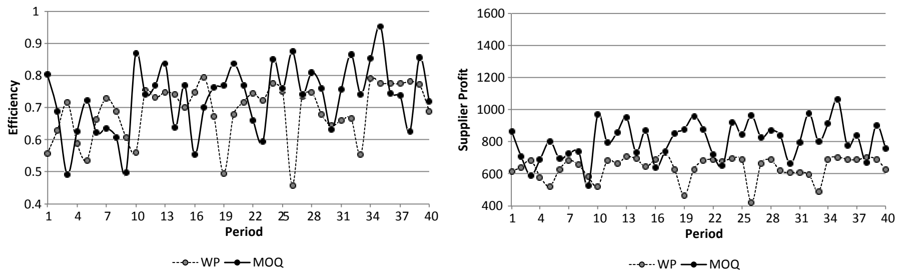


Fig. 5.4 Comparing efficiency and supplier in the MOQ and WP treatments. (a) Efficiency and (b) Supplier Profit

different from the efficiency of the MOQ contract, which is 73%. But supplier's average profit under the MOQ contract is 807.38, which is significantly higher than the supplier's profit under the WP contract, which are 640.18.

Figure 5.4 shows efficiency (a) and supplier profit (b) as they evolve over time in the MOQ and WP treatments. We can see that efficiency levels in the two treatments improve over time, but MOQ does not become more efficient than WP. In contrast, in the second half of the sessions, supplier profit becomes higher in the MOQ treatment than in the WP treatment. So flexible contracts do not improve efficiency (due to rejections) but they do improve the supplier's profit.

To test the implication of the model that rejections are caused by incomplete information about fairness, the authors conducted three additional MOQ treatments in which human suppliers were matched with computerized retailers programmed to behave the way human retailers did in the MOQ treatment. In all three treatments suppliers were told that the retailers are computerized. The three treatments differed in what supplier's knew about the retailers propensity to reject offers. In the treatment labeled MOQ-A-BR (A for automated, BR for bounded rationality) suppliers were not given any additional information relative to the MOQ treatment. They only knew that they were dealing with computerized retailers programmed to act like humans acted in an earlier session. In that treatment, rejection rates are 29.44%, the efficiency is 68%, and average supplier profits are 771.39; neither significantly different from the MOQ treatment.

In a treatment labeled MOQ-A-F-BR (F for fairness information) supplier were given the probability that the retailer they are matched with in a round would reject a particular offer. This setting corresponds to a hypothetical setting in which suppliers have full information about α, β, and τ of the individual with whom they are matched. In the MOQ-A-F-BR treatment rejections significantly dropped to 9.17%, efficiency significantly increased to 90%, and supplier average profit significantly increased to 1,108.64. Comparison between the MOQ-A-F-BR and the MOQ-A-BR treatment shows the detrimental effect of incomplete information on contract performance.

To measure the effect of bounded rationality we compare the MOQ-A-F-BR treatment to a treatment we label MOQ-A-F, in which retailers care about fairness but are programmed to be fully rational (in other words, they accept offers that yield positive utility 100% of the time). Suppliers received feedback for each offer of whether the retailer will accept it or not. Efficiency increased to 98%, rejections dropped to 0.28%, and supplier profits stayed about the same at 1,246.90. Figure 5.5 shows average efficiency over time. In all four treatments in the figure, retailers have identical fairness concerns (by design). Yet channel efficiency is radically different. These results demonstrate that it is both, the incomplete information about fairness preferences, and the bounded rationality, that contribute to the loss of efficiency. So the Katok and Pavlov [74] model captures the causes of inefficiency well.

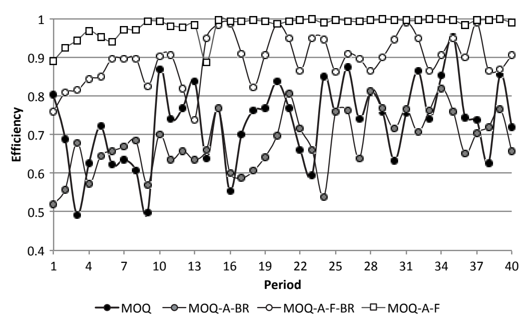


Fig. 5.5 Average channel efficiency plotted over time in four MOQ treatments. Average efficiency levels for the entire session are:
MOQ: 0.69; MOQ-A-BR: 0.68; MOQ-A-F-BR: 0.90; MOQ-A-F: 0.98.

5.3 The Bargaining Process

All experiments that deal with contracts that I discussed up to this point share one feature in common: the bargaining between the supplier and the retailer takes the form of the Ultimatum game (the supplier proposes the contract and the retailer can accept or reject it). If the inefficiency is driven primarily by incomplete information about fairness, then the ultimatum structure of the bargaining process in the experiments probably exacerbates the problem relative to the real world, because in the real world parties usually have the opportunity for a back-and-forth exchange of offers and counteroffers, which may well mitigate the incomplete information problem.

Haruvy et al. [54] tested the assertion that a bargaining process that allows for more communication may improve channel efficiency. Using the same basic underlying parameters as Katok and Pavlov [74], they compared three contracts: the minimum order quantity (MOQ), the two-part tariff (TPT), and the wholesale price (WP) using two bargaining protocols, the ultimatum bargaining (UB) protocol that has been used in pervious studies, and the structured bargaining (SB) protocol. Under the structured bargaining protocol, participants have 5 minutes in which the supplier can make proposals and the retailer can either accept them, or reject them. If the proposal is accepted, the round ends, but if it is rejected within the 5-minute time window, the supplier can make a new proposal that has to be at least as good for the retailed as the rejected proposal (it can be the same). If the 5 minutes expire without an agreement, the bargaining reverts to the ultimatum protocol, under which the supplier can make the last and final offer, and the retailer can either reject it, or declare the impasse. After the impasse both players earn zero.

The structured bargaining mechanism is strategically equivalent to the ultimatum bargaining because the supplier has the option to either wait until the 5 minutes expire to make his offer, or make his offer early and never improve it. However, if the 5 minutes can be used for the supplier to collect useful information about the retailer's fairness preferences, channel efficiency can be improved.

Table 5.1. Average efficiency in the treatments in the Haruvy and Katok [53] study.

Bargaining Protocol	Contract: Wholesale Price (WP)	Two-Part Tariff (TPT)	Minimum Order Quantity (MOQ)
Ultimatum Bargaining (UB)	0.66	0.80	0.72
Structured Bargaining (SB)	0.81	0.90	0.92

Table 5.1 reports average channel efficiency in all six treatments of the Haruvy et al. [54] study. Under UB there is no statistically significant difference between the efficiency of the WP contract and the MOQ contract ($p = 0.6068$), but the TPT contract is significantly more efficient than the WP contract ($p = 0.0191$). Under the SB protocol, both coordinating contracts are more efficient than the WP contract ($p = 0.0263$ for MOQ and $p = 0.0948$ for TPT). At the same time, the MOQ and TPT efficiency is significantly below 100%, strongly so under the UB protocol ($p = 0.0007$ for MOQ and $p = 0.0002$ for TPT), and weakly so under the SB protocol ($p = 0.0524$ for MOQ and $p = 0.0789$ for TPT).

Structured bargaining leads to higher efficiency for all three contracts ($p = 0.0106$ for MOQ, $p = 0.0775$ for TPT, and $p = 0.0065$ for WP). The efficiency of about 90% of the MOQ and TPT contracts under structured bargaining is the same as the efficiency in the MOQ-A-F-BR treatment in the Katok and Pavlov [74] study, which is a treatment with full information and bounded rationality. The underlying parameters in the two studies are the same, so this result is suggestive of the fact that structured bargaining eliminated incomplete information, and the remaining loss of efficiency is due to bounded rationality.

6

Procurement Auctions

6.1 Buyer-Determined Procurement Auctions

Procurement auctions are reverse auctions, in which suppliers bid on price. Jap [64] notes that one of the key differences between forward auctions, as modeled by auction theorists, and procurement auctions, is that usually procurement auctions do not determine winners. This key aspect of procurement auctions has become the first topic of interest in BOM research on procurement auctions. Some of the other procurement auction issues BOM research examines include testing the expected cost equivalence between various formats (this is analogous to the early stream of research on forward auctions that tested expected revenue equivalence [68]), and investigating the extent to which bidders follow equilibrium bidding strategies.

As Jap [64] pointed out, in practice, most procurement auctions are buyer-determined (BD), that is, the buyer-determines the winner based on the price bid as well as on some other, non-price attributes. Examples of non-price attributes are reputation, location, access to expertise, establish relationship, the incumbent status, etc. In practice, procurement auctions are generally not multi-attribute, meaning that bidders do not include non-price attributes as part of their bids.

