

**DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY**

PASADENA, CALIFORNIA 91125

ECONOMIC RESEARCH CENTER
DEPARTMENT OF ECONOMICS
UNIVERSITY OF MINNESOTA

DIMENSIONS OF PARALLELISM: SOME
POLICY APPLICATIONS OF EXPERIMENTAL METHODS

Charles R. Plott



SOCIAL SCIENCE WORKING PAPER 569

April 1986

DIMENSIONS OF PARALLELISM: SOME
POLICY APPLICATIONS OF EXPERIMENTAL METHODS

Charles R. Plott*
California Institute of Technology

This paper outlines several strategies that have been employed in attempts to use experimental methods for actual policy decisions. The topic to be explored is how policy issues have been posed in policy related studies so that experimental methods could be applied. The discussion is limited to ten instances that have involved me personally at some level of effort.

Many different opinions exist about experimental methodology and the relationship between laboratory work, field studies, and policy decisions. My impression is that the opinions are strongly held and are just as likely to be held by those with no experience at all in applying the methods as by those with much experience. For example, the new textbook by Samuelson and Nordhaus, 1983 (p. 8) boldly claims that experiments in economics are impossible. Presumably they have some sort of field study in mind as the only possibly relevant way to approach an application of experimental methods. Referee reports frequently reflect methodological philosophies. Every experimentalist who has submitted a paper to a professional journal has seen a referee report aggressively claiming that the experiments had nothing to do with the "real world" or that the experiments were not "relevant" for

some reason or another. My own impression is that the connections between laboratory and naturally occurring situations come in several different forms.

The impression brings me to my point. Economists should keep an open mind about experimental methodology. Methodological principles should evolve from our experiences with what works and what does not work. That point is reflected in the title and organization of this paper. The topic is policy research as opposed to basic research. The issues are what was attempted? what seemed to work and why? what was a flop and why?

The material is organized according to what seems to have been the principal strategy for using the experiments. Five different strategies are identifiable. Each section treats a different strategy. The discussion includes a general description of the strategy, the context of the policy problem, and the role of the experiments in the final decision if any decision resulted.

I. EX POST EVALUATION OF A DECISION: THE FLYING CLUB

A policy decision was made. An act was taken. A result was observed. What influence did the policy have on the observed result? The issue is the possible role of experiments in the ex post evaluation of a policy decision.

The ex post evaluation of a policy decision motivated the agenda experiments reported in Levine and Plott (1977). The policy decision was to promote the use of a particular agenda by a large

flying club that was selecting a fleet of aircraft. Unknown to all members of the club but one, the agenda was designed to influence the club to choose the fleet of aircraft most preferred by the one. The theory used to design the agenda was untested and the actual preferences of the members of the club were largely unknown. Nevertheless, the club used the agenda and chose the fleet that the agenda was designed to get. Was the agenda responsible for the group choice? The question can never be answered but educated guesses are possible.

Some preference data for members of the club became available through a questionnaire that was circulated after the decision. A series of experiments was conducted with a set of induced preferences. The induced preferences were basically the same as those of the club members that became available. Changes made in the preferences for purposes of experimentation reflected an attempt to make the situation somewhat harder for the agenda than would have been the case with the reported preferences.

The experiments involved a smaller set of options than the flying club actually faced. Letters of the alphabet were used as options without any reference to airplanes, the flying club, etc. Monetary incentives were used to induce preferences over options. The amount a particular individual would get given the group choice of an option was known only to that individual reflecting the fact that the magnitude of happiness or satisfaction by any member of the flying club was unknown to others. The neutral labels on the option were used to avoid the possibility that experimental control would be lost due to subjects

acting on motivations stimulated by the option labels. The agenda used was mathematically identical to the one used by the flying club.

Experiments were conducted. The influence of the agenda was decisively demonstrated by changing the agenda while holding preferences fixed. The agenda influences the decisions in experiments. Little doubt exists about that.

Did the agenda influence the outcome of the flying club meeting? Two assumptions are necessary for an inference. (1) The preferences of the flying club were near those that were induced in the laboratory. (2) The relationship between the act of voting and preferences was the same for the individuals in the flying club meeting as it was for those individuals in the laboratory. If both assumptions are accepted then the agenda must have had an influence on the outcome of the flying club.

So the question of relevance becomes focused on (1) and (2). If (1) is doubted new experiments can be conducted in which the induced preferences more closely approximate those hypothesized for the club members. In principle all preference patterns could be examined so assumption (1) provides no problem for the application of experiments. Any criticism along the lines of (1) is not an objection to the use of experiments. Quite the contrary, it is a call for more experiments.

