JOURNAL OF BUSINESS LOGISTICS
STRATEGIC SUPPLY CHAIN RESEARCH

Journal of Business Logistics, 2013, 34(1): 6–14

Using Laboratory Experiments in Logistics and Supply Chain Research

Cary Deck[1] and Vernon Smith[2]

[1]*University of Arkansas*
[2]*Chapman University*

Laboratory experiments provide an underutilized methodology for subjecting research in logistics and supply chain management to rigorous scientific scrutiny. As discussed in this paper, by following established procedures researchers can create an economic system in which behavior can be observed and replicated. With the ability to control the institution and the environment, researchers have complete information and can exogenously manipulate treatment variables, neither of which may be feasible in field work. We also address many of the reservations that skeptics of laboratory experiments often express including issues of realism, participant sophistication, and payoff stakes. We then provide several examples where experiments have been used to study issues relevant to logistics and supply chain management including auctions, wholesale practices in gasoline markets, inventory replenishment, liberalization of the electric power industry, and deregulation of the natural gas markets. Finally, we identify several additional areas where laboratory experiments could be informative.

Keywords: laboratory experiments; procurement and replenishment; energy markets

INTRODUCTION

It is not surprising when physicists conduct experiments to test the predictions of a particular model or when engineers construct prototypes to test performance. Science is a continuing process of formulating hypotheses, designing experimental tests of those hypotheses, and then updating the hypotheses before scrutinizing them with ever more stringent tests. However, learning is at its greatest when the unexpected happens. In economics, laboratory experiments only became a widely recognized practice after about 1985 and are now viewed as an accepted methodology.[1] Controlled experiments afford the same opportunity to rigorously test theories and develop institutional prototypes in logistics and supply chain research as in economics. Although there are some specific logistics and supply chain topics for which experiments have been used extensively, such as procurement auctions, their use on other problems is not so well known, partly because published articles have appeared in journals specialized in the areas of application and partly because researchers have nagging questions about the ability to extrapolate from the lab to the natural economy.

We briefly summarize some common academic arguments expressing skepticism about the relevance of lab experiments to the "real world" of logistics and supply chain research. We argue that some of the common academic arguments are much too conservative and naïve about what experimental methods can bring to the table. We then take up some examples illustrating how these issues were confronted by teams of experimentalists and practitioners in industry and government, and the degree of progress that resulted.

Corresponding author:
Cary Deck, Department of Economics, University of Arkansas, Walton College of Business, Fayetteville, AR 72701, USA; E-mail: cdeck@walton.uark.edu

[1]Samuelson wrote the most popular of all introductory economics texts, now (with Nordhaus) in its 17th edition. We have chosen the date, 1985, as a key, but conservative, marker in that it is the last (12th) edition containing the statement: "Economists … cannot perform the controlled experiments of chemists and biologists because they cannot easily control other important factors" (Samuelson and Nordhaus 1985, 8).

COMMON CRITICISMS OF LABORATORY EXPERIMENTS

Skepticism about what is often referred to as the "validity" of lab experiments is perhaps part of the reason for their lack of prominence in the fields of supply chain management and logistics. This skepticism often takes the form of one of the following arguments, which will be familiar to anyone who has conducted lab experiments in a social science.[2]

Lab experiments lack realism or complexity

Far from being a substantive criticism, this is actually a major virtue of lab experiments. When scientists build a model or theory to explain a phenomenon, they abstract away what are believed to be the superfluous details to focus on the critical elements. Thus, theoretical models necessarily lack realism. This is what allows one to make predictions without needing knowledge of unobservable inputs. A model for predicting inventory replenishment does not incorporate a store manager's mood or the level of training of the head of IT support, although these factors may in fact influence choices. These details are assumed to be irrelevant and are thus excluded in providing a useful means of conceptualizing the problem, framing questions, and seeking to answer them.

Related to the lack of realism, criticism of experiments is the question of how results apply to the outside world. Ultimately, this is a question about the comparability of data from two dif-

[2]Falk and Heckman (2009) make similar arguments defending laboratory experiments in the social sciences.

ferent economic environments. Comparability issues also arise for experiments performed in different labs with different subjects, and when we use field experiments or natural observations on say, the electric power industry in one country, and inquire as to their relevance in understanding the same industry in different countries or regions. These are empirical questions that are answered subject to sampling variation by comparing experience and observations from the different sources. No amount of conjectural argument can substitute for comparative experience.

As anyone who has tried to set up an experiment knows, experimental environments are commonly richer than the models that are being tested. For example, many logistics models and economics models simply assert that when buyers and sellers interact in the market, there is a competitive price. In an actual market in or out of the lab, there is an institution that governs how the parties interact and what outcomes will occur as a result of those interactions. The institution may be a convention that buyers and sellers negotiate private deals or there may be an explicit auction format that is used. What we know from years of laboratory experiments in economics is that institutions matter, in the sense of the rules that govern how actions, such as bids, translate into outcomes (see Smith 1994, 116–118).

Those aspects of the natural world judged to be critical must be incorporated into a model. For example, a hydroelectric company needs to account for line loss and other constraints when determining how much power to generate. By the same token, if a feature of the real world is important, it can be incorporated into a lab experiment model and its effect can be studied directly. Backerman et al. (2000) do just that when they explicitly consider how transmission constraints impact market performance in the generation of electric power. This early demonstration experiment translated into much more complex electricity experiments later (Denton et al. 2001; Olson et al. 2003).

