Experimental Economics, 5:91–110 (2002)

Method in Experiment: Rhetoric and Reality

VERNON L. SMITH
Professor of Economics and Law, Interdisciplinary Center for Economic Science, George Mason University, Truland Building 400-J, 3330 Washington blvd, Arlington Virginia

Abstract

The methodological ideal of experimentalists, E, is easily stated: derive a testable hypothesis, H, from a well-specified theory, T; implement experiments with a design; implicitly in the latter are auxiliary hypotheses, A, that surface in the review/discussion of completed research reports (payoffs are 'adequate,' Ss are 'relevant,' instructions, context are 'clear,' etc.). We want to be able to conclude, if statistical test outcomes support not-H, that T is 'falsified.' But this is not what we do; rather we ask if there is a flaw in the test, i.e. not-A is supported, and we do more experiments. This is good practice—much better than the statistical rhetoric of falsificationism. Undesigned social processes allow E to accumulate technical and instrumental knowledge that drive the reduction of experimental error and constitute a more coherent methodology than falsificationism.

Keywords: experimental economics, science, methodology

JEL Classification: B4, C90

1. Introduction

What does it mean to test a theory, or a hypothesis derived from a theory? Scientists everywhere believe that the unique feature of science is that theories, if they are to be acceptable, require rigorous support from facts based on replicable observations. But the closer one examines this belief the more elusive becomes the task of giving it precise meaning and content. This is my first message, which I claim is interesting and worth examining. I believe, however, that the significance of this message for the operating scientist is easily exaggerated: what scientists do is better than what they say about what they do. Hence, the second message is that the practitioners daily overcome this imprecision with their growing repertoire of practical knowledge. Their natural instincts and lively professional interactions lead them perpetually to design new experiments that examine the right questions, but such methods are not part of their rhetoric of science. Finally, the Machine Builders, who invent and fabricate each new generation of tools and instruments, finesse it, but with each new device the same methodological issues emerge at a deeper observational level.

2. Can we derive theories directly from observations?

Prominent scientists through history have believed that the answer to this question is "yes," and that this was their modus operandi. Thus, quoting from two who are among the most influential scientists in history:

> I frame no hypotheses; for whatever is not deduced from the phenomena . . . have no place in experimental philosophy . . . (in which) . . . particular propositions are inferred from the phenomena . . . (Isaac Newton, Principia, 1687; quoted in Segrè, 1984, p. 66);

> Nobody who has really gone into the matter will deny that in practice the world of phenomena uniquely determines the theoretical system, in spite of the fact that there is no theoretical bridge between phenomena and their theoretical principles (Einstein, 1934, pp. 22–23).

But these statements are without validity. "One can today easily demonstrate that there can be no valid derivation of a law of nature from any finite number of facts" (Lakatos, Vol. 1, 1978, p. 2).

Newton passionately believed that he was not just proffering speculative hypotheses, but that his laws were derived directly, by logic, from Kepler's "discovery" that the planets moved in ellipses. But, of course, Newton only showed that the path was an ellipse if there were $n = 2$ planets. Kepler was wrong in thinking that the planets followed elliptical paths, and to this day there is no exact solution for the n (>2)-body problem; in fact the paths can be chaotic. Thus, when he published the Principia, Newton's model was born falsified, for it could not account for the orbit of our nearest and most easily observable neighbor, the moon, whose orbit is strongly influenced by both the sun and the earth.

Newton's sense of his scientific procedure is commonplace: one studies an empirical regularity (e.g. the "trade-off" between the rate of inflation and the unemployment rate), and proceeds to articulate a model from which a functional form can be derived that yields the regularity. This methodology survives in the thinking of some under the name of "phenomenology." Economists and psychologists often speak of the need for more data, particularly direct observations of individual behavior as a means of deriving better, i.e. a "truer," theory of behavior. In the above confusing quotation, Einstein seems to agree with Newton. At other times he appears to articulate the more qualified view that theories make predictions, which are then to be tested by observations (see his insightful comment below on Kaufmann's test of special relativity theory), while on other occasions his view is that reported facts are irrelevant compared to theories based on logically complete meta theoretical principles, coherent across a broad spectrum of fundamentals (see Northrup, 1969, pp. 387–408). Thus, upon receiving the telegraphed news that Eddington's 1919 eclipse experiments had "confirmed" the general theory, he showed it to a doctoral student who was jubilant, but Einstein commented, unmoved: "I knew all the time that the theory was correct." But what if it had been refuted? "In that case I'd have to feel sorry for God, because the theory is correct" (Fölsing, 1997, p. 439). Actually, two of Eddington's three eclipse sites supported Einstein and one, Newton, and, as usual, a scientific controversy occurred (Mayo, 1996, pp. 278–293).