Assumption (2) really involves a theory of behavior. In essence it requires the hypothesis that the voting decisions of all people including those of the flying club and those in our experiments, can be reasonably captured by the model used to design the agenda. The

acceptance of this general theory is a key to the application. To the extent that it can be demonstrated as unreliable the conclusions drawn from the theory can be challenged. But again we discover that the challenge does not involve an objection to the use of experimental method. Instead the challenge is a call for additional theory and perhaps more experiments. Additional theory is simply an improved replacement of the old and the additional experiments are tests of it.

As it turns out the basic theory of the agenda appears to work well in the laboratory setting. The extent to which one is prepared to assert something about the flying club seems to be an unavoidable matter of subjective judgment concerning the confidence one places in the two assumptions. A variety of preferences have been checked and no exceptions to the behavioral theory have been exhibited (so far).

II. DEMONSTRATION: LANDING SLOT ALLOCATIONS

On occasion the implications of theory are so clear and the results of previous experiments are so unambiguous that professionals have little to learn from experiments. Nevertheless, the theory that seems so obviously relevant to professionals is frequently abstract and complicated to those who have the power and responsibility to make decisions. Sometimes in a policy-making environment even the word "theory" is pejorative. In such cases experiments provide a way of demonstrating the ideas without resort to theoretical constructions. The role of experiments as a demonstration was the basis of the Polinomics report (Grether, Isaac, and Plott, 1979) on the allocation

of airline landing slots.

After the airline deregulation act of 1976 staff members of the Civil Aeronautics Board became concerned about the method of allocating the right to land at four major airports (Washington National Airport, Kennedy, La Guardia, O'Hare). The allocation decisions were made by committees. Each airport had a separate committee consisting of those airlines that had been certificated by the CAB to operate at the airport.

In 1968 the Federal Aviation Administration had limited the number of slots (takeoffs and landings per hour) that could be conducted at each of these airports. The committees were charged with the task of determining by "agreement" the allocation of slots among the certificated carriers. What might happen if the committees failed to reach unanimity was unclear. The FAA might have administratively allocated the slots but the criterion that the FAA might have used and the role of politics in the process made the exact consequences of default uncertain. Since the committees had successfully achieved agreement every six months from 1968 until the time of the study (1978), what might have happened if the committees had failed was only a matter of speculation.

The situation changed with the Airline Deregulation Act. The CAB staff became concerned that the committees could be used as a barrier to new competition. I was contacted to study the committees because of my previous work on committee behavior. The question posed related to the degree to which the committee process of allocating slots was

compatible with the Airline Deregulation Act.

After some study and attendance at committee meetings, the nature of the committee behavior became clear and the structure of a reasonable process that could replace the committees became apparent. The appropriate model of the committee would have become obvious immediately to anyone with some game-theoretic and economic training. The committee operating under unanimity would attain some point in the core (imputations) in the appropriate game without side payments. The location of the core would be very sensitive to the beliefs about what would happen should the committee default. Those beliefs determined the threat points which bound the imputations. Since there were no side payments, the allocation would not necessarily be related to the efficient allocations. That is, the airlines that would acquire the slots under the committee process would not necessarily be the carrier that valued the slots the most in a cost/benefit sense. Furthermore the allocation would be sensitive to the beliefs about the consequences of default. Reasonable replacement processes involved markets with initial allocations determined by auction, or by lottery, or perhaps grandfathered.

The role of the experiments was twofold. First, the experiments were used to demonstrate the implications of the game-theoretic model used to evaluate the committee process. The emphasis is upon the word demonstrate, because the implications of the model were fully understood theoretically at the time, and previous experiments left little doubt that the substantive implications of the model would be predictive of

committees operating under laboratory conditions. For those who had previously studied a wide range of committee experiments, very little was to be learned from additional experimentation.

The audience which consisted of CAB staff, DOT staff, the FAA staff and the airlines had no previous experiences with committee experiments. Nor did the audience understand or have a tendency to accept game theory. Thus, given the political and controversial nature of the issue, some sort of demonstration that the theory had content was necessary. The structure of the experiments was to demonstrate the major consequences of the theory when applied to the slot allocation committee process. The intent was to demonstrate the theory while avoiding any detailed discussion of the content of the theory, the axioms or mathematical structure, and also avoiding any long and academic discussions about why the theory might be true. The strategy was one of demonstration.