Laboratory experiments are not recreations of the naturally occurring world any more than a NASA training capsule does not reproduce a moon landing, nor should they be. Experiments are operating representations of how managers, engineers. or technicians in the world of application think about their decision environment. Not only are experiments often richer than standard theories, the lab allows one to highlight the shortcomings of a current theory or how its representation has failed to identify important parts of the problem. Ultimately, the lack of realism critique boils down to a call for modification in the experiment to explore particular components. In any meaningful application, this can only be accomplished with research teams made up the knowledgeable operating stakeholders who are accountable for the validity and relevance of the results.

The subjects lack sophistication

Many experiments use undergraduates out of *convenience*, and this is generally not a problem. Normally, in an experiment, the researcher is concerned with the direction of a treatment effect. That is, experiments are useful for evaluating comparative static effects, not levels. Just as, in a field study, knowing how much profit Company X made from switching from supplier A to supplier B does not tell you how much profit Company Y will make from a similar switch, the lab cannot be used to identify how much profit Company Y will make. However, both the lab and the experience of Company X can be used to make an educated inference about Company Y's likely outcomes. Unless there is reason to believe that sophistication will have a differential effect across treatments, then this criticism is of little significance.

It is worth stating explicitly that undergraduate students are *not required* for an experiment, and it is common, once replicable results have been established, for experimentalists to examine its robustness with respect to different subject pools. If a particular experience is believed necessary in a particular setting and is expected to interact with the treatment effect, then a researcher can conduct experiments with people who have the requisite experience. If such experience is empirically important then this needs to be captured in the theory. However, theory is commonly silent on agent identity, but such theoretical deficiencies should not prevent explorations through further experiments.

When researchers have brought more sophisticated subjects into the lab, the results have generally been similar to the results found with undergraduates.[3] For example, King et al. (1993) report the results of an asset market experiment in which the subjects were either professional over-the-counter traders or corporate personnel. Asset markets experiments are meant to capture the key features of securities that pay dividends over several periods. Undergraduate behavior in these experiments is very regular—novice sellers create a pricing bubble (see Smith et al. 1988). King et al. (1993) report that "the general pattern of trading … is not altered when the subjects are corporate personnel or when they are stock market dealers" (p. 197).

The use of undergraduates offers some advantages, and it is convenient to begin with them to create a baseline before extending the work to other subject pools. For one, their opportunity cost is relatively low, at least as compared to professionals. This is important both for encouraging someone to become a participant and for maintaining control of the experiment as discussed in the next section. Also, undergraduates do not typically come with the baggage associated with having been in an industry. In a set of experiments on common value auctions, Dyer et al. (1989) find that professionals had become proficient in avoiding the winner's curse not because they understood the problem well, but because they had developed rules of thumb. Once these professionals were taken out of a familiar context, their performance was no better than that of undergraduates. Hence, the finding that professionals did not generalize their adaptation to other contexts, a condition not commonly part of theoretical modeling, nor even discoverable without controlled comparisons.

The stakes are too low

By this point, the response to this argument should be clear—if the level of the stakes is thought to be important, then one can run experiments with high stakes. The effects of changing stakes are generally minimal (for a review, see Camerer and Hogarth 1999). Carpenter et al. (2005) (also see Hoffman et al. 1996) find that dictator and ultimatum game behavior remain essentially unchanged if one was allocating \$10 or \$100. Besedes et al. (2012) study age

[3] The professional traders in King et al. (1993) actually created some of the largest bubbles that were observed. More recently, Haigh and List (2005) also find that professionals exhibited behavior at least as extreme as undergraduates.

effects in a choice overload problem. Their main conclusion is that senior citizens perform worse than younger adults and this remains unchanged when the stakes are increased. While both of these studies find no changes in the level or the treatment effect from increasing stakes, in some instances, the level of behavior does seem to vary with the size of the stakes. For example, Holt and Laury (2002) find that people became more risk averse when several hundred dollars was on the table as compared to only a few dollars.

While the level of stakes may be important in some situations, most often this criticism is leveled after the experimental results have been viewed and found to be contradictory to a particular model or one's own belief or expectations. Unfortunately, the criticism can always be made.[4] That is, one can argue that if \$M stakes does not generate the desired outcome maybe $\$M \times K$ would, but if the results do not change if $K = 10$ one can still argue that K was not large enough. As with subject sophistication, the stakes criticism identifies a shortcoming in any theory where an entity is maximizing an objective function.[5] Absent a clear theoretical or behavioral justification for claiming that the stakes in an experiment are too low, the key issue comes down to whether or not the stakes outweigh the subjective costs faced by the participants, an issue addressed in the next section.

Field experiments are more informative

This criticism implicitly assumes that lab and field experiments are substitutes, as although the two tools provide the same information. In actuality, field experiments provide invaluable information that is *complementary* to the information that lab experiments can provide. Field experiments offer greater natural contextual richness, participants who have relevant experience, and, by definition, stakes that are comparable to those in the field but do so at the expense of control and replicability. The initial conditions, including endowments, abilities, opportunities, and feasibility constraints, are not observable to the researcher in the field much less controllable to the point where a researcher can compare two situations that differ in only one of these dimensions. In contrast, lab experiments do offer control over relevant factors and enable researchers to replicate conditions. Replicability allows a researcher to evaluate stability and regularity of outcomes without having to hope that a ceteris paribus assumption is accurate, as is necessarily the case in the field. This means that lab experiments are better able to identify the particular cause for a theory's success or failure. Lab experiments also offer the advantage of being relatively inexpensive, due to the size of stakes, the costs of implementation, and the potential for negative unintended consequences. These considerations call for research programs that use field experiment results to inform the design of lab experiments that probe more deeply into the underlying behavior, which in turn can help inform the design of better field experiments.