There are two important features of the buyer-determined reverse auctions that make them different from standard auctions. First, the best price bid does not necessarily win, so winning may seem random, but the probability of winning is affected by the price bid. In a dynamic (open-bid) buyer-determined auction bidders do not know either their winning status, or by exactly how much they may be winning or losing. As a result, bidders in open-bid auctions do not have the dominant bidding strategy, and open-bid buyer-determined auctions begin to have some features of sealed-bid auctions, in the sense that bidders have to decide on their final bid without knowing their winning status with certainty.

Engelbrecht-Wiggans et al. [36] model the bidders in the BD auction as having bidder-specific non-price attributes that can be measured using a parameter Q_i. The Q's are independent across bidders, can be arbitrarily related to bidder i's cost C_i; the bidder i knows his own Q_i and C_i but not the cost or quality of her competitors. Engelbrecht-Wiggans et al. [36] also assume that the buyer knows the qualities of all the bidders, and awards the contract to the bidder whose bid results in the highest *buyer surplus* level $Q_i - B_i$, where B_i is bidder i's bid. The assumption that the buyer knows the Q's is reasonable because in the end the buyer has to determine the winner of the auction based in part on the qualities, so at some point in time he has to learn the Q's. Engelbrecht-Wiggans et al. [36] compare the sealed bid BD auction to a standard (price-based sealed bid) reverse sealed auction and show analytically that BD auctions result in higher buyer surplus levels as long as there are enough suppliers competing and the positive relationship between Q and C is not too strong.

In the lab they conduct the experiment to test the model. In all treatments $C_i \sim \mathrm{U}(0,100), Q_i = C_i + \gamma X_i$ where $X_i \sim \mathrm{U}(0,1)$. The model predicts that the buyer surplus level in the buyer-determined auction is below than that of the price-based auction if and only if the number of bidders $N = 2$ and $\gamma > 200$. Therefore, the treatments in the experiment are: $N = 2$ and $\gamma = 100$; $N = 2$ and $\gamma = 300$; and $N = 4$ and $\gamma = 300$.

Engelbrecht-Wiggans et al. [36] conducted these three treatments with both mechanisms (price-based and buyer-determined) and also with experienced and inexperienced bidders. Experienced bidders came to the lab after having participated in a session in which they bid against computerized rivals.

Table 6.1 summarizes actual and theoretical buyer surplus levels in the Engelbrecht-Wiggans et al. [36] study. The results are consistent with the model's qualitative predictions:

1. When $N = 2$ and $\gamma = 100$ the buyer surplus is significantly higher under BD than under PB.
2. When $N = 2$ and $\gamma = 300$ the buyer surplus is significantly lower under BD than under.
3. When $N = 4$ and $\gamma = 300$ the buyer surplus is again significantly higher under BD than under PB.

The actual buyer surplus levels are above predicted in all treatments. Since the auctions in this study are sealed-bid, these higher than predicted average buyer surplus levels are consistent with overly-aggressive bidding that has been reported in forward auction experiments (the "sealed-bid" effect).

6.2 The Effect of Feedback

Haruvy and Katok [52] consider the effect of information that bidders have in terms of price visibility during the auction, and in terms of their knowledge about the non-price attributes of the other bidders (Q). The study manipulates auction format (open-bid vs. sealed bid) and whether or not the non-price attributes of all bidders are known to all bidders, or whether they are the bidder's private information. In all treatments bidders continue to know their own Q's. In the open-bid format, bids are entered dynamically and the contract is awarded to the bidder who generates the highest buyer surplus $Q_i - B_i$ with her final bid. In the sealed-bid auction (which is also called a request for proposals (RFP) in the procurement literature) each bidder places a single bid B_i, and the contract is awarded to the bidder whose bid generates the highest buyer surplus $Q_i - B_i$. The open-bid format has full price visibility, and the sealed-bid format has no price visibility.

Table 6.1. Actual and predicted buyer surplus levels in the Engelbrecht-Wiggans et al. [36] study. Median buyer surplus levels are in square brackets, and standard deviations are in parenthesis. The unit of analysis is a session; each treatment contains two sessions.

		Experienced			Inexperienced		
Mechanism		$N = 2/\gamma = 100$	$N = 2/\gamma = 300$	$N = 4/\gamma = 300$	$N = 2/\gamma = 100$	$N = 2/\gamma = 300$	$N = 4/\gamma = 300$
Buyer-determined	Actual	[49.00] 45.93 (12.32)	[123.00] 123.31 (30.12)	[235.00] 196.36 (29.00)	[42.00] 39.74 (12.72)	[132.00] 120.41 (37.16)	[202.50] 226.12 (40.56)
	Theoretical	[36.75] 34.30 (11.22)	[110.50] 102.88 (33.71)	[195.00] 184.25 (33.53)	[36.75] 34.30 (11.22)	[110.50] 102.88 (33.71)	[195.00] 184.25 (33.53)
Price-based	Actual	[27.50] 28.29 (32.66)	[136.50] 132.37 (88.37)	[123.00] 130.59 (89.08)	[31.50] 30.79 (33.08)	[137.00] 136.71 (88.37)	[150.00] 143.72 (90.18)
	Theoretical	[20.05] 19.51 (32.06)	[123.50] 123.54 (88.29)	[134.75] 133.80 (89.09)	[20.50] 19.51 (32.06)	[123.50] 123.54 (88.29)	[134.75] 133.80 (89.09)

Quality transparency is the second factor that Haruvy and Katok [52] manipulate. Bidders always know their own Q's, and in the full information condition (F) they also know the Q's of their competitors, while in the private information condition (P) they do not.

There is an expected-buyer surplus equivalence for risk neutral bidders that holds between the sealed bid auction with private information and the open-bid auction with information. This result follows from the expected-buyer surplus equivalence between the sealed-bid first- and second-price buyer-determined auctions (see [36] for the proof), and the strategic equivalence between the sealed-bid second price buyer-determined auction and the open-bid buyer-determined auction with full information. So Haruvy and Katok [52] have analytical benchmarks for two of the treatments in their study. They also show that as long as the score $Q_i - C_i$ and the quality Q_i are not independent, bids in the sealed-bid buyer-determined auction with full information depend on the qualities of the competitors. For the open-bid buyer-determined auction with private information, Haruvy and Katok [52] show that in equilibrium bids cannot fully reveal either the quality Q_i or the score $Q_i - C_i$ of bidder i.

In their experiment, Haruvy and Katok [52] use parameters similar to Engelbrecht-Wiggans et al. [36], with $C_i \sim \mathrm{U}(0,100)$ and $Q_i \sim \mathrm{U}(C_i, C_i + \gamma)$, with $\gamma = 300$. In all of their treatments auctions had four bidders ($N = 4$), and they used a 2×2 full factorial design in which they varied the auction mechanism (sealed- or open-bid) and the quality information (full or private). Each treatment included 4 cohorts of 8 participants randomly re-matched in groups of 4 for 30 auctions.

It is worthwhile at this point to understand the risk-neutral equilibrium bidding behavior. In the open-bid auction with full information bidders know their winning status, so they have the dominant strategy to bid down in the smallest allowable bid decrements as long as they are losing, and to stop bidding when they are winning, so for losing bidders we have:

$$B_i = C_i. \tag{6.1}$$

In the sealed-bid auction with private information, Engelbrecht-Wiggans et al. [36] derived the risk-neutral Nash equilibrium to be:

$$B_i = C_i + \frac{1}{N}(Q_i - C_i). \tag{6.2}$$

In the other two treatments, however, equilibrium bidding strategies cannot be derived due to the lack of bidder symmetry.

The authors then proceed to use the equilibrium bid functions, to compute predicted average buyer surplus levels, and the predicted proportion of efficient allocations in the two treatments in which theoretical predictions are available, using the actual realizations of costs and qualities.

Table 6.2 shows that the average surplus in the open-bid auctions with full information is in line with theory, while in the sealed-bid private format, average buyer surplus levels are higher than predicted. The efficiency levels are slightly lower than predicted in both cases. A notable consequence of these deviations from equilibrium predictions is that the expected-buyer surplus equivalence between the open-bid full and the sealed-bid private conditions fails to hold.

In terms of the individual bidding behavior, Haruvy and Katok [52] find that bidding behavior in the open-bid auction with full information is mostly in line with theoretical predictions. In the sealed-bid private treatment, however, the coefficient on the score variable $Q_i - C_i$ is

Table 6.2. Average buyer surplus levels, proportion of efficient allocations, and where possible, the comparison between actual and estimated theoretical buyer surplus levels and efficiency.