The first theme I want to develop, particularly for experimental economics, is not new; it is captured by the following quotation from a lowbrow source, the mythical character Phaedrus in Zen and the Art of Motorcycle Maintenance, ". . . the number of rational hypotheses that can explain any given phenomenon is infinite" (Pirsig, 1981, p. 100).

The wellspring of testable hypotheses in economic (game) theory are to be found in the marginal conditions defining equilibrium points or strategy functions that constitute a theoretical equilibrium. In games against nature the subject-agent is assumed to choose among alternatives in the feasible set that which maximizes his/her outcome (reward, utility or payoff). Strategic games are solved by the device of reducing them to games against nature, as in a noncooperative (Cournot-Nash) equilibrium (pure or mixed). The (symmetric) equilibrium strategy when used by all but agent i reduces i's problem to an own maximizing choice of that strategy. Hence, in economics, all testable hypotheses come from the marginal conditions (or their discrete opportunity cost equivalent) for maximization that define equilibrium. These conditions are implied by the theory from which they are derived, but given experimental observations consistent with (supporting) these conditions there is no way to reverse the steps used to derive the conditions, and deduce the theory from subject observations. Behavioral point observations conforming to an equilibrium theory cannot be used to deduce or infer either the equations defining the equilibrium, or the logic and assumptions of the theory used to derive the equilibrium conditions. Hence, the importance of guarding against the belief that the theory has got to be true, given corroborating evidence.

Suppose, however, that the theory is used to derive noncooperative best reply functions for each agent that maps one or more characteristics of each individual into that person's equilibrium decision. Suppose next that we perform many repetitions of an experiment varying some controllable characteristic of the individual's environment, such as their assigned values for an auctioned item, and obtain an observed response for each value of the characteristic. This repetition, of course, must be assumed always to support equilibrium outcomes. Finally, suppose we use these data to estimate response functions obtained from the original maximization theory. First order conditions defining an optimum can always be treated formally as differential equations in the original criterion function. Can we solve these equations and "deduce" the original theory?

Here is an example from first price auction theory (Smith et al., 1991). Each of N individuals in a repeated auction game is assigned values $v_i(t)$ $(i = 1, \ldots, N; t = 1, 2, \ldots, T)$ from a rectangular distribution, and bids $b_i(t)$ on each trial. Each i is assumed to

$$\max_{0 \leq b_i \leq v_i} (v_i - b_i)^{r_i} G_i(b_i), \tag{1}$$

where $r_i (0 < r_i \leq 1)$ is i's measure of constant relative risk aversion, and $G_i(b_i)$ is i's probability belief that a bid of b_i will win. This leads to the first order condition, which can be reduced to

$$(v_i - b_i)G_i'(b_i) - r_i G_i(b_i) = 0. \tag{2}$$

If all i have common rational probability expectations (Nash, 1950)

$$G_i(b_i) = G(b_i). \tag{3}$$

This leads to a closed form equilibrium bid function (See Cox, Smith, and Walker; reprinted in Smith, Chapter 29, 1991).

$$\begin{aligned} &b_i = (N-1)v_i/(N-1+r_i), \quad b_i \leqq \bar{b} = 1 - G(\bar{b})/G'(b), \quad \text{and} \\ &G(b_i)/G'(b_i) = b_i/(N-1), \quad \text{for all } i. \end{aligned} \tag{4}$$

The data from experimental auctions strongly support linear subject bid rules of the form

$$b_i = \alpha_i v_i, \tag{5}$$

that can be estimated by linear regression of b_i on v_i using the T observations on (b_i, v_i), for given N, with $\alpha_i = (N-1)/(N-1+r_i)$. Can we reverse the above steps, integrating (2) and (3) to get (1). The answer is no. Equation (2) can be derived from maximizing either (1), or the criterion

$$(v_i - b_i)G(b_i)^{1/r_i}, \tag{1$'$}$$

in which subjective probabilities, rather than profit, are discounted. In (1′) we have $G_i(b_i) = G(b_i)^{1/r_i}$ instead of (3), which is not ruled out by the data.

In fact instead of (1) or (1′) we could have maximized

$$(v_i - b_i)^{\beta_i} G(b_i)^{\beta_i/r_i}, \quad \text{where } r_i \leq \beta_i \leq 1, \tag{1$''$}$$

This implies (as asserted by Phaedrus) an infinite mixture of subjective utility and subjective probability models of bidding that can support the same set of observations on bidding.

Thus, in general, we cannot backward induct from empirical equilibrium conditions, even when we have a large number of experimental observations, to arrive at the original model. The purpose of theory is precisely one of imposing much more structure on the problem than can be inferred from the data. This is because the assumptions used to deduce the theory contain more information, such as (3), than the data. The next time you report experimental data supporting a hypothesis, someone may note that the result might be due to "something else". Of course, this is necessarily, and trivially true; there are an infinite number of them.