Parameters for the committee experiments were chosen to reflect beliefs about the actual committee parameters. Part of the study involved a demand curve estimation for a certain period at Washington National Airport. The experiment used those parameters scaled down appropriately. Subjects in most experiments were adults preferably with some connection to the aerospace industry (e.g., aeronautical engineers). These decisions were made in anticipation of a criticism or claim that the committees under the field parameters would behave differently and/or that people from the industry are different. In addition, participants in some committees made several decisions

together thus reflecting the fact that sequences of decisions are characteristic of the slot committees.

The experimental design for committees made three points. (1) The outcome of the committee process is sensitive to the consequences of default. The point was made through experiments with identical preference parameters but different default rules. (2) The committee processes with different committee meetings for different airports could not deal efficiently with interdependencies between the airports. A carrier might not want a slot to take off from O'Hare if it did not have a slot to land at DCA. This point was made by creating preference interdependencies or complementarities between the choice variables of committee meetings. (3) The committee process would be insensitive to profitability of carriers and thus not promote an efficient allocation of resources. This point was made along with (1) by inducing very high values for slots for some participants and very low values for others.

The results of the committee experiments were not controversial. All three points were clearly evident in the data. Under unanimity a great pressure exists for equality of distribution and unless a large allocation could be protected by a threat point participants had difficulty keeping it. For example, large carriers that should grow according to the economics of the situation never did so and usually contracted under the committee process. Inefficient carriers that should leave the airport never did so under the committee. The experiments provided a means by which the consequences of the theory for the allocation of airport resources could be communicated without

reference to the theory itself.

The report proposed the creation of a market for slots to replace the committees. Slots were to be auctioned by means of a first rejected bid, sealed-bid auction with an aftermarket. Markets and auctions had received some attention in the trade literature. Almost uniformly authors predicted disastrous consequences that might follow should markets be used to allocate the slots. This literature provided an excellent background for the experiments. The questions posed were (1) how do the committees perform relative to the proposed market process and (2) do any of the disastrous predictions found in the trade literature actually occur if auctions are used?

Specifically the market experiments were conducted to demonstrate (1) that "rampant speculation" did not occur; (2) that the values placed on slots by the large carriers do not dictate slot prices because price is determined by the marginal buyer; and (3) that the problems poorly solved by the committee process would be solved more efficiently by a particular type of market process. All three points were clearly demonstrated by the experiments.

The experiments served the purpose well. The report was adopted and promoted by the CAB. It was the subject of many public hearings and a notice of proposed rule making. The recommendations to replace the committee with a market process were very controversial but the experiments were never criticized. In fact, the economic analysis was accepted in the sense that the critics chose to question the CAB's authority to implement such a scheme, and the tools used by critics to

back up such claims were congressional and international political pressure.

The exact role of the experiments in this process is difficult to ascertain. The report was supplemented by detailed transcripts of three of the committee meetings. Quotations from these meetings were used to buttress the results of the theory and experiments. No doubt these were read carefully and from these texts alone the logic of the theory could be detected. My guess is that the experiments prevented certain types of claims from surfacing in policy debates and also gave confidence to governmental staff who needed to buttress their views with data. Something other than pure theory was necessary.

Staff at the FAA were opposed to market policies from the beginning. They were certainly not convinced by the experiments and funded experiments from another group that they hoped would disconfirm our findings. The follow-up experiments conducted by another group were so complicated that no conclusions could be drawn from them. They attempted to use members of the industry who applied their own valuations brought from the field. In the sense of modern experimental economics the study had no controls at all.

The recommendations of the Polinomics report were not implemented in 1979. However, attempts to implement variations of the recommendations have existed almost every year since then. The analysis of the committees has been almost completely accepted by all now, including the airlines and even the FAA. The committees themselves began to deadlock by 1982. By 1984, the airlines had

recommended that the committees be replaced with a modification of the Polinomics recommendation that I proposed as an alternative (Aviation Daily, 1983). This alternative grandfathered airlines at current slot holdings, created a market for slots, and provided that new capacity be auctioned. The FAA, which had assumed a leadership role in opposing all forms of market allocation, aggressively opposed this proposal in favor of its own plan to administratively allocate slots. In the fall of 1985 the DOT issued a notice of proposed rule-making to implement the proposal. That rule became law April 1, 1986.