	Open-Bid Full	Sealed-Bid Private	Open-Bid Private	Sealed-Bid Full
Actual Buyer surplus	186.11	224.60	211.85	205.88
(Standard error)	(4.17)	(2.60)	(6.12)	(7.90)
Actual Proportion of efficient allocations	86.88%	88.43%	85.94%	84.38%
Deviation of actual surplus from predicted (Standard error)	-2.39 (2.11)	40.35^{**} (1.04)		
Deviation of actual efficiency from predicted.	$-13.12\%^{**}$	$-11.57\%^{**}$		

$^{*}p < 0.05$; $^{**}p < 0.01$.

significantly lower than it should be, while the other coefficients are mostly in line with predictions. In the sealed-bid full and open-bid private treatments, bidders pay attention to the available information. This result suggests that in the three treatments in which bidders do not have the dominant bidding strategy, they bid too aggressively, primarily because they fail to mark up their bids sufficiently based on their high quality. The phenomenon is clearly related to the "sealed-bid effect" because it is related to the overly aggressive bidding in sealed-bid first price auctions.

Figure 6.1 compares the average number of bids bidders place in the two open-bid treatments. We can clearly see that while bidders place a large number of bids in open-bid auctions with full information, many place a very small number of bids in the open-bid auction with full information (the mode is 1). Many bidders do not use the price information from the auction because this information does not tell them their bidding status. Instead, they select what amounts to a sealed-bid, and simply place it.

The main conclusion from the Haruvy and Katok [52] study is that giving bidders less information (price information or quality information) appears to be better for the buyer. Katok and Wambach [77] stress-test this assertion by examining what happens when bidders do not know even their own quality. In procurement events, bidders often

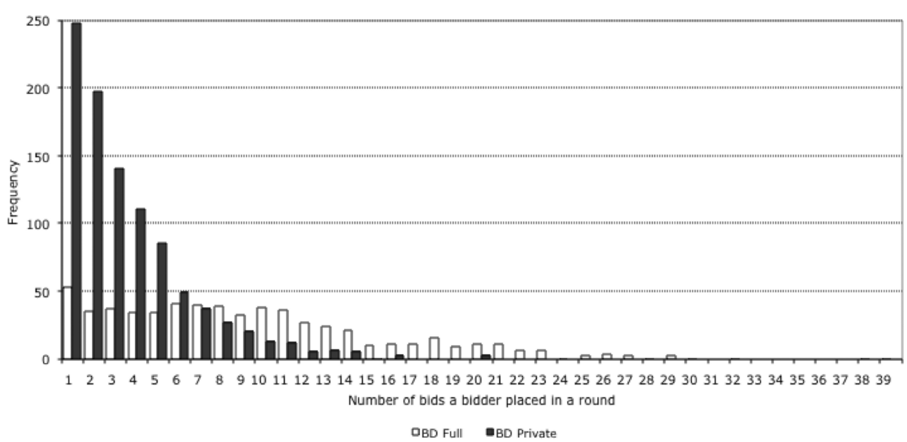


Fig. 6.1 Average number of bids bidders place in the two open-bid treatments.

know the attributes that are important to the buyer, but often do not know the exact trade-offs between those attributes. In fact, sometimes buyers do not even know their own trade-offs until they evaluate the bids after the auction ends [34]. In this setting, when bidder i does not know Q_i, winner determination begins to look random to the bidders.

Katok and Wambach [77] show that in this setting there exists an equilibrium in which all bidders stop bidding at a point at which everyone has the same ex ante probability of winning (i.e., $1/N$). The reserve price has to be high enough relatively to the differences in privately known parameters for this equilibrium to exist.

The Katok and Wambach [77] experiment included three treatments. In all treatments two bidders ($N = 2$) whose cost is known to be 50 ($C_i = 50 \forall i$) compete in auctions with the reserve price of 200. The bidders only differ in their quality ($Q_i \sim \mathrm{U}(0, 10)$), and compete in 30 open-bid descending auctions (with random re-matching) that last for 1 minute and have a 10 second soft close. The three treatments are as follows:

1. Binding: all bidders know their own Q_i and Q_j of their competitor.
2. Non-binding: bidders do not know any Q_j.
3. Non-binding (know own): bidders know their own Q_j but not their competitor's.

Figure 6.2 shows the average prices over time in the Katok and Wambach [77] study. It is clear from the figure that after the initial 13 rounds, bidders in the non-binding treatment learn to implicitly collude, driving prices essentially to the reserve level. So giving bidders less information is not always better for the buyer, but the Katok and Wambach [77] counter-example is quite extreme because auctions have only two bidders (fewer players generally make it easier to collude) and their cost is constant and known to all.

Elmaghraby et al. [35] examine the effect of price visibility, but unlike Haruvy and Katok [52] who only examine the two extreme forms of price visibility, Elmaghraby et al. [35] look at the effect of rank-based feedback in buyer-determined auctions. With rank feedback, bidders in open-bid auctions are told the rank of their price bids, but they do

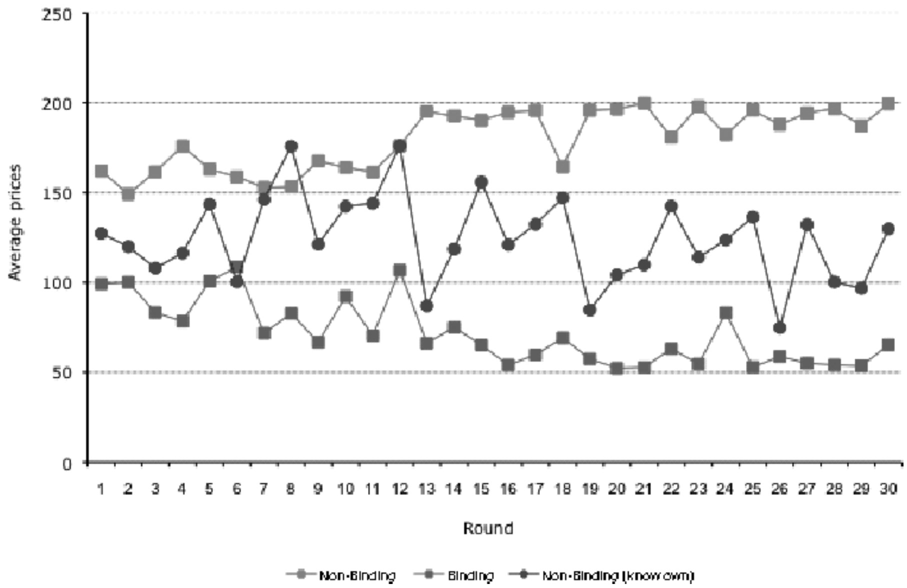


Fig. 6.2 Average prices in the Katok and Wambach [77] study.

not know the actual bids of their competitors. Elmaghraby et al. [35] examine a simple setting with two quality types (high and low) and full information about the competitor's quality type. When a bidder is bidding against competitor of the same type (called symmetric market), he has the dominant strategy of bidding down one bid decrement at a time as long as he is not in rank one. When a bidder is bidding against a competitor of the opposite type (called asymmetric market), in the simplified Elmaghraby et al. [35] setting, he should simply bid down to the level of what his sealed bid would have been. Rank-based feedback is prevalent in practice. There is a perception that it leads to less adversarial relationships between buyers and suppliers [65, 66], and suppliers prefer it because it reveals less information about their cost to their competitors. Buyers also prefer it because they believe that it leads to more competition.

All treatments include auctions with two bidders ($N = 2$). There are two types of bidders, called High and Low types. The quality for high types is 200 and the quality for low types is 100 ($Q^H = 200$, $Q^L = 100$). The costs of the two types come from different distributions,

Table 6.3. Summary of the average prices, bid decrements (standard deviations in parenthesis) and theoretical predictions [in square brackets].

	Prices			Bid Decrements	
Treatment	Overall	Symmetric	Asymmetric	Symmetric	Asymmetric
Full	67.11 (3.06) [67.60]	69.03 (3.93) [71.24]	65.65 (3.07) [65.34]	6.35** (1.26) [1]	5.82** (1.14) [1]
Rank	58.73** (3.59) [69.04]	63.68** (3.03) [71.24]	55.66** (4.41) [67.67]	12.61** (3.59) [1]	21.86 (4.65) [$\gg$1]
Sealed Bid	57.49* (2.20) [68.26]	58.21* (2.86) [69.20]	57.04* (1.93) [67.67]	N/A	N/A

Notes: H_o: Data = theoretical prediction; $^{*}p < 0.10$; $^{**}p < 0.05$.

for high types, $C_i^H \sim \mathrm{U}(100, 200)$, and for low types $C_i^L \sim \mathrm{U}(0, 100)$. The auctions are open-bid with a 1 minute duration and a 10 second soft close. The three treatments differ in their feedback: Full feedback, Rank feedback, and Sealed bid.