Proposition 1. *Particular hypotheses derived from any testable theory imply certain observational outcomes*; *the converse is false* (Lakatos, 1978, Vol. 1, pp. 2, 16, passim).

3. Economics: Is it an experimental science?

Those of us who do experiments in economics think that we have helped to transform economics, in part, into an experimental science. I would expect, however, that an opinion poll among all economists would show a substantial majority replying "no" to the above question. Although, surely, most have heard that there is a growing subfield of economists who do experiments, this does not, in their perception, make it an experimental science. Similarly, noneconomist professionals do not perceive economics as one of the experimental sciences (Friedman and Sunder, 1995, p. 1).

All editions of Paul Samuelson's Principles of Economics refer to the inability of economists to perform experiments. This continued for a while after William Nordhaus joined Samuelson as a coauthor. Thus, "Economists ... cannot perform the controlled experiments of chemists and biologists because they cannot easily control other important factors" (Samuelson and Nordhaus, 1985, p. 8). My favorite quotation, supplied by one of the century's foremost Marxian economists, Joan Robinson, is, "Economists cannot make use of controlled experiments to settle their differences" (Robinson, 1979, p. 1319). Of course she was wrong—economists do controlled experiments—but how often have they, or their counterparts in any science, used them to settle their differences? Here she has fallen victim to the popular image of science, which is indeed one in which "objective" facts are the arbiters of truth that in turn "settle" differences. The image is that of two scientists, who, disagreeing on a fundamental principle, go to the lab, do a "crucial experiment," and learn which view is right.[1]

Why this catatonic state of the profession's knowledge of the development of experimental methods in economics during the past half century? I doubt that there are many non-experimentalists in economics that understand or appreciate the essence of our methodology: (1) to motivate behavior in laboratory economic environments whose equilibrium properties are known to the experimental researcher or designer; and (2) to use the experimental observations to test predictive hypotheses derived from one or more formal or informal models of these equilibrium properties. The way economics is commonly researched, taught and practiced implies that it is a à prior science, in which economic problems come to be understood by thinking about them. This generates logically correct, internally consistent, theories. Field data are then used for "testing" between alternative econometric model specifications within basic equilibrium theories that are not subject to challenge. If experiments are used, the purpose is to get better field observations on demand, labor supply, investment and other avenues of agent choice behavior, not to study market performance under alternative institutions of contract.

I want to report two examples indicating how counterintuitive it has been for prominent economists to see this function of laboratory experiments in economics. The first example is contained in a quotation from Hayek whose Nobel citation was for his theoretical conception of the price system as an information system for coordinating agents with dispersed information in a world where no single mind or control center possesses, or can ever have knowledge of, this information (Hayek, 1945). In interpreting competition as a discovery process, rather than a model of equilibrium price determination, he argues (brilliantly, in my view, except for the last sentence),

> . . . wherever the use of competition can be rationally justified, it is on the ground that we do not know in advance the facts that determine the actions of competitors . . . competition is valuable only because, and so far as, its results are unpredictable and on the whole different from those which anyone has, or could have, deliberately aimed at . . . . The necessary consequence of the reason why we use competition is that, in those cases in which it is interesting, the validity of the theory can never be tested empirically. We can test it on conceptual models, and we might conceivably test it in artificially created real situations, where the facts that competition is intended to discover are already known to the observer. But in such cases it is of no practical value, so that to carry out the experiment would hardly be worth the expense (Hayek, 1978/1984, p. 255).

He recognizes an important use that has been made of experiments, testing competitive theory "in artificially created real situations, where the facts which competition is intended to discover are already known to the observer," then proceeds to completely fail to see how such an experiment could be used to test his own proposition that competition is a discovery procedure, under the condition that neither agents as a whole nor any one mind needs to know what each agent knows. Rather, his concern for dramatizing what is arguably the most important socioeconomic idea of the 20th century, seems to have caused him to interpret his suggested hypothetical experiment as "of no practical value," since it would (if successful) merely reveal what the observer already knew! I find it astounding that one of the most profound thinkers in the 20th century, could not see the demonstration potential and testing power of the experiment he suggests.[2]