III. SHIFTS IN THE BURDEN OF PROOF

Experimental data can influence the burden of proof in an ongoing policy debate. In this context the use of experiments seems to be as much tactical as fact gathering about the nature of the actual situation. The objective of the experiment is to establish the need for the other side of the argument to prove or disprove something before a policy discussion can proceed in their favor. Specifically in the cases below the other side had made claims about a complex situation based on a very general economic model. The experiments were designed to check the accuracy of the model. If the model advocated as being very general failed to be reliable in the simple case of the experimental markets, then the burden of proof would presumably rest on the advocate to explain why it did not work. If the model is so general that it can be applied to some very complex situations then why can't it be reliable in the simple situations? The burden is on the

model's advocate to explain why the experiments were "special" or "different" from the complex case. Failing that the generality of the model is in question and the application to the complex case is in doubt. Thus the experiments do not address the field situation directly. Rather, the experiments address theory that one side or the other has used to analyze the field situation.

The shift of burden of proof arguments has been explicitly used in two studies. Some of the demonstration arguments used by Grether, Isaac, and Plott in the Polinomix airport slot study could be counted as a third instance of shift of burden of proof strategies.

An interesting feature of all three attempts to use the shift of burden of proof strategy is that the experiments were designed to mirror the industry as closely as was possible. Relative sizes of buyers and sellers, demand elasticities, numbers of participants, etc., were all similar to those of the target industries. This was done to prevent the application of theories that attempt to show that the behavior of the laboratory industry will differ from that of the industry. Each imaginable difference between the experimental setting and the field is the starting point for a potential theory. An infinity of such theories necessarily exist.

The logic is as follows. Individual A has used a general theory T to infer something about the industry and its performance. Individual B has noted that under experimental circumstances E, T is not reliable. Thus T is not reliable in general because it is not reliable in E. B then asks A to explain why T can be applied to the

industry. That is B places a burden of proof on A to show why T is applicable to the industry but not under condition E. Now B does not want A to have readily available some specified property of E that might explain why E is an (uninteresting) exception to the general reliability of T. Each difference with the industry serves as a basis of a potential theory. For example, the laboratory results are due to the special concentration ratios used in the experiment that differ from the industry's. The laboratory did not use people experienced in the industry. The laboratory demand elasticity differed from the industry, etc. Such theories can be checked through additional experiments but time and money is involved. The best strategy is to eliminate as many potential theories as is possible so the burden of proof is not easily shifted back to its original position.

Inland Waterways Barge Traffic

Railroad companies were lobbying a high-level administrator to require barges to post rates with the Interstate Commerce Commission. The railroads argued that the public information feature of posted rates would make the industry more competitive, allow the railroads to compete better, and aid the small barge owners who were allegedly secretly being undersold by the large barge companies. The administrator was skeptical of the arguments but had no basis to express his skepticism. He commissioned an experimental study (Hong and Plott, 1982) that became the first attempt to apply the recently developed experimental methods to an actual policy problem.

The inland waterways barge industry is complex. Traffic exists on both coasts and in the Great Lakes region. Much of the industry exists on the Mississippi River and its tributaries. A great variety of products are hauled with boats, and firms are specialized accordingly. The first task of the study was to isolate a portion of the industry that had minimal complications. That portion would serve as a model for the creation of a laboratory industry.

A small portion of the Mississippi River was chosen. Only dry bulk cargo was used for the basic model. Dry bulk was the major product for this stretch of the river. Parameters from governmental studies, judgments by industry people, and judgments by the researchers were used to characterize that portion of the industry during the specific year for which the best data were available. A laboratory experiment was conducted that represented a dramatically scaled down version of the snapshot of the industry.¹

The industry had about fifteen grain shippers. All are about equal size. Somewhere between 25 and 35 larger companies operated. The size of a company could be measured by the number of boats it operated. Rough estimates of the volume of cargo shipped were available and served as the basis for demand and supply estimates.

The difficult part was determining an appropriate scale. Since a tow down the river took about a month, the number of boats translated in the number of tows a company could undertake. A unit in the experiment became one-half tow and a period in the experiment represented two weeks. A participant with a capacity to deliver five

units in the experiment represented a company with over five boats. The costs associated with units for sale in the experiment corresponded to engineering cost estimates for barges. An upward sloping supply curve reflected a high marginal cost of upgrading marginal equipment and entry into the grain hauling business by firms ordinarily hauling something else. The overall elasticity of supply was structured to reflect the best guesses about the industry. Elasticity of demand was similarly chosen. When the scale parameters were applied to the experimental parameters the known industry aggregates were recovered.

A price posting institution was used for two experimental sessions. The price posting institution previously studied in laboratory work has many features similar to the rate posting procedures used by the ICC. A second two experiments were conducted with a telephone market. Buyers and sellers were all located in different rooms with telephones, and orders were negotiated and placed by phone. The telephone market was arguably analogous to the existing form of organization. Parameters were identical across all four markets.