Table 6.3 summarizes the average prices, bid decrements, and theoretical predictions in the Elmaghraby et al. [35] study. Here we see again that sealed-bid prices are lower than open-bid prices. The consequence of this observation is that we observe the “sealed-bid effect” in asymmetric auctions with rank feedback, where average prices are very close to sealed-bid prices. Surprisingly, in symmetric auctions, average prices in auctions with rank feedback are lower than full feedback prices, which should not be the case because symmetric bidders have the same dominant bidding strategy under both formats.

The explanation Elmaghraby et al. [35] propose for overly aggressive bidding in symmetric auctions with rank feedback is bidder impatience. Even though bidders should be bidding down one bid decrement at a time, the average bid decrement in these auctions is 12.61. While significantly smaller than the average bid decrement of 21.86, in asymmetric auctions, it shows that that jump bidding due to bidder impatience in these auctions is highly prevalent.

The Elmaghraby et al. [35] results complement the findings of Isaac et al. [61, 62]. The authors study both real-world data from the FCC spectrum auctions [62] and from the lab experiments [61]; they find that auction formats that allow for jump bids can help increase the bid-taker's revenue (alternatively, decrease her procurement costs). In addition, their data suggest that jump bidding arises as a result of bidder impatience rather than effort by bidders to deter competition (signaling). Kwasnica and Katok [81] report similar results in experiments in which bidder impatience was deliberately induced.

Elmaghraby et al. [35] conducted a robustness check of their results by conducting treatments in which bidders do not know the type of their competitor, as well as treatments in which the cost support of low and high type bidders overlaps. The results continue to hold.

6.3 Qualification Screening and Incumbency

Wan et al. [125] focus on one particularly important non-price attribute in buyer-determined auctions — incumbency status. In their model an incumbent supplier competes against an entrant supplier whose probability of being able to meet the requirements to perform the contract (pass qualification screening) is $0 \leq \beta < 1$. The buyer has a choice of screening the entrant before the auction, and if he fails the qualification screening, which happens with probability $1 - \beta$, renewing the incumbent's contract at the current price of R, or waiting to screen him after the auction. In the latter case, the buyer will have to screen the entrant (which costs K to do) only in the event the entrant wins the contract. But the auction between an incumbent and an entrant who may or may not be qualified is less competitive than an auction between two qualified bidders, because the incumbent may lose the auction but win the contract (with probability $1 - \beta$).

Whether the buyer is better off to screen the entrant before or after the auction is the central question that Wan et al. [125] pose, and the answer hinges on the incumbent's bidding behavior when competing against an unscreened entrant. Wan et al. [125] derive equilibrium bidding strategy for the (risk neutral) incumbent supplier in this situation. In equilibrium, a high cost incumbent should bid the reserve, a very low cost incumbent, should bid to win, and at intermediate cost levels, incumbents should stop bidding at some threshold above their costs.

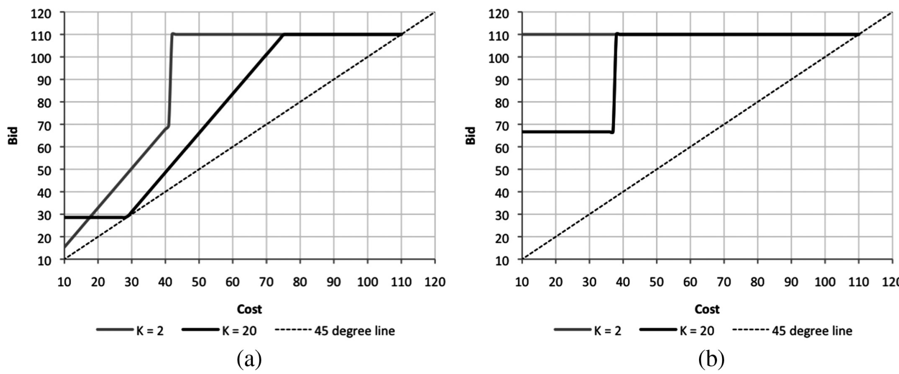


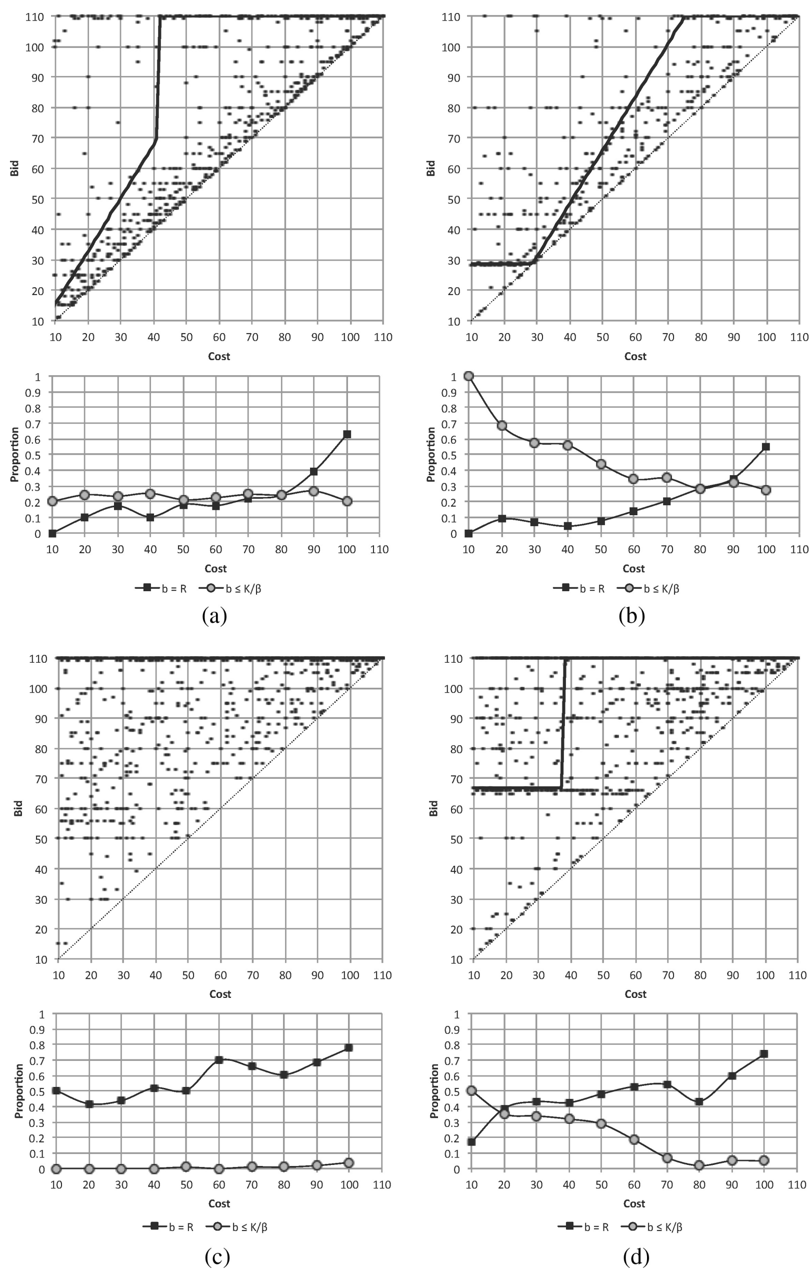
Fig. 6.3 A risk-neutral incumbent's bidding functions when $K = 2$ and $K = 20$. Plots assume $x_i \sim \mathrm{U}[10, 110], x_e \sim \mathrm{U}[0, 100]$, and $\beta = 0.7$ (left panel) or $\beta = 0.3$ (right panel).

Figure 6.3 shows the equilibrium bidding strategy for a risk-neutral incumbent whose cost $x_i \sim \mathrm{U}(10, 110)$ bidding against an entrant whose cost $e_i \sim \mathrm{U}(0, 100)$. The graphs represent the parameters in the Wan et al. [125] experiment that varies β at 0.3 and 0.7 and the cost of screening K at 2 or 20. The reserve price is set at 110 in all treatments.

When competing against an entrant who may not be qualified, the incumbent often bids less aggressively than he would against a qualified competitor. Sometime, the incumbent may boycott the auction entirely (always place the bid of R), as should happen when the qualification cost is low, and the entrant's probability of being qualified is also low ($\beta = 0.3, K = 2$ in the Wan et al. [125] experiment).