In Smith (1982; reprinted in Smith, 1991, pp. 221–235), I assembled a number of experiments illustrating what I called the Hayek Hypothesis: strict privacy together with the trading rules of a market institution (the oral double auction in this case) is sufficient to produce efficient competitive market outcomes. The alternative was called the Complete Knowledge Hypothesis: competitive outcomes require perfectly foreseen conditions of supply and demand, a statement attributable to many economists and game theorists. In this empirical comparison the Hayek Hypothesis was strongly supported. This theme had been visited earlier in Smith (1976, reprinted in Smith, 1991, pp. 100–105; 1980, pp. 357–360), wherein eight experiments comparing private information with complete information, showed that complete information was neither necessary nor sufficient for convergence to a competitive equilibrium: complete information interfered with, and slowed, convergence compared with private information. Shubik (1959, pp. 169–171) had noted earlier the confusion inherent in ad hoc claims that perfect knowledge is a requirement of pure (or sometimes perfect) competition.[3] The experimental proposition that private information increases support for noncooperative, including competitive, outcomes applies not only to markets, but also to the two-person extensive form repeated games reported by McCabe et al. (1998). Hence it is clear that without knowledge of the other's payoff, it is not possible for players to identify and consciously coordinate on a cooperative outcome, such conscious coordination being essential in two person interactions, but irrelevant, if not pernicious, in impersonal market exchange. I note in passing that the large number of experiments demonstrating the Hayek Hypothesis in no sense implies that there may not be exceptions.[4] It's the other way around: this large set of experiments demonstrates clearly that there are exceptions everywhere to the Complete Knowledge Hypothesis and these exceptions were not part of a prior design created to provide a counter example. The exceptions are important, not because they constitute a disproof of general principles, analogous to a counterexample to an alleged theorem, but because they enable the behavioral limits to such principles to be identified.

My second example derives from a personal conversation with another Nobel Laureate in economics, a prominent theorist. I was describing the experimental public goods research I had done comparing the efficacy of various public good mechanisms, Lindahl; Groves-Ledyard; etc. (See the public goods papers reprinted in Smith, 1991.) He was puzzled as to how I had achieved control over the efficient allocation as a benchmark in these comparisons. I explained what I thought was commonly understood: each subject receives a payoff function (table) defined jointly over variable units of a public and a private good. This allows the experimenter to solve for the social optimum and judge the comparative performance of alternative public good incentive mechanisms. Incredibly, he objected that if, as the experimenter, I have sufficient information to know what is the socially optimal allocation, I don't need a mechanism. I can just impose the optimal allocation! I felt like an anthropologist on Mars, unable to convey to one of the best and brightest, that the whole idea of laboratory experiments was to evaluate mechanisms in an environment where the Pareto Optimal outcome was known by the experimental designer but not the agents so that performance comparisons could be made; that in the field such knowledge was never possible, and we had no criteria, other than internal theoretical properties such as incentive compatibility, to make a judgement. His deficit was driven by a common mind-set: if you have incentive compatibility, then there is nothing to test; if you don't, then the mechanism is not worth testing.

The issue of whether economics is an experimental science is moot among experimental economists who are, and should be, too busy having fun doing their work to reflect on the methodological implications of what they do. But when we do, as in comprehensive introductions to the field, what do we say? Two quotations from impeccable sources will serve to introduce the concepts to be developed next. The first emphasizes that an important category of experimental work " . . . includes experiments designed to test the predictions of well articulated formal theories and to observe unpredicted regularities, in a controlled environment that allows these observations to be unambiguously interpreted in relation to the theory" (Kagel and Roth, 1995, p. 22). Experimental economists strongly believe, I think, that this is our most powerful scientific defense of experimental methods: we ground our experimental inquiry in the firm bedrock of economic (game) theory. A second crucial advantage, recognizing that field tests involve hazardous joint tests of multiple hypotheses, is that "Laboratory methods allow a dramatic reduction in the number of auxiliary hypotheses involved in examining a primary hypothesis (Davis and Holt, 1993, p. 16).

Hence, the belief that, in the laboratory, we can test well-articulated theories, and interpret the results unambiguously in terms of the theory, and that we do so with minimal dependence on auxiliary hypotheses.

4. What is the scientist's image of what he does?

The standard experimental paper within and without economics uses the following rhetorical format in outline form: (1) state the theory; (2) implement it in a particular context (with "suitable" subject monetary motivation in economics); (3) summarize the implications in one or more testable hypotheses; (4) describe the experimental design; (5) present the data and results of the hypothesis tests; (6) conclude that the experiments reject or fail to reject the theoretical hypothesis. This format is shown in figure 1. In the case in which we have two or more competing theories and corresponding hypotheses, then one concludes as to which one is supported by the data using some measure of statistical distance between the data and each of the predictive hypotheses, and reporting which distance is statistically the shortest.

Suppes (1969; also Mayo, 1996, Chapter 5) points out that there exists a hierarchy of models behind the process in figure 1. The primary model or theory is contained in (1) and (2) which generates particular topical hypotheses that address primary questions. Experimental models are contained in (3) and (4). These serve to link the primary theory with data. Finally, we have data models, steps (5) and (6) that link operations on raw data (not the raw data itself) to the testing of experimental hypotheses.

This process describes much of the rhetoric of science, and reflects the self-image of scientists as in the quote above from Robinson (1979), but it does not adequately articulate