In the price posting markets as compared to the telephone markets, prices were higher, efficiencies were lower, and the small sellers made less profits. The results were exactly opposite the predictions made by the railroads. Furthermore the experiment provided estimates of the amount of business that would be shifted to the railroads should posted prices be required. The study was sufficient to make the administrator wary of the claims of the railroads. In private

conversations they were challenged to explain the results. The administrator claimed that with evolving scientific evidence against their case he was not in a position to help them. The lobbying effort was diminished and the policy as advocated by the railroads was never pursued. The fact that a presumption existed against their case was sufficient to deter further lobbying efforts.

The administrator's use of this study was not widely supported within DOT. After the administrator left, the study was to have been published but a staff economist who feared it would earn them Proxmire's Golden Fleece Award blocked it. He thought the whole idea of doing laboratory experiments in economics was silly. At the time (1977) one could not use the authority of a large number of published papers to contest his belief. The study itself was then rejected by Sam Peltzman, editor of the Journal of Political Economy, who suggested that a paper with the details of the barge application removed and replaced by survey-oriented material might be acceptable. This suggestion came after he had mistakenly claimed that the results were due to an artifact of the experimental procedures. It was then accepted and published by The Bell Journal of Economics.

The Ethyl Case

The FTC brought action against the major manufacturers of tetraethyl and tetramethyl lead, the lead-based gasoline additives that increase octane levels. The basis of the lawsuit was four practices widely used in the industry (1) delivered pricing, (2) most favored

nations clauses, (3) automatic matching of competitor's prices and (4) advance notification of price increases.

The government's claim was that these four practices when taken as a group served to increase prices in an "anticompetitive" fashion. The logic is as follows. Delivered pricing removes potential for under-the-table price discounts in terms of free services. Delivery is the only major service provided customers by producers. Most favored nations assure customers that no other customer is buying at a lower price. This policy eliminates a source of small price concessions in response to individual customer pressure. It is like the posted-price of a rate bureau. What you see is what you get--no negotiations. Meet or release clauses tie prices to that of a competitor. It is the precommitment to a trigger-price policy. A company will not win customers away from a competitor by price concessions because as soon as the lower price becomes known the competitor lowers price automatically. Advance notification requires a thirty-day notice in advance of price increases. By giving a forty-day advance notice competitors were aware of a ten-day window to bring prices up to the new level. If they failed to act in ten days the company that made the notice would necessarily retract it because of the practice of matching prices. Thus by giving a forty-day notice a company gives competitors a choice between all competitors having the higher price and no price increases by anyone.

The four defendants' reply to the charge that the practices had an anticompetitive effect was that they were an oligopoly. The practices

had no effect on industry performance because there was no room for an effect. The industry enjoyed high (but not illegal) prices fostered by industrial concentration. According to the defendants' claim, any high prices and/or profits were accounted for entirely by industrial structure and therefore were unrelated to the four practices.

The experiments were designed for the government to serve as a basis of rebuttal testimony. Was the industry's claim true? Is it a fact that the practices necessarily have no influence when the industrial organization is that of the industry? If the answer is no, then the defendants cannot claim that the high profits and prices are necessarily unrelated to the government's case. A major tenet of their argument would be damaged.

Laboratory experiments were designed to match the numbers of suppliers and demanders, concentration ratios, demand elasticities, excess capacity, etc., that are known properties of the industry. Special care was taken to anticipate questions that one could imagine during cross examination. Would the attorney attempt to make the experimental argument look silly? Subjects were not undergraduates. For the most part they were employed adults preferably with an engineering background and/or some connection to the oil industry. Subjects participated in more than one experiment. Several different variations of the practices were studied. Consistency in design with previous experiments was sought so the weight of the authority of different types of experiments conducted by others could be used. Wherever possible the consistency of the results with "the tradition of

experimental research" was established. Testimony of respondents' experts was studied carefully so different forms of the rebutted theory would be recognizable within the experimental design. Many replications were done. Some experiments were blind in the sense that the experimenter conducting the experiment did not know the parameters. A double blind experiment was considered but the experiment was so complicated that it could not be conducted by novices.

The experiments were conducted. The results were decisive in showing that the practices could have a substantial impact. The results were circulated to the respondents but the government decided (correctly) that the case was won without the need of the rebuttal testimony provided by the experiments. Since the experiments were novel and having never before been introduced in a court, a decision was made not to use the results from the stand. Presumably no problem would have existed getting the experiments admitted as evidence (Kirkwood, 1981). The government won the case on the first round but the defendants won on appeal.