The main lab finding is that in this dynamic auction incumbents, bidding against computerized entrants programmed to follow the dominant strategy, bid with a great deal of noise, and on average bid more aggressively than they should in equilibrium. Figure 6.4 summarizes the incumbent bidding data.

Each part of the figure displays behavior in one treatment. The top panel of each part of the figure shows a scatter plot of actual bids as a function of x_i and the equilibrium bid function. The bottom part of each panel shows the proportion of bids (as a function of x_i) that are either boycotting ($Bid = R$) or bid all the way down as low as needed to win the auction outright $Bid \leq \max(x_i, K)$. It turns out that incumbents



(a) (b) (c) (d)

Fig. 6.4 Bidding behavior: Bids as a function of x_i, proportion of $Bid = R$, proportion of $Bid \leq \max(x_i, K)$. (a) $K = 2/\beta = 0.7$, (b) $K = 20/\beta = 0.7$, (c) $K = 2/\beta = 0.3$ and (d) $K = 20/\beta = 0.3$.

do not boycott enough when they should, and on average usually bid more aggressively than they should (the "sealed-bid effect").

As a result of the "sealed-bid effect," the buyer strategy of not qualifying the entrant supplier until after the auction is even more attractive than it should be in theory.

In summary, buyer-determined procurement auctions are a class of reverse auctions that are prevalent in procurement. The bidding takes place during a dynamic event, butwinners don't know if they are winning and losers don't know if they are losing or by how much. The resulting bidding behavior exhibits the "sealed bid effect" — bidding is overly aggressive. Giving bidders less information appears to result in lower prices, unless there is so little information that bidders can profitable implicitly collude. Rank feedback results in lower prices, primarily because it promotes the sealed-bid effect, as well as bidder impatience.

6.4 The Bid-Taker Behavior

In a simple (single item) auction, the bid-taker makes two main decisions: the auction format and the reservation price (R). The standard risk neutral theory states that the auction format does not matter [122], but the laboratory evidence seems to indicate quite strongly that if the auction format is up to the bid-taker, the sealed-bid auction is likely the best option. If an open-bid format is required (due to industry standard for example) then using rank feedback achieves much of the benefits associated with the sealed bid. But how should bid-takers set the reserve prices? In this subsection I will describe a simple (forward auction) model and a laboratory test of this model. Bidders are assumed to have independent privately-known values drawn from a distribution $F()$ with the density $f()$. In this setting bidders have a weakly dominant strategy to bid up to their value as long as they are not winning, so the bidders' behavior can be modeled as a sealed-bid second price auction, in which a bidder with a value v_i places a bid $b(v_i) = v_i$. The bid-taker with a value v_0 for the object has to set a reservation price R. If the best bid is below R, the object reverts to the bid-taker. If the best bid is above R and the second-best bid is below R, the bidder who submitted the best bid wins the object and pays R. If the two best bids are above R, the bidder who submitted the best bid wins the object and pays the amount of the second best bid. Myerson [91], and Riley and Samuelson [102] show that the optimal reserve price under risk neutrality is given by

$$R^* = v_0 + \frac{1 - F(R^*)}{f(R^*)}. \tag{6.3}$$

If the seller is risk averse with the Bernoulli utility function $u(x)$, then the optimal reservation price is given by

$$\frac{u(R^*)}{u'(R^*)} = \frac{1 - F(R^*)}{f(R^*)}. \tag{6.4}$$

An interesting implication of both of these expressions is that the optimal reserve price is independent of the number of bidders for either a risk-neutral or a risk averse seller.

Davis et al. [30] are the first to test this model in the lab. All of their treatments used human sellers with $v_0 = 0$ who had to set a reserve price for auctions with computerized bidders programmed to follow the weakly dominant strategy. They used two distributions of values: in the *Cuberoot* treatment

$$F(v) = \left(\frac{v}{100}\right)^{\frac{1}{3}} \tag{6.5}$$

and $R^* = 42$. In the *Cube* treatment

$$F(v) = \left(\frac{v}{100}\right)^{3} \tag{6.6}$$

and $R^* = 63$.

Davis et al. [30] conducted six experimental treatments. The treatments varied in $F()$ (*Cuberoot* or *Cube*), the number of bidders in the auction, and the amount of decision support provided to the bid takers. The number of bidders was $\{1, 2, 3, 4\}$ in two of the treatments and $\{1, 4, 7, 10\}$ in the other four treatments. In two of the treatments with $\{1, 4, 7, 10\}$ bidders, called *FullInfo*, the bid takers had a calculator built into the software that compute, for any reserve price the probability of not selling the object, the probability of selling it at R, the probability of selling it above R, and the average selling price conditional on

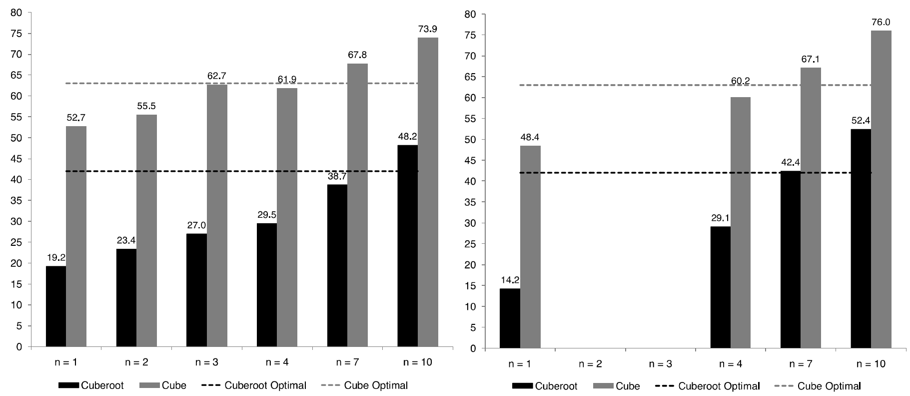


Fig. 6.5 Average reserve prices for each number of bidders in the Davis et al. [30] study.

it being above R (in other words, the bid takers were given all information they need to calculate their expected profit from each R they entered).

Figure 6.5 shows the average reserve prices for each number of bidders in the Davis et al. [30] study. The striking regularity that we can observe from the figure is that sellers set higher reserve prices for auctions with more bidders, regardless of the underlying distribution of values, and regardless of the additional decision support provided to them in the *FullInfo* condition.

A regression model confirms that there is a positive relationship between the number of bidders and the reserve prices in all treatments. This relationship is inconsistent with either the risk-neutral or a risk averse model. The authors test two behavioral models: regret and probability weighting.

The authors conclude that a simple regret model, in which the seller regrets setting R too high if he fails to sell the item, and otherwise regrets setting R too low, provides the best overall fit. Fitting models to individuals, Davis et al. [30] find that between 40% and 50% of the subjects are best described by the regret model, about 30% are best described by the probability weighting model, and the rest are best described with a risk aversion model. Only a single subject is best described by the risk-neutral model.

7

Future Trends in BOM Research

In this section I conclude the monograph with some of my personal thoughts about the future directions and trends in the BOM field. Let us start by briefly looking back to the three purposes of laboratory experiments I mentioned in the introduction, and ask how have we done so far? The three purposes are: (1) To test and refine existing theories; (2) To characterize new phenomena leading to new theory; and (3) To test new institutional designs.

Much of the effort up to this point has been devoted to (1). Many of the studies I described in Sections 4–6 test existing analytical models and often refine them, to include, for example, random errors. There has also been some effort devoted to (2), with studies that identified new phenomenon, such as loss aversion, or regret aversion. In the future I anticipate more work devoted to testing more sophisticated Operations Management models.

Trend 1: Undertake the testing of more sophisticated Operations Management models. For example, revenue management is a field ripe for laboratory investigation. I refer interested readers to Özer and Zheng [96] for a thorough discussion of behavioral issues in pricing management.

Less BOM work has so far focused on (3), testing of new institutional designs, and I expect this kind of work to be a future BOM trend. After all, Operations Management is by its nature a practical field, devoted to improving operations. The laboratory is ideal for cleanly and inexpensively testing supply chain mechanisms.

Trend 2: Explore behavioral mechanism design. The laboratory is ideal for better understanding how human decision-makers behave, and using this knowledge to design better systems that take into account how human decision-makers are likely to behave in reality, as opposed to how they should behave in theory. Mechanisms that take human behavior into account are more likely to be implemented and to work as advertised. The laboratory is also an inexpensive way to compare alternative new designs. One example of how this approach was applied in practice is the work by Bolton et al. [9].

The next trend is one that I would like to see BOM researchers to consciously pursue.