CONDITIONS FOR CREATING A LABORATORY ECONOMY

Economic models postulate agents who have preferences for certain outcomes and are attempting to maximize their own situation through their chosen actions. In naturally occurring settings, preferences are endogenous. A company naturally seeks to maximize its profits, but preferences must be induced on subjects in a lab using a reward medium in order for the researcher to form hypotheses about how agents are expected to behave. Typically, the medium is money for people.[6] In part, the need to induce preferences in the lab is because the people who are subjects in the lab arrive with their own endogenous preferences, but simultaneously, it is these endogenous preferences that enable the researcher to induce other preferences. Using a reward medium for which subjects have strictly increasing personal preferences, a researcher can align the subject's personal interests with that required for control of the experiment.

Smith (1982) lays out a framework for creating an economic system in the laboratory by identifying five precepts: nonsatiation, saliency, dominance, privacy, and parallelism.

The first two precepts pertain to the ability of the researcher to influence the preferences of the subject. Nonsatiation means the subject prefers to receive more of the reward medium to less. Saliency means that the level of the reward medium a subject receives is tied to the actions the subject takes within the artificial economy. Hence, paying people to complete a survey does not constitute an economic system in that the compensation is not dependent upon the responses given thus violating the precept of saliency.

The third and fourth precepts help distinguish between instances where the researcher has control over a subject's preferences as opposed to simply influencing them. Dominance means that the reward structure within the experiment overwhelms any other considerations faced by a subject. For example, if the subject is worried about impressing the experimenter or finds the task too cognitively complex, the researcher does not have control over the subject's preferences. For this reason, experiments conducted with the researcher's students could be problematic if the students believe they can convey information to the instructor that might influence a course grade or a letter of recommendation. If the reward for intense concentration during a 2-hr experiment is an increase in payoff of only a few cents, then subjects may not find it worthwhile to pay attention. It is in this sense that an experiment's stakes may be too low. However, it is important to distinguish between the level of the reward (average payment) and the gradient in rewards based on subject choices. It is the latter that is critical for controlling preferences. It should be clear that criticisms regarding the size of the stakes involved in an experiment are really a concern about dominance. As cognitive costs and reputation concerns are subjective and thus difficult to value in dollar terms, one cannot use a hard and fast rule in determining if dominance is met and instead

[4] Ariely et al. (2009) find that excessive large stakes can actually have a detrimental effect of performance.

[5] While \$10 might seem trivial to many researchers in supply chain and logistics, it likely represents a greater share of an undergraduate's wealth than \$100,000 does for many companies or governments all of whom are modeled as making optimal choices given a set of constraints.

[6] For researchers working with other animals, the reward medium is typically food. See Brosnan et al. (2011) for a direct comparison of behavior in strategic games when the subjects are humans, capuchin monkeys, and chimpanzees. The question as to whether strategic undergraduate behavior is more similar to that of nonhuman primates or professionals remains unresolved.

must rely upon reason. The fourth precept, privacy, holds that subjects should be given only information regarding their own payoffs. This is not to say that people do not care about the allocations of others, in fact, many lab experiments have shown that people are concerned about the payoffs of others. Rather, the reward medium in the laboratory replaces the endogenous preferences of the subject and one cannot observe the preferences of others in the naturally occurring economy. Sometimes agents in the naturally occurring economy can observe the allocation that is received by other agents, and this information may also be observable within an experiment if appropriate for the given context.

The fifth precept, parallelism, pertains to the ability to extrapolate from the laboratory to another economic system, typically one that has arisen naturally in the world. Parallelism holds that behavior is the result of agent incentives for a given institution and that this behavior holds in any economy where the appropriate ceteris paribus assumptions hold. This raises the issue of what is meant by an institution. An *institution* is the collection of rules, procedures, and property rights that govern what actions are available to an agent, including the allowable sequencing and observability of those actions. The institution also specifies how the selected actions of the agents are converted into outcomes and the rights of the agents to claim the reward medium. The notion of an institution is distinct from that of the environment. *Environment* refers to the set of characteristics of all agents in the economy, including endowments and preference. Within this framework, the researcher can control the institution and the environment and observe the resulting behavior or choices of the subjects within the created economy (for a more detailed discussion, see Smith 1989). A researcher can systematically vary aspects of the institution and the environment to explore how these factors impact behavior. If the first four precepts hold, the researcher can infer that that observed behavioral changes are due to the changes in the institution or the environment. If the precept of parallelism holds, then the results can be extended to other economic settings as well.

EXAMPLES OF CONTROLLED EXPERIMENTS WITH RELEVANCE TO LOGISTICS AND SUPPLY CHAIN

Procurement Auctions

Numerous laboratory experiments have investigated how specific auction formats impact prices (for reviews, see Kagel and Roth 1995; Plott and Smith 2008). Two of the most prevalent auction types for procurement purposes are the first price sealed bid auction and the English auction. In the first price sealed bid auction, each bidder privately submits a bid. The winner of the auction is the bidder offering the most favorable terms to the party holding the auction and the transaction price is equal to the winning bid. In an English auction, the price starts in a position favorable to the bidders so that multiple bidders are willing to trade at the starting price. The price is gradually incremented in a direction that is favorable to the party holding the auction until a single bidder remains. This sole remaining bidder is declared the winner and the transaction price is set at the prevailing price when the other bidders dropped out of the auction. In a procurement setting, the bidders are sellers so that low prices are deemed favorable to the party holding the auction, the buyer. Thus, in the English procurement auction the prices decline leading to the term "reverse auctions." This is an auction format distinct from the Dutch auction, which has a declining price when the party holding the auction is the seller. In a Dutch auction, prices begin in a position favorable to the party holding the auction and become more favorable to the bidders over time. The winner of a Dutch auction is the first bidder to indicate willingness to trade at the current price and the trade is completed at that price.