After the trial a seminar on the experimental results was conducted at the FTC. Discussions with the defense lawyers following the seminar revealed some of their thinking. They had studied the several variations of the practices reported in the paper. One of the variations in which the practices were not well enforced resulted in prices slightly above the competitive equilibrium. Counsel for the defense asked if that particular treatment wasn't the best approximation of the actual practices. The questions made sense

because the practices as found in the industry were not perfect. Evidently his first line of defense would have been to attempt to use the experimental data as evidence in support of his position. In retrospect our experimental case could have been a better tool for the prosecution if we had built the detailed exceptions to perfect enforcement into the design. My guess is that given the nature of the imperfect practices found in the field, the experimental results would not be changed at all.

IV. DIRECT EXTRAPOLATION: AIR FREIGHT POSTING

Policy choices require decisions, and the weight of the evidence is a subjective issue that rests with the decision maker. Studies designed to answer one set of questions might provide a decision maker with sufficient insight to act on a completely different set of issues. Such was the case with the CAB air freight rate decision.

Prior to 1980 air freight forwarders were required to post their rates with the CAB. Having studied the influence of posted prices in the early Plott and Smith study (1978) and the barge study (Hong and Plott, 1982), a reasonable presumption existed that posted prices reduce market efficiency. Based on existing laboratory results and within the spirit of deregulation, the CAB issued a notice of proposed rule making to eliminate air freight rate posting. Seeing no claims that the presumption was incorrect, the CAB acted and eliminated the policy of rate posting.

V. POTENTIAL DESIGN: PREPOLICY RESEARCH

Two experimental studies (Plott and Wilde, 1982; and Lynch, Miller, Plott, and Porter, 1984) were developed as tools to study policy options. Both were financed by the Federal Trade Commission, which has an interest in consumer protection. The staff of the commission is exposed to many competing policies aimed at correcting alleged market failures.

The problem faced by the staff is that neither the "market failure" nor the influence of a "proposed remedy" can be clearly observed with field data. The experimental strategy was to create markets that would reliably "fail." Such markets can then be used to study the properties of proposed policy remedies as implemented in those markets. The experiments conducted were not focused on any particular industry or potential decision. Rather, the experiments had characteristics of a variety of markets and alleged market problems that were the concern of the commission.

No policy recommendation resulted from either study. The researchers learned much about the limitations of broad statements about behavior that have accompanied past policy decisions. The experiments also provided many insights about the nature of models that are being applied to consumer protection problems. The hope is that these background experiments will be the basis for additional experimentation and policy analysis.

VI. DESIGN

Sometimes new economic problems surface that require completely new types of organization and decision processes. In such cases experiments can provide some experiences upon which to base judgments about the nature of appropriate organizations and policy. Such situations are properly called problems of organizational design. Three instances are outlined here.

Slot Exchanges

Following the slot allocation process study by Grether, Isaac, and Plott, the air controllers' strike occurred and the committees at the various airports began to deadlock. The number of constrained airports expanded from four to twenty-two. A decision was made to create a "slot exchange." Air carriers were given temporary grandfather rights to their historic level of slots. The plan was to allow carriers to meet and exchange slots within and across airports on a one for one basis.

How should such a process be organized? The problem was non-trivial because the size of the exchange was staggering. The number of commodities measured in the thousands and the number of agents measured in the hundreds. The politics of the situation dictated that no buying or selling was to be allowed so no numeraire existed. The logistics problem was enormous.

My role in this process was as a consultant for a carrier that wished to trade away from one airport to get to another. As a

participant in all organizational meetings, I was involved in the design of the process.

The ultimate process was constructed on the basis of experiences with experiments with one-sided oral auctions. The only difference was that bids were to be tendered in writing rather than orally. Each carrier listed slots that they wished to acquire together with slots that they would take in exchange. The form of these proposed trades was any slot in column A will be exchanged for any slot in column B. These lists of bids were collected and circulated to all carriers. With the list of proposed trades, carriers searched for chains of trades that involved their own proposals. At this stage of searching for trades, carriers were free to add new proposed exchanges that had the effect of accepting a proposal or completing a chain of trades.

The process was not well understood. However, pilot experiments had been conducted at Caltech. The carrier that had employed me had practiced and had a strategy for dealing effectively with the process. Since our "team" had well-defined ideas about how the logistics of the process might work, we had little difficulty in convincing the group of all carriers to adopt the process. The process and improved variations were used many times, including a brief period when monetary transactions could take place and a period in which "many for one" trades were permitted. As an organizer and participant the whole affair was similar to a large experiment. Given the constraints it worked rather smoothly.