Trend 3: Become more sophisticated about proposing new explanations for observed phenomena. Behavior usually deviates from predictions of standard neo-classical models, and often it does so in systematic ways. Currently the trend is to insist on full explanations, and the expectation is that a single paper should both, identify and explain a new phenomenon. But in fact this kind of knowledge and insights should be acquired in a sequence of papers, not in a single paper. For example, consider the case of social preferences literature in economics. The first paper about the Ultimatum Game, Güth et al. [51] was published in 1982. Theory papers with the full explanation of the phenomenon did not appear until Fehr and Schmidt [39] and Bolton and Ockenfels [11]. Since 1982, hundreds, if not thousands, of related papers have been published, refining models and reporting on new experimental results (see [24] for a review).

Whether an explanation is a valid one should not be based on whether it seems plausible, and even not on whether a model fits better for one explanation than for another. Instead, experiments should be designed to directly test explanations, and when appropriate, compare them directly. This can only be done through a sequence of experiments, but such an approach requires a more nuanced understanding of what it really means to "explain" a behavioral regularity.

Trend 4: Undertake systematic investigation of the differences between students and professional. Operations Management researchers as a group seem to be more concerned with the questions regarding the effects related to the subject pool than some other social scientists are (economists, psychologists), that generally accept undergraduate students as perfectly acceptable subjects. Partly this skepticism on the part of OM researchers may have to do with the fact that business decisions are usually made by trained managers. While other social scientists are interested in more basic research questions. I anticipate seeing more systematic studies of the subject pool effect in the future.

So far most of studies that looked into this question failed to find differences. Nevertheless, many non-experimentalists have a very strong intuition that the subject pool matters a great deal (specifically, that the student subjects are less informative than managers would be). Rather than having hypothetical arguments, I suggest that the profession should undertake systematic studies to understand in which domains the subject pool makes a difference.

Trend 5: Explore cultural differences. Most supply chains are multinational, but very few laboratory experiments systematically examine cultural differences (Roth et al. [107] is an example of such a study in economics). A recent work by Özer et al. [98] investigating cultural differences between US and China, in terms of trust and trustworthiness, is a beginning of what I consider a new and useful trend in BOM. For our behavioral insights to be useful, we need to better understand which ones hold across cultures, which ones differ, and why.

Trend 6: Become more sophisticated about experimental design. The beauty and power of laboratory experiments is control. The reason the most influential economic experiments tend to be very simple (such as the Ultimatum Game) is because economists often take an analytical model as their staring point, and design the experiment to match the model. In contrast, because OM is at its core a practical discipline, BOM researchers are more interested in explaining reality (this tends to affect their analytical models as well, because OM models are expected, at least at first glance, to be grounded in practice). This concern with practice and realism sometimes leads to influential experimental paradigms that are too complicated for an analytical model (such as the Beer Game). The field of BOM will have to find its own balance between analytical tractability, which will provide lab experiments with theoretical guidance, and realism, which will enhance external validity of lab experiments.

Acknowledgments

I gratefully acknowledge the support from the National Science Foundation (Award 0849054) and thank Bernardo F. Quiroga for helping proofread this monograph. All remaining errors are my own.

References

[1] M. Allais, "Le comportement de l'homme rationnel devant le risqe: Critique des postulats et axioms de l'ecole americane," *Econometrica*, vol. 21, pp. 503–546, 1953.

[2] A. C. Atkinson and A. N. Donev, *Optimum Experimental Designs*. Oxford, England: Clarendon Press, 1992.

[3] R. Axelrod, *The Evolution of Cooperation*. Basic Books, 1984.

[4] S. B. Ball and P. Cech, "Subject pool choice and treatment effects in economic laboratory research," *Experimental Economics*, vol. 6, pp. 239–292, 1996.

[5] M. Becker-Peth, E. Katok, and U. Thonemann, Designing Contracts for Irrational but Predictable Newsvendors, URL: http://www.personal.psu.edu/exk106/Paper_090904MB_INFORMS.pdf, 2011.

[6] U. Benzion, Y. Cohen, R. Peled, and T. Shavit, "Decision-making and the newsvendor problem — An experimental study," *Journal of the Operational Research Society,* advance online publication 11 July 2007, doi: 10.1057/palgrave.jors.2602470, 2007.

[7] D. Bernoulli, "Specimen theoroae novae de mensura sortis, Comentarii Academiae Scientiarum Imperialis Petropolitanae, 5 175–92," *English Translation in Econometrica*, vol. 22, no. 1954, pp. 23–36, 1738.

[8] S. Blount, "When social outcomes aren't fair: The effect of causal attributions on preferences," *Organizational Behavior and Human Decision Processes*, vol. 63, pp. 131–144, 1995.

[9] G. Bolton, B. Greiner, and A. Ockenfels, "Engineering trust — reciprocity in the production of reputation information," Working Paper, University of Cologne, Germany, 2011.

[10] G. Bolton and E. Katok, "Learning-by-doing in the newsvendor problem: A laboratory investigation," *Manufacturing and Service Operations Management*, vol. 10, no. 3, pp. 519–538, 2008.

[11] G. Bolton and A. Ockenfels, "A theory of equity, reciprocity, and competition," *American Economics Review*, vol. 90, no. 1, pp. 166–193, 2000.

[12] G. Bolton, A. Ockenfels, and U. Thonemann, "Managers and students as newsvendors: How out-of-task experience matters," Working Paper, University of Cologne, Germany, 2008.

[13] G. E. Bolton, "A comparative model of bargaining: Theory and evidence," *The American Economic Review, American Economic Association*, vol. 81, no. 5, pp. 1096–1136, December 1991.

[14] G. E. Bolton and Z. Rami, "Anonymity versus punishment in ultimatum bargaining," *Games and Economic Behavior*, vol. 10, no. 1, pp. 95–121, 1995.

[15] A. J. A. Bostian, C. A. Holt, and A. M. Smith, "The newsvendor "pull-to-center effect": Adaptive learning in a laboratory experiment," *Manufacturing and Service Operations Management*, vol. 10, no. 4, pp. 590–608, 2007.

[16] A. O. Brown and C. S. Tang, "The impact of alternative performance measures on single-period inventory policy," *Journal of Industrial and Management Optimization*, vol. 2, no. 3, pp. 297–318, 2006.

[17] G. Cachon and M. Lariviere, "Supply chain coordination with revenue sharing contracts," *Management Science*, vol. 51, no. 1, pp. 30–44, 2005.

[18] G. P. Cachon, "Supply chain coordination with contracts," in *Handbooks in Operations Research and Management Science: Supply Chain Management*, Chapter 6, (S. Graves and T. de Kok, eds.), North-Holland, 2003.

[19] C. Camerer, "Individual decision Making," in *The Handbook of Experimental Economics*, vol. 1, (J. H. Kagel and A. E. Roth, eds.), pp. 587–704, Princeton University Press, 1995.

[20] O. Carare and M. H. Rothkopf, "Slow dutch auctions," *Management Science*, vol. 51, no. 3, pp. 365–373, 2005.

[21] T. N. Cason, "An experimental investigation of the seller incentives in the EPA's emission trading auction," *The American Economic Review*, vol. 85, no. 4, pp. 905–922, September 1995.

[22] E. H. Chamberlin, "An experimental imperfect market," *Journal of Political Economy*, vol. 56, no. 2, pp. 95–108, 1948.

[23] L. Chen, A. G. Kok, and J. Tong, "The effect of payment timing on inventory decisions in a newsvendor experiment," Working Paper, Fuqua School of Business, Duke University, 2010.

[24] D. J. Cooper and J. H. Kagel, "Other-regarding preferences: A selective survey of experimental results, 1995–2008," in *The Handbook of Experimental Economics*, vol. 2, (J. H. Kagel and A. E. Roth, eds.), Princeton University Press, 2008. in preparation URL: http://www.econ.ohio-state.edu/kagel/Other_Regarding%20Preferences_survey.pdf.

[25] J. C. Cox, B. Roberson, and V. L. Smith, "Theory and behavior of single object auctions," in *Research in Experimental Economics*, (V. L. Smith, ed.), pp. 1–43, Greenwich, CT: JAI Press, 1982.

[26] J. C. Cox, V. L. Smith, and J. M. Walker, "Theory and individual behavior of first-price auctions," *Journal of Risk and Uncertainty*, vol. 1, no. 1, pp. 61–99, 1988.

[27] D. Croson, R. Croson, and Y. Ren, "How to manage an over confident newsvendor," Working Paper, Cox School of Business, Southern Methodist University, USA, 2009.

[28] T. H. Cui, J. S. Raju, and Z. J. Zhang, "Fairness and channel coordination," *Management Science*, vol. 53, no. 8, pp. 1303–1314, 2007.