As described here, these institutions are incompletely specified. To actually implement an auction in or out of the laboratory, one must specify how much time a bidder has to submit a bid in a first price auction as well as the increment in which one can bid and a tie breaking rule. In an English or Dutch auction, one must specify a starting price and the speed and increment for adjusting prices, among other factors. While most theoretical treatments of auctions are silent on such matters, laboratory evidence suggests that these aspects of an auction institution can matter greatly. For example, Katok and Kwasnica (2008) report that faster speeds in the Dutch auction lead to prices that are more favorable to bidders. This also highlights the fact that there is not a single kind of first price auction and English auction, but rather a whole class of institutions that can be thought of as first price auctions or as Dutch auctions. Deck and Wilson (2008b) demonstrate the care that should be taken in designing experiments to ensure a neutral comparison across institutions and failure to do so can have detrimental effects.

Traditional procurement auctions were conducted using sealed bids in which bidders not only specified a price, but could also indicate a level of quality, customization in specification, or flexibility in delivery. When the bids were opened a buyer could choose the best bid based on some possibly unspecified multidimensional criteria. With the dawning of the Internet and electronic commerce, many procurement auctions began using an English reverse price institution. The perceived advantages included opening bidding to a larger number of participants, quicker resolution of the outcome, and the possibility of fiercer competition leading to better prices. However, if price is the only dimension on which bids could be communicated within the institution then one cannot distinguish quality or other properties of a supplier's offering.[7] This is not problematic if bidders have homogenous products, but if there are varying quality levels or if flexibility is costly then these higher costs will preclude suppliers of these preferred products from being able to win the auction, resulting in low quality, low price suppliers winning the auction. Thus, the specific form of the English auction can force the focus to be on price at the expense of other attributes. The same would have been true if sealed bids were forced to only specify a price. Similarly, one could construct an English auction that takes quality into account by allowing there to be concurrent prices for different levels of quality. Thus, any general claim

[7]Quality measures—make, year, model, mileage, etc.—of automobiles and trucks enable online auctions to function exceptionally well. See, for example, http://www.copart.com/c2/copart_home_page.html for upcoming and active regional auctions in the United States.

that first price sealed bid auctions are preferable to reverse price auctions due to quality issues while ignoring the specific implementations is a bit like saying a starving person prefers potatoes to steak because he chooses two large baked potatoes over a single bite of steak.

The environment in an auction includes the induced value of each bidder, the information that each bidder has about the values of the other bidders, and both the actual and perceived number of bidders. In an experiment that systematically changed the number of bidders, Cox et al. (1982, 569) "…conjectured that the deviant results for the case $N = 3$ in both first- and second-price auctions are due to failure of the assumptions of noncooperative behavior…" in the theory. However, this deviant behavior for $N = 3$ failed to replicate in a subsequent series of experiments (Cox et al. 1988) showing the importance of replication in testing conjectured explanations of cases where a theory fails. Symmetrically, the same methodological caution applies also where the theory does not fail. Pari passu, field experiments need to be subjected to the same caution. Moreover, if results differ between a field and a lab experiment, *both* should be replicated.

One of the most prominent theoretical results in the auction literature is the revenue equivalence theorem, which predicts equal prices across the four standard auction formats under the assumption of risk neutrality and a uniform distribution of bidder values.[8] In the laboratory, the experimenter can easily control both the distribution of bidder values and the number of bidders. However, laboratory results consistently reveal that revenue equivalence does not hold. In comparing the theoretically isomorphic first price and Dutch auctions, bidders regularly act as if they are more risk averse in the first price auction (see e.g., Cox et al. 1982; Turocy et al. 2007; Deck and Wilson 2008b). Field experiments as conventionally designed seem unlikely to be able to rescue these negative results, as the findings are dependent on independent knowledge of individual values for chosen items.

The laboratory can also offer insight beyond what is observable in the field. For example, Cox et al. (1996) study procurement auctions where sellers have an unknown cost at the time of the auction and can make unobservable investments to reduce costs. By definition, unobservable costs cannot be measured in field data, but the researcher can perfectly measure such costs in the lab, as the action must be undertaken by the seller and can thus be made observable to the researcher while hidden from the buyer. The laboratory experiments reveal that higher cost-sharing rates lead to less discretionary cost-savings measures by the selected suppliers. The results also indicate that cost-sharing contracts lead to lower efficiency than fixed price contracts.

Divorcement and zone pricing in wholesale gasoline markets

Deck and Wilson (2008a) create a laboratory economy, which mimics characteristics of the market for refined gasoline, but which can be applied to other industries as well. In this economy, branded retailers (gas stations) are located at geographically dispersed locations along a city grid of avenues and streets. Customers are also located at different points in the city and some customers have idiosyncratic preferences for certain brands, meaning that a shopper's willingness to pay may be greater for some particular brand. Customers incur travel costs to visit a seller and purchase from the retailer offering the best combination of price and location. The values, brand preferences and travel costs of the customers, all constitute a part of the environment in this experiment and are controlled by the experimenters. Retailers in this economy operate multiple outlets; one in the city center close to rivals and another in the more isolated periphery. As each retailer offers a specific brand, that retailer is contractually obligated to purchase inventory from the wholesaler (refiner) who produces that brand. In the experiment, the wholesaler's cost (world oil prices) followed a random walk, also a part of the environment and under the control of the researcher. The economy's institution allowed wholesalers to observe their own costs, but not that of the other wholesalers, and the prices set at all retail outlets. With this information, wholesalers could set the retail outlet-specific price at which the retailer was obligated to purchase inventory. Retailers observed their rival's prices and their own wholesale price, but not that of their rivals, and in turn set their retail prices. The institution allowed everyone to see the purchase decision of a series of customers and adjust their own prices at will in real time.