Westchester County Airport

The county of Westchester in New York decided to auction access to its airport terminal. The terminal is small and safety codes limited passenger usage to a maximum of forty enplane and forty deplane passengers in any fifteen-minute period. In addition, a maximum of four aircraft could use the parking pads at any time and, at most, two of these could be large aircraft.

When the county had taken action to limit the use of the terminal to the stated capacity limitations, it became involved in a lawsuit. The judge ordered the county to devise a mechanism for allocating the available capacity that was consistent with the airline deregulation act. The county chose to develop an auction system that was to be used in the event that a settlement could not be attained.

The auction was designed by Glen George and me. It was important to avoid many potential criticisms of auction processes that contestants might assert and it was necessary to conform to the realities of politics. The carriers might more readily accept a process that tended to allocate rents to carriers so a one-price auction was used. The continuum of time was unwieldy so the day was divided into fifteen-minute segments. Capacity was divided into five passenger enplane blocks and five passenger deplane blocks. So, two separate markets existed every fifteen-minute period of the day in which eight passenger blocks were sold in each. Carriers desired to tie purchases together so provisions for special constraints that tied enplane purchases to deplane purchases were designed. Carriers were

also allowed to submit multiple bids tied together with a constraint that canceled all other tied bids if one of the set was accepted.²

The number of markets together with the possibility of constraints resulted in a very large and potentially complex auction. Experiments were conducted using demand conditions similar to those believed to exist at Westchester County Airport. Some very practical questions formed the purpose of the experiments. (1) Were the instructions clear about how to tender bids and use the constraints? What types of confusions were we likely to encounter? (2) Did unusual strategies exist that might undermine any efficiency properties of the auction process? (3) Were we likely to encounter a computational problem in determining the winning bids? We could imagine problems that would exceed our computer capacity. The solution to the auction involves a large integer programming problem, the dimensions of which are very sensitive to the number of constraints. The use of bids and constraints is not governed by the logic of the problem so we had no a priori way of determining the size of the computing program without actually trying the auction.

The experiments were invaluable. Many problems were uncovered at every stage. The whole process was redesigned several times after bugs of one form or another surfaced during experiments. Experiments are still underway to improve the process

The process was not used at Westchester County. The respondents settled by adopting the process I proposed for Washington National (Aviation Daily, 1983) discussed above. The original laboratory

experiments and the field experiments with the slot exchanges have provided convincing evidence that markets in slots will "work." All of this evidence made carriers happy with a market for slots although the FAA remained adamantly opposed. Now the New York Port is considering an auction process for the three major airports in New York City. The research and refinements on the Westchester County problem are relevant to the Port's problem.

Space Station

Several years from now NASA will place a manned space station in orbit. The station will be used as a research laboratory, as a manufacturing facility, and for a variety of other purposes. The users will be the U. S. government, foreign governments, and private corporations. The Reagan administration wants the capacity to be allocated by some sort of pricing system. A team of economists at the Jet Propulsion Laboratory have been given the task of designing a pricing mechanism.

The task is complicated by the existence of nonconvexities, externalities, large common costs, much uncertainty, etc. In addition, NASA cannot operate for a profit or even take money for that matter, so profit centers and related decentralized schemes do not seem to be feasible. Matters are further complicated by the fact that the space station design is in a continuous state of evolution and the design of the system itself should be affected by the pricing scheme.

The proposed role of experiments in this project is much different

from that of previous research. Some testing of institutional influences are under way but the central role is to be a little different.

The experiments are to be used as a heuristic--a tool to organize thoughts and questions as opposed to a tool for answering questions. Most experimenters have noticed that the process of designing experiments makes the researchers aware of complications and interdependencies that would have otherwise escaped notice. The space station project is intended to capitalize on this feature of the method.

The space station is just finding its way to the drawing boards. The variables are not even known, much less the costs. The experimental plan is to conduct simulations of pricing policies under experimental conditions that reflect much of the physical, institutional, and motivational aspects of the space station. Ultimate subjects will be the NASA personnel who are building the station. The purpose will be to instruct the personnel on the nature of competing policy options by providing them with some experiences with their operation. Hopefully, such exercises will generate insights about the features of the station, its cost, engineering structure, service capacities, etc., to make simulation of policy options useful.