[29] A. M. Davis, "An experimental investigation of pull contracts," Penn State Working Paper, URL: http://www.personal.psu.edu/amd361/Paper_PullExperiment_100410.pdf, 2010.

[30] A. M. Davis, E. Katok, and A. M. Kwasnica, "Do auctioneers pick optimal reserve prices?," *Management Science*, vol. 57, no. 1, pp. 177–192, 2011.

[31] D. D. Davis and C. A. Holt, *Experimental Economics*. Princeton: Princeton University Press, 1993.

[32] A. DeBruyn and G. E. Bolton, "Estimating the influence of fairness on bargaining behavior," *Management Science*, vol. 54, no. 10, pp. 1774–1791, 2008.

[33] T. Deng and Z. J. M. Shen, "Asymmetries in the pull-to-center effect of the newsvendor experiment," U.C. Berkeley Working Paper, 2010.

[34] W. Elmaghraby, "Auctions within e-Sourcing Events," *Production and Operations Management*, vol. 15, no. 4, pp. 409–422, 2007.

[35] W. Elmaghraby, E. Katok, and N. Santamaria, "A laboratory investigation of rank feedback in procurement auctions," *Manufacturing & Services Operations Management*, in press, 2011.

[36] R. Engelbrecht-Wiggans, E. Haruvy, and E. Katok, "A comparison of buyer-determined and price-based multi-attribute mechanisms," *Marketing Science*, vol. 26, no. 5, pp. 629–641, 2007.

[37] R. Engelbrecht-Wiggans and E. Katok, "Regret in auctions: Theory and evidence," *Economic Theory*, vol. 33, pp. 81–101, 2007.

[38] R. Engelbrecht-Wiggans and E. Katok, "Regret and feedback information in first-price sealed-bid auctions," *Management Science*, vol. 54, no. 3, 2008.

[39] E. Fehr and K. M. Schmidt, "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics*, vol. 114, no. 3, pp. 817–868, 1999.

[40] T. Feng, L. R. Keller, and X. Zheng, "Decision making in the newsvendor problem: A cross-national laboratory study," Omega, in press, 2010.

[41] E. Filiz-Ozbay and E. Y. Ozbay, "Auctions with anticipated regret: Theory and experiment," *The American Economic Review*, vol. 97, no. 4, pp. 1407–1418, 2007.

[42] U. Fischbacher, "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, vol. 10, no. 2, pp. 171–178, 2007.

[43] R. A. Fisher, *The Design of Experiments Games*. Edinburgh, Scotland: Oliver and Boyd, 1935.

[44] M. M. Flood, "Some experimental games," *Management Science*, vol. 5, pp. 5–26, 1958.

[45] R. Forsythe, S. Horowitz, and M. Sefton, "Fairness in simple bargaining experiments," *Games and Economic Behavior*, vol. 6, no. 3, pp. 347–369, 1994.

[46] D. Friedman and S. Sunder, *Experimental Methods: A Primer for Economists*. Cambridge University Press, 1994.

[47] S. Gavirneni and A. M. Isen, "Anatomy of a newsvendor decision: Observations from a verbal protocol analysis," Working Paper, Cornell University, 2008.

[48] S. Gavirneni and Y. Xia, "Anchor selection and group dynamics in newsvendor decisions," *Decision Analysis*, vol. 6, no. 2, pp. 97–97, 2009.

[49] J. K. Goeree and C. A. Halt, "Hierarchical package bidding: A paper & pencil combinatorial auction," *Games and Economic Behavior*, forthcoming, 2009.

[50] J. K. Goeree, C. A. Halt, and T. R. Palfrey, "Quantal response equilibrium and overbidding in private-value auctions," *Journal of Economic Theory*, vol. 104, pp. 247–272, 2002.

[51] W. Güth, R. Schmittberger, and B. Schwarze, "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, vol. 3, no. 4, pp. 367–388, 1982.

[52] E. Haruvy and E. Katok, "Increasing revenue by decreasing information in procurement auctions," Penn State Working Paper, 2010.

[53] E. Haruvy and E. Katok, "Increasing revenue by decreasing information in procurement auctions," Penn State Working Paper, 2011.

[54] E. Haruvy, E. Katok, and V. Pavlov, "Can coordinating contracts improve channel efficiency?," Penn State Working Paper, 2011.

[55] T. Ho and J. Zhang, "Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter?," *Management Science*, vol. 54, no. 4, pp. 686–700, 2008.

[56] R. Hogarth, *Judgement and Choice*. New York: John Wiley and Sons, 2nd ed., 1987.

[57] C. A. Holt, "Competitive bidding for contracts under alternative auction procedures," *Journal of Political Economy*, vol. 88, no. 3, pp. 435–445, 1980.

[58] C. A. Holt, "Industrial organization: A survey of laboratory results," in *Handbook of Experimental Economics*, (J. Kagel and A. Roth, eds.), pp. 349–443, Princeton, NJ: Princeton University Press, 1995.

[59] C. A. Holt and S. K. Laury, "Risk aversion and incentive effects," *The American Economic Review*, vol. 92, no. 5, pp. 1644–1655, 2002.

[60] R. M. Isaac and D. James, "Just who are you calling risk averse?," *Journal of Risk and Uncertainty*, vol. 20, pp. 177–187, 2000.

[61] R. M. Isaac, T. C. Salmon, and A. Zillante, "An experimental test of alternative models of bidding in ascending auctions," *International Journal of Game Theory*, vol. 33, no. 2, pp. 287–313, 2005.

[62] R. M. Isaac, T. C. Salmon, and A. Zillante, "A theory of jump bidding in ascending auctions," *Journal of Economic Behavior & Organization*, vol. 62, no. 1, pp. 144–164, 2007.

[63] J. Jamison, D. Karlan, and L. Schechter, "To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments," *Journal of Economic Behavior & Organization*, vol. 68, pp. 477–488, 2008.

[64] S. D. Jap, "Online reverse auctions: Issues, themes and prospects for the future," *Journal of the Academy of Marketing Science*, vol. 30, no. 4, pp. 506–525, 2002.

[65] S. D. Jap, "An exploratory study of the introduction of online reverse auctions," *Journal of Marketing*, vol. 67, pp. 96–107, 2003.

[66] S. D. Jap, "The impact of online reverse auction design on buyer-supplier relationships," *Journal of Marketing*, vol. 71, no. 1, pp. 146–159, 2007.

[67] J. Kagel and D. Levin, "The winners curse and public information in common value auctions," *The American Economic Review*, vol. 76, no. 5, pp. 894–920, 1986.

[68] J. H. Kagel, "Auctions: A survey of experimental research," in *The Handbook of Experimental Economics*, (J. H. Kagel and A. E. Roth, eds.), pp. 501–585, Princeton, NJ: Princeton University Press, 1995.

[69] J. H. Kagel, R. M. Harstad, and D. Levin, "Information impact and allocation rules in auctions with affiliated private values: A laboratory study," *Econometrica, Econometric Society*, vol. 55, no. 6, pp. 1275–1304, 1987.

[70] J. H. Kagel and D. Levin, "Independent private value auctions: Bidder behaviour in first-, second- and third-price auctions with varying numbers of bidders," *Economic Journal, Royal Economic Society*, vol. 103, no. 419, pp. 868–879, 1993.

[71] J. H. Kagel and D. Levin, "Auctions: A survey of experimental research, 1995–2008," in *The Handbook of Experimental Economics*, vol. 2, (J. H. Kagel and A. E. Roth, eds.), Princeton University Press, 2008. in preparation URL: http://www.econ.ohio-state.edu/kagel/Auctions_Handbook_vol2.pdf.

[72] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, pp. 263–291, 1979.

[73] E. Katok and A. M. Kwasnica, "Time is money: The effect of clock speed on sellers revenue in dutch auctions," *Experimental Economics*, vol. 11, no. 4, pp. 344–357, 2008.

[74] E. Katok and V. Pavlov, "Fairness and coordination failures in supply chain contracts," Working Paper, Penn State University, 2009.

[75] E. Katok and E. Siemsen, "The influence of career concerns on task choice: Experimental evidence," *Management Science*, accepted, 2011.

[76] E. Katok, D. Thomas, and A. Davis, "Inventory service level agreements as coordination mechanisms: The effect of review periods," *Manufacturing & Service Operations Management*, vol. 10, no. 4, pp. 609–624, 2008.

[77] E. Katok and A. Wambach, "Collusion in dynamic buyer-determined reverse auctions," Penn State Working Paper, 2008.