Deck and Wilson (2008a) compare behavior in the baseline condition as described above to two alternative institutions that had been proposed by various groups as policies that would benefit consumers and ensure greater competition. The first policy proposal involved divorcement, forbidding a wholesaler from operating at the retail level. This policy prescription was tested by creating vertically integrated firms that operated in the laboratory market. The results clearly demonstrated that vertical integration actually lowered prices in both the competitive central areas and in the periphery, consistent with the theoretical prediction driven by the elimination of a double mark-up.

The second policy proposal was to ban the practice of zone pricing under which a wholesaler could set differential prices based on the geographic location of the retail outlet. This policy was tested by forcing each wholesaler in the laboratory economy to set a single wholesale price. The results showed that the practice of zone pricing did not impact retail prices (which were allowed to differ by location consistent with the legislation that was being proposed) and thus was not harmful to consumers. What the experiments did reveal is that zone pricing transfers profit from retailers to wholesalers; hence the fact that retailer associations were lobbying to ban zone pricing is not surprising in retrospect. The laboratory economies also displayed the "rockets and feathers" phenomenon that has been observed in gasoline and other markets. While prices that increase at a faster pace than they fall has been held up as evidence of collusion in naturally occurring markets, the subject sellers in the laboratory had no ability to collude and thus the asymmetric pricing dynamic appears to be a result of the strategic pressures in an economy and not evidence of clandestine back room deals.

Inventory replenishment

Several studies have used the beer distribution game to investigate inventory replenishment decisions in a supply chain (for a

[8] The fourth institution is the second price sealed bid auction in which the winner is the bidder submitting the most favorable bid from the perspective of the party holding the auction, but the transaction price is set at the second most favorable bid.

review, see Croson and Donohue 2002). Subjects each play the role of a decision maker at different echelons of the supply chain and order inventory from the immediate upstream supplier to fill orders placed by the immediate downstream customer. When there are delays in filling orders due, for example, to production or shipping times and each decision maker incurs costs for holding inventory and penalties for delays in filling orders, the resulting pattern of orders resembles a bullwhip with oscillation (variance at each echelon) and amplification (increasing variance across echelons up the supply chain). In these experiments, the researcher can control whether or not information can be shared across the supply chain (as determined by the institution) or the lag length (as specified in the environment). As discussed in Croson and Donohue (2002) and references therein, reducing lags and sharing information tends to tame the bullwhip.

A more recent experiment by Tokar et al. (2011) demonstrates how specific informational sharing practices can impact inventory replenishment in the face of positive demand shocks and lags in order fulfillment. The experiment looked at a single echelon of the supply chain via a newsvendor problem because the focus was on individual behavior in response to information rather than the strategic interaction present in the beer distribution game. The derived demand environment facing the subject decision maker featured a retail promotion-driven spike in a particular period. The treatment conditions varied the institutional information that could be shared by the automated retailer and the human supplier. In some instances, the supplier knew the size of the demand spike and in others, the supplier knew of the timing of the spike. As in research on the bullwhip, Tokar et al. (2011) conclude that more information sharing enhances the efficiency of the supply chain.

Liberalization of the electric power industry

Previous electric power experiments serve as a clear counter example to two propositions: (1) "this problem is too complicated for a lab experiment" and (2) "field experiments are more informative" than lab experiments. It is also an example in which metaphors, such as wind tunnels and space simulators, serve to convey the sense of a methodology that combines experiment, system design, and management policy issues in an appropriate learning environment. This methodology can require years of intensive effort combining the specialized skills of industry managers, government representatives, technicians, engineers, and whoever else brings needed input to the project. It is these requirements that are often ignored, contributing to a misunderstanding of the role of lab experiments in applications to the economy. Any innate skepticism that members of these research teams might bring initially to the project is overcome by open give-and-take data driven discussion sessions and answers based on demonstration. They are overcome largely because there are no comparable competing methodologies other than the try-it-and-correct-it approach in which responsibility for the relevance, reliability, and replicability of the results is shared by all members of the team.

Electric power networks offer mind boggling complexity: their dynamic transient behavior is far from being fully understood by engineers; it can take days if not months to identify the causes and transient circumstances of a disturbance that cascades into a major blackout. Measured in terms of the population size affected, the largest blackout in U.S. and Canadian history occurred on August 14, 2003, but the final report identifying its multiple interdependent causes was not issued until April 2004 (U.S.-Canada Power System Outage Task Force 2004). The U.S. and Canadian blackout was a consequence of four control-area management failures: (1) of agent assessment and understanding of system inadequacies, (2) of agent situational awareness, (3) of adequate agent control of tree growth in transmission rights-of-way, and (4) of reliability organization diagnostic support (U.S.-Canada Power System Outage Task Force 2003, 18–19). The initiating circumstance happened to be a high voltage transmission line fault due to tree invasion, but faults are ubiquitous from storms, and it was the cascade of inadequate agent responses afterward that brought down the system as backup safety devices shed loads and interconnects to protect network damage. Any experimental economist will see that all these failures are related to human incentives, and experiments—lab or field—are most prominently about the conjunction of rules and incentives, and their role in operationally successful decentralized control over system performance. Power systems today are not controlled from the top down, but from the bottom up with the interdependent parts, all governed by rules and incentives as well as decentralized sensing, monitoring, and control devices. The original and continuing design challenge is to evaluate and test alternative rule/incentive specifications, which are what constitutes property rights, defined as rights to take action.