CLOSING REMARKS

The theme of this paper is that "parallelism" involves many different dimensions. There may be many different types of parallelism. The use of laboratory methods in policy contexts seems to be an art involving skillful and very subjective analogies. The laboratory results are sources of experiences in situations that hopefully will be useful to those responsible for decisions. I tend to view the role of experiments in these contexts as an activity more akin to practice than to some sort of scientific pursuit of truth. It is a source of experience similar to the experience one gets as one practices the piano before a concert or as a team practices before a game.

Having compared this type of research with practice, it may come as no surprise that I am particularly uncomfortable with the concept of external validity so popular in psychology. The concept of external validity may be appropriate when the problem is a statistical problem of inference about a population from a sample. One can speak about the "validity" of inferences one makes about a population based upon the properties of a sample. The analogy with statistical inference is inappropriate when applied to the policy problems reviewed here. In these applications theory plays a special role. The policy decision will constitute a unique event in history. Therefore there is in principle no way to "validate" theories about what might happen. Simple judgment cannot be avoided. The experiments simply shape the thought processes and arguments that form that judgment.

My approach to applied work has been to forget the concept of external validity and not to take the concept of parallelism too literally. Instead the approach has been pragmatic in the sense that the use of experiments in each project was justified by whatever arguments seemed appropriate at the time. The purpose of this note was to survey some of what was done to see to what extent some order or method actually existed after all.

FOOTNOTES

- * The financial support of the National Science Foundation and the Caltech Program for Enterprise and Public Policy is gratefully acknowledged. I also wish to thank Alvin Roth and Howard Kunreuther who provided comments on an early draft.
1. My own thoughts about how one might learn something relevant by using experimental methods were influenced by Caltech environmental engineers. The engineers were attempting to learn about the flow of effluents in the ocean near Los Angeles by studying currents in a large pool constructed in the basement of a building on the campus. Of course their pool looked nothing like the Pacific. Yet the pool taught them something about their models and it was the models that taught them something about the Pacific.
2. Many of the ideas were motivated by Rassanti, Smith and Bulfin (1982). The combinatorial auction was not feasible because of practical and technical problems.

REFERENCES

- Aviation Daily, Washington, D.C. July 25, 1983, back of pages 124 and 127.
- Banks, Jeffrey S.; Plott, Charles R.; and Porter, David P. "An Experimental Analysis of Public Goods Provision Mechanisms with and without Unanimity." Social Science Working Paper no. 595. Pasadena: California Institute of Technology, January 1986.
- Grether, David M.; Isaac, R. Mark; and Plott, Charles R. "Alternative Methods of Allocating Airport Slots: Performance and Evaluation." Prepared for the Civil Aeronautics Board. Polinomics Research Laboratories, Inc., Pasadena, California, 1979.
- Grether, David M.; Isaac, R. Mark; and Plott, Charles R. "The Allocation of Landing Rights by Unanimity among Competitors." American Economic Review 71 (May 1981):166-71.
- Grether, David M. and Plott, Charles R. "The Effects of Market Practices in Oligopolistic Markets: An Experimental Examination of the Ethyl Case." Economic Inquiry 22 (October 1984):479-507.

- Hong, James and Plott, Charles R. "Rate Filing Policies for Inland Water Transportation: An Experimental Approach." Bell Journal of Economics 13 (Spring 1982):1-19.
- Kirkwood, John B. "Antitrust Implications of the Recent Experimental Literature on Collusion." In Strategy, Predation, and Antitrust Analysis, edited by Steven C. Salop, pp. 608-621. Federal Trade Commission, September 1981.
- Levine, Michael E. and Plott, Charles R. "Agenda Influence and Its Implications." Virginia Law Review (May 1977):561-604.
- Lynch, Michael; Miller, Ross M.; Plott, Charles R.; and Porter, Russell. "Product Quality, Informational Efficiency and Regulations in Experimental Markets." Social Science Working Paper no. 518. Pasadena: California Institute of Technology, March 1984.
- Plott, Charles R. and Smith, Vernon L. "An Experimental Examination of Two Exchange Institutions." Review of Economic Studies 45 (February 1978):133-153.

- Plott, Charles R. and Wilde, Louis L. "Professional Diagnosis Versus Self Diagnosis: An Experimental Examination of Some Special Features of Markets with Uncertainty." In Research in Experimental Economics, vol. 2, edited by Vernon Smith. Greenwich, Conn.: JAI, 1982.
- Rassanti, S. J.; Smith, V. L.; and Bulfin, R. L. "A Combinatorial Auction Mechanism for Airport Time Slot Allocation." Bell Journal of Economics 13 (Autumn 1982):402-417.
- Samuelson, Paul A. and Nordhaus, William D. Economics. 12th ed. New York: McGraw-Hill, 1983.