[78] E. Katok and D. Y. Wu, "Contracting in supply chains: A laboratory investigation," *Management Science*, vol. 55, no. 12, pp. 1953–1968, 2009.

[79] M. Kremer, S. Minner, and L. Van Wassenhove, "Do random errors explain newsvendor behavior," *Manufacturing and Services Operations Management*, vol. 12, no. 4, pp. 673–681, Fall 2010.

[80] V. Krishna, *Auction Theory*. San Diego, CA: Academic Press, First ed., 2002.

[81] A. M. Kwasnica and E. Katok, "The effect of timing on jump bidding in ascending auctions," *Production and Operations Management*, vol. 16, no. 4, pp. 483–494, 2007.

[82] N. Lim and T.-H. Ho, "Designing price contracts for boundedly-rational customers: Does the number of blocks matter?," *Marketing Science*, vol. 26, no. 3, pp. 312–326, 2007.

[83] S. A. Lippman and K. F. McCardle, "The competitive newsboy," *Operations Research*, vol. 45, no. 1, pp. 54–65, 1997.

[84] J. A. List, S. Sadoff, and M. Wagner, "So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design," *Experimental Economics*, vol. 14, no. 4, pp. 439–457, 2010.

[85] C. H. Loch and Y. Wu, "Social preferences and supply chain performance: An experimental study," *Management Science*, vol. 54, no. 11, pp. 1835–1849, 2008.

[86] D. Lucking-Reiley, "Using field experiments to test equivalence between auction formats: Magic on the internet," *The American Economic Review*, vol. 89, no. 5, pp. 1063–1079, 1999.

[87] N. H. Lurie and J. M. Swaminathan, "Is timely information always better? The effect of feedback frequency on decision making," *Organizational Behavior and Human Decision Processes*, vol. 108, no. 2, pp. 315–329, 2009.

[88] M. Machina, "Choice under uncertainty: Problems solved and unsolved," *Journal of Economic Perspectives*, vol. 1, no. 1, pp. 121–154, 1997.

[89] R. D. McKelvey and T. R. Palfrey, "Quantal response equilibria for normal form games," *Games and Economic Behavior*, vol. 10, pp. 6–38, 1995.

[90] B. B. Moritz, A. V. Hill, and K. Donohue, "Cognition and individual difference in the newsvendor problem: Behavior under dual process theory," Working Paper, University of Minnesota, 2008.

[91] R. Myerson, "Optimal auction design," *Mathematics of Operations Research*, vol. 6, no. 1, pp. 58–73, 1981.

[92] T. Neugebauer and R. Selten, "Individual behavior of first-price auctions: The importance of information feedback in computerized experimental markets," *Games and Economic Behavior*, vol. 54, pp. 183–204, 2006.

[93] J. Ochs and A. E. Roth, "An experimental study of sequential bargaining," *The American Economic Review*, vol. 79, pp. 355–384, 1989.

[94] A. Ockenfels and R. Selten, "Impulse balance equilibrium and feedback in first price auctions," *Games and Economic Behavior*, vol. 51, pp. 155–179, 2005.

[95] A. Ortmann and R. Hertwig, "The costs of deception: Evidence from psychology," *Experimental Economics*, vol. 5, pp. 111–131, 2002.

[96] Ö. Özer and Y. Zheng, "Behavioral issues in pricing management," in *The Oxford Handbook of Pricing Management*, (Ö. Özer and R. Phillips, eds.), Oxford University Press, 2011.

[97] Ö. Özer, Y. Zheng, and K. Chen, "Trust in forecast information sharing," *Management Science*, vol. 57, no. 6, pp. 1111–1137, 2011.

[98] Ö. Özer, Y. Zheng, and Y. Ren, "Forecast information sharing in China and the U.S.: Country effects in trust and trustworthiness," UT Dallas Working Paper, 2011.

[99] N. C. Petruzzi and M. Dada, "Pricing and the newsvendor problem: A review with extensions," *Operations Research*, vol. 47, pp. 183–194, 1999.

[100] C. Plott, "Dimensions of parallelism: Some policy applications of experimental methods," in *Experimental Economics: Six Points of View*, (A. Roth, ed.), New York, NY: Cambridge University Press, 1987.

[101] E. L. Porteus, "Stochastic inventory theory," in *Handbook in OR & MS*, vol. 2, (D. P. Heyman and M. J. Sobel, eds.), pp. 605–652, Elsevier, North–Holland: The Netherlands, 1990.

[102] J. G. Riley and W. F. Samuelson, "Optimal auctions," *The American Economic Review*, vol. 71, no. 3, pp. 381–392, 1981.

[103] A. E. Roth, "The evolution of the labor market for medical interns and residents: A case study in game theory," *Journal of Political Economy*, vol. 92, pp. 991–1016, 1984.

[104] A. E. Roth, "Bargaining experiments," in *The Handbook of Experimental Economics*, vol. 1, (J. H. Kagel and A. E. Roth, eds.), pp. 253–248, Princeton University Press, 1995.

[105] A. E. Roth, "Introduction to experimental economics," in *The Handbook of Experimental Economics*, vol. 1, (J. H. Kagel and A. E. Roth, eds.), pp. 3–109, Princeton University Press, 1995.

[106] A. E. Roth, "Form and function in experimental design," *Behavioral and Brain Sciences*, vol. 24, pp. 427–428, 2001.

[107] A. E. Roth, V. Prasnikar, M. Okuno-Fujiwara, and S. Zamir, "Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: An experimental study," *The American Economic Review*, vol. 81, no. 5, pp. 1068–1095, 1991.

[108] A. Rubinstein, "Perfect equilibrium in a bargaining model," *Econometrica*, vol. 50, no. 1, pp. 97–109, January 1982.

[109] N. Rudi and D. Drake, "Level, adjustment and observation biases in the newsvendor model," Working Paper, INSEAD, France, 2008.

[110] A. G. Sanfey, J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen, "The neural basis of economic decision-making in the ultimatum game," *Science*, vol. 300, pp. 1755–1758, 2003.

[111] J. P. W. Scharlemann, C. C. Eckel, A. Kacelnik, and R. K. Wilson, "The value of a smile: Game theory with a human face," *Journal of Economic Psychology*, vol. 22, pp. 617–640, 2001.

[112] S. Schiffels, J. Brunner, A. Fuegener, and R. Kolisch, "Profit- vs. cost-orientation in the newsvendor problem: Insights from a behavioral study," Technische Universitaet Muenchen, Germany Working Paper, 2011.

[113] K. L. Schultz, J. O. McClain, L. W. Robinson, and L. J. Thomas, "The use of framing in inventory decisions," Working Paper, Cornell University, 2007.

[114] M. Schweitzer and G. Cachon, "Decision bias in the newsvendor problem: Experimental evidence," *Management Science*, vol. 46, no. 3, pp. 404–420, 2000.

[115] V. L. Smith, "An experimental study of competitive market behavior," *The Journal of Political Economy*, vol. 70, no. 2, pp. 111–137, April 1962.

[116] V. L. Smith, "Experimental economics: Induced value theory," *The American Economic Review*, vol. 66, no. 2, pp. 274–279, 1976.

[117] V. L. Smith, "Microeconomic systems as an experimental science," *The American Economic Review*, vol. 72, pp. 923–955, 1982.

[118] J. Sterman, "Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment," *Management Science*, vol. 35, no. 3, pp. 321–339, 1989.

[119] X. Su, "Bounded rationality in newsvendor models," *Manufacturing & Service Operations Management*, vol. 10, no. 4, pp. 566–589, 2008.

[120] L. L. Thurstone, "The indifference function," *Journal of Social Psychology*, vol. 2, pp. 139–167, 1931.

[121] A. Tversky and D. Kahneman, "The framing of decisions in the psychology of choice," *Science*, vol. 211, no. 4481, pp. 453–458, 1981.

[122] W. Vickrey, "Counterspeculation, auctions and competitive sealed tenders," *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961.

[123] L. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press, 1944.

[124] W. A. Wallis and M. Friedman, "The empirical derivation of indifference functions," in *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz*, (O. Lange, F. McIntyre, and T. O. Yntema, eds.), pp. 175–189, Chicago, IL: Chicago University Press, 1942.

[125] Z. Wan, D. Beil, and E. Katok, "When does it pay to delay supplier qualification? Theory and experiments," Penn State Working Paper, 2010.

[126] J. Weimann, "Individual behavior in a free riding experiment," *Journal of Public Economics*, vol. 54, pp. 185–200, 1994.

[127] E. Winter and S. Zamir, "An experiment on the ultimatum bargaining in a changing environment," *Japanese Economic Review*, vol. 56, pp. 363–385, 2005.