Electric power research programs using laboratory and field research methods and interdisciplinary teams of industry executives, engineers, and government representatives began over a quarter of a century ago (1984–85) and are continuing today. The Arizona Corporation Commission report by Block et al. (1985; hereafter, ACC Report) initiated the development of laboratory investigations into the feasibility and implementation of the design of markets for the control of electric power production, transmission and delivery to retail customers. Because a published review of this history is available, we provide only a very brief summary of its main developments, and an update of more recent references (see Rassenti et al. 2002):

1. Engineering cost minimization algorithms for loading dispersed generators on a transmission line in regulated or government owned systems could be implemented to computer assist an electric power market in which wholesale buyers submitted real time location-specific bids for delivery, and generator companies submitted location-specific offers to sell and inject power into the grid. This became known as one of many examples of a "smart computer assisted market" (McCabe et al. 1991).
2. Facilities subject to economies of scale, capital indivisibilities or common pool "externalities" could be structured as cotenant joint ventures and governed by property right rules that share capacity drawing rights in proportion to the fixed capital cost contributed by each cotenant and variable costs in proportion to real time shares of output taken. As it was learned later, the concept had emerged historically in the newspaper industry where morning and

evening city newspapers shared the construction and fixed operating costs of a printing facility each paying for the set up and operating cost of their morning and evening print runs. Suitably ruled, each could compete for customers in the market, however, avoid duplicating separate printing facilities. Hence, institutional restructuring could allow market competition to discipline prices, whereas economies of scale, productive input, and capital cost sharing could substitute for central ownership or regulatory control.

3. It was recognized that pipes and wires are separable from the energy delivered by means of them and one could retain or modify elements of the federal or state regulatory framework, but apply it only to the pipes or wires. In the late 1980s, at the wholesale level, this concept informed and guided the Department of Energy and Federal Energy Regulatory Commission (FERC) in their concerted efforts in deregulating wellhead gas and its transmission to city gates, where state regulations applied. By the 1990s, FERC was applying this approach to wholesale electricity, and many states were instituting retail customer choice among competing electricity providers, with separate billing for energy from the provider and for wire rental from the wires company.
4. The ACC Report led to the development of a prototype electricity network in the early 1990s used initially for demonstrating the feasibility of creating a market for wholesale power that could enlist industry and government representatives as participants and provide a specific focus enabling all parties to jointly address the question: What are the strengths and weakness in this experimental model and how can it be transformed into a system that could be implemented in the field?[9] This helped to set the agenda for subsequent reform in New Zealand and Australia. In Australia, the model led to a design for trading based on the parameters of that country's grid, the first field implementation in Victoria and New South Wales in 1996, and the Australian National Electricity Market initiated trading on December 13, 1998.[10] In the intervening years, this market has undergone almost continuous adaptation and revision in the light of operating experience and technological change.

Deregulation of the natural gas industry

Based on the success of experiments in electric power markets, FERC and the Energy Information Administration moved to test the performance properties of the competitive-ruled joint venture institution in a prototype natural gas pipeline network for the production, transportation, and delivery of gas in the wholesale market.[11] The objectives were "to develop an exchange mechanism for natural gas pipeline networks that will correct the three most common criticisms of the industry: gas is inefficiently allocated; gas is unfairly priced (e.g. due to 'excessive' price discrimination by pipelines); prices are excessively volatile" (Rassenti et al. 1994, 42). A pipeline network is technically less complex than an electric network as the flows from producers to consumers can be fully controlled through choke valves as well as compressors.

FERC personnel specified the network studied initially in McCabe et al. (1989), which led to a subsequent nine node variation in Rassenti et al. (1994). These networks had several features reflective of conditions in the industry: (1) most buyers have multiple sources and (2) the commodity can be delivered over alternative possible transportation paths from wellhead sources to wholesale buyers. Because pipelines at the time had ownership or contract interests in the wells they served, the experiments endowed pipeline agents with less than one-half of the production capacity in each of the three producing fields. The trading procedure designed for study was a smart computer assisted mechanism that accepted real time location-specific price and quantity bids to buy delivered gas by each buyer, location-specific offers to sell gas inject by each producer, and leg-specific offers to transport gas between leg origin and leg end points. As bids and offers arrived in real time, a system-wide linear optimization program computed prices at each node, and allocations to each agent that maximized system surplus based on these dispersed bid revelations. The algorithms have no knowledge of, and therefore could make no use of, the private buyer valuations of gas or supply chain private costs of producers and transporters. Experimenter knowledge of all induced values and costs was used to evaluate the efficiency of the system design under alternative treatments.

The primary treatment condition introduced cotenants who shared capacity ownership. In this treatment, some of the pipeline segments were assigned three cotenants: one was an owner as in the baseline twin comparison experiment, who was also an owner of one of the producing well fields, and two cotenants owned no well production capacity. What was most important about the experiments was that each cotenant submitted independent competing offers. Explicit collusion was prevented, although nothing prevented forms of tacit collusion that might emerge to manipulate prices, restrict allocations, and reduce efficiency. Cotenants on each pipeline leg shared common variable costs in proportion to their share of transportation throughput allocated to them.

[9] A version of this prototype design and many of the experiments run with it, were subsequently published in Backerman et al. (2000). Far more complex experiments growing out of interactions with interested industry groups were published in engineering and science journals (see Olson et al. 2003; Rassenti et al. 2003).

[10] In contrast with the Australian experience, the same approach was later used in an attempt to frame the design of California's aborted effort to deregulate electricity, but failed to enlist industry and government support (see Smith 2008, 303–306).

[11] Georgia deregulated the retail distribution of natural gas in 1997. The local pipeline distribution system remained regulated, but gas was supplied by individual providers competing for customer accounts (see http://www.psc.state.ga.us/gas/ngdereg.asp). For an investigation and evaluation of the Georgia gas market, see Blue Ribbon Natural Gas Task Force (2002).

The experimental results were consistent with the proposition that competitively ruled cotenancy property right regimes were capable of improving the performance efficiency of a natural gas supply chain network. In addition to overall system performance, the data permit comparisons between the equilibrium shares of producers, transporters, and buyers. The experiments showed that all economic interest groups in the network market gained under the cotenancy treatment.

DIRECTIONS FOR FUTURE BUSINESS LOGISTICS RESEARCH IN THE LABORATORY

The potential applications of controlled laboratory experiments in the supply chain and logistics fields are as varied as the field itself. Possible directions include looking at transaction costs as a driver for adopting environmentally friendly practices (discussed theoretically by Tate et al. 2011), how a supply chain serving multiple retailers impacts inventory decisions (discussed based on simulations by Wan and Evers 2011), and how order fulfillment impacts customer retention (discussed empirically by Rao et al. 2011). Laboratory experiments can be used to inform retailers about how customers will react to realized or anticipated stock-outs, such as under what conditions will customers substitute for missing products and under what circumstances will they simple go to another provider. Laboratory experiments can be used to evaluate alternative inventory management systems. For example, a multioutlet retailer could allow store level managers to bid on inventory allowing these intermediaries to express the intensity of their needs. Alternatively, one could allow store level managers to trade inventory between themselves. Experiments could also inform policy makers about the differential economic impact proposed fuel efficiency measures might have on national long haul carriers versus smaller regional trucking companies. The laboratory could also serve as testing ground for various public-private partnerships, such as might be used to help the ports of Los Angeles and Long Beach improve their rail access to compete with a widened Panama Canal for handling shipping between Asia and the Eastern United States.

Laboratory experiments have the potential to inform a myriad of research questions in logistics and supply chain management. We hope that this study both provides a structural foundation for the use of experimental economics methods in this area and serves as inspiration for researchers to develop ever more creative applications of laboratory experiments.

REFERENCES

Ariely, D., Gneezy, U., Lowenstein, G., and Mazar, N. 2009. “Large Stakes and Big Mistakes.” *The Review of Economic Studies* 76:451–69.

Backerman, S., Rassenti, S., and Smith, V. 2000. “Efficiency and Income Shares in High-Demand Energy Networks: Who Receives the Congestion Rents When a Line is Constrained?” *Pacific Economic Review* 5:331–47.

Besedes, T., Deck, C., Sarangi, S., and Shor, M. 2012. “Age Effects and Heuristics in Decision Making.” *Review of Economics and Statistics* 94:580–95.

Block, M., Cox, J., Isaac, R.M., Pingry, D., Rassenti, S., and Smith, V. 1985. “Alternatives to Rate of Return Regulation.” Final Report of the Economic Science Laboratory, University of Arizona, to the Arizona Corporation Commission, February 15.

Blue Ribbon Natural Gas Task Force. 2002. “Final Report to Governor Roy E. Barnes and General Assembly of the State of Georgia, February 5.

Brosnan, S., Parrish, A., Beran, M., Flemming, T., Heimbauer, L., Talbot, C., Lambeth, S., Schapiro, S., and Wilson, B. 2011. “Responses to the Assurance Game in Monkeys, Apes, and Humans Using Equivalent Procedures.” *Proceedings of the National Academy of Sciences* 108:3442–47.

Camerer, C., and Hogarth, R. 1999. “The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework.” *Journal of Risk and Uncertainty* 19 (1–3):7–42.

Carpenter, J., Verhoogen, E., and Burks, S. 2005. “The Effect of Stakes in Distribution Experiments.” *Economics Letters* 86:393–98.

Cox, J., Isaac, M., Cech, P., and Conn, D. 1996. “Moral Hazard and Adverse Selection in Procurement Contracting.” *Games and Economic Behavior* 17:147–76.

Cox, J., Roberson, B., and Smith, V.L. 1982. “Theory and Behavior of Single Object Auctions.” In *Research in Experimental Economics*, Vol. 2, edited by V.L. Smith, 537–79. Greenwich: JAI Press.

Cox, J.C., Smith, V.L., and Walker, J.M. 1988. “Theory and Individual Behavior in First-Price Auctions.” *Journal of Risk and Uncertainty* 1(1):61–99.

Croson, R., and Donohue, K. 2002. “Experimental Economics and Supply-Chain Management.” *Interfaces* 32(5):74–82.

Deck, C., and Wilson, B. 2008a. “Experimental Gasoline Markets.” *Journal of Economic Behavior and Organization* 67(1):134–49.

Deck, C., and Wilson, B. 2008b. “Fixed Revenue Auctions.” *Economic Inquiry* 46(3):342–54.

Denton, M., Rassenti, S., and Smith, V.L. 2001. “Spot Market Mechanism Design and Competitivity Issues in Electric Power.” *Journal of Economic Behavior and Organization* 44(4):435–53.

Dyer, D., Kagel, J.H., and Levin, D. 1989. “A Comparison of Naive and Experienced Bidders in Common Value Offer Auctions: A Laboratory Analysis.” *The Economic Journal* 99:108–15.

Falk, A., and Heckman, J. 2009. “Lab Experiments Are a Major Source of Knowledge in the Social Sciences.” *Science* 326:535–38.

Haigh, M., and List, J. 2005. “Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis.” *The Journal of Finance* 60(1):523–34.

Hoffman, E., McCabe, K., and Smith, V.L. 1996. “On Expectations and Monetary Stakes in Ultimatum Games.” *International Journal of Game Theory* 25(3):289–301.

Holt, C., and Laury, S. 2002. “Risk Aversion and Incentive Effects.” *The American Economic Review* 92:1644–55.

Kagel, J., and Roth, A. (eds.) 1995. *The Handbook of Experimental Economics*. Princeton, NJ:: Princeton University Press.

Katok, E., and Kwasnica, A. 2008. “Time is Money: The Effect of Clock Speed on Seller’s Revenue in Dutch Auctions.” *Experimental Economics* 11:344–57.

King, R., Smith, V., Williams, A., and Van Boening, M. 1993. "The Robustness of Bubbles and Crashes in Experimental Stock Markets." In *Nonlinear Dynamics and Evolutionary Economics*, edited by R.H. Day and P. Chen, 183–200. New York, NY: Oxford University Press, USA.

McCabe, K., Rassenti, S., and Smith, V.L. 1989. "Designing 'Smart' Computer Assisted Markets: An Experimental Auction for Gas Networks." *European Journal of Political Economy* 5:259–83.

McCabe, K., Rassenti, S., and Smith, V.L. 1991. "Smart Computer-Assisted Markets." *Science* 254:534–38.

Olson, M., Rassenti, S., Smith, V., and Rigdon, M. 2003. "Market Design and Motivated Human Trading Behavior in Electricity Markets." *Institute of Industrial Engineering Transactions* 35(9):833–49.

Plott, C., and Smith, V. (eds.) 2008. *Handbook of Experimental Economics Results*. Vol. 1. Amsterdam: North-Holland.

Rao, S., Goldsby, T., Griffis, S., and Iyengar, D. 2011. "Electronic Logistics Service Quality (e-LSQ): Its Impact on Customer's Purchase." *Journal of Business Logistics* 32 (2):167–79.

Rassenti, S., Reynolds, S., and Smith, V.L. 1994. "Cotenancy and Competition in an Experimental Auction Market for Natural Gas Pipeline Networks." *Economic Theory* 4:41–65.

Rassenti, S., Smith, V.L., and Wilson, B. 2002. "Using Experiments to Inform the Privatization/Deregulation Movement in Electricity." *The Cato Journal* 21:515–44.

Rassenti, S., Smith, V.L., and Wilson, B. 2003. "Controlling Market Power and Price Spikes in Electricity Networks: Demand-Side Bidding." *Proceedings of the National Academy of Sciences* 100(5):2998–3003.

Samuelson, P., and Nordhaus, W. 1985. *Economics*. 12th ed. New York: McGraw-Hill.

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

Smith, V.L. 1989. "Theory, Experiment and Economics." *The Journal of Economic Perspectives* 3(1):151–69.

Smith, V.L. 1994. "Economics in the Laboratory." *The Journal of Economic Perspectives* 8(1):113–31.

Smith, V.L. 2008. *Discovery: A Memoir*. Bloomington, Ind.: Authorhouse.

Smith, V.L., Suchanek, G., and Williams, A. 1988. "Bubbles, Crashes and Endogenous Expectations in Experimental Spot Asset Markets." *Econometrica* 56:1119–51.

Tate, W., Dooley, K., and Ellram, L. 2011. "Transaction Costs and Institutional Drivers of Supplier Adoption of Environmental Practices." *Journal of Business Logistics* 32(1):6–16.

Tokar, T.J., Aloysius, M.W., and Williams, B. 2011. "Retail Promotions and Information Sharing in the Supply Chain: A Controlled Experiment." *International Journal of Logistics Management* 22:5–25.

Turocy, T., Watson, E., and Battalio, R. 2007. "Framing the First-Price Auction." *Experimental Economics* 10(1):37–51.

U.S.-Canada Power System Outage Task Force. 2004. "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations." April.

Wan, X., and Evers, P. 2011. "Supply Chain Networks With Multiple Retailers." *Journal of Business Logistics* 32(1):27–39.

SHORT BIOGRAPHIES

Cary Deck (PhD, The University of Arizona) is a Professor of Economics in the Sam M. Walton College of Business at the University of Arkansas where he serves as the Director of the Behavioral Business Research Laboratory. He has published numerous articles in leading journals like *American Economic Review*, *Information Systems Research*, *Journal of Conflict Resolution*, *Review of Economic Studies*, and *Review of Economics and Statistics*.

Vernon Smith (PhD, Harvard University) is the George L. Argyros Endowed Chair in Finance and Economics in the Economic Science Institute at Chapman University. In 2002 he was awarded the Nobel Prize in Economic Science. He is a member of the National Academy of Sciences and President of the International Foundation for Research in Experimental Economics. He has published numerous articles in leading journals, edited many collected volumes, and authored several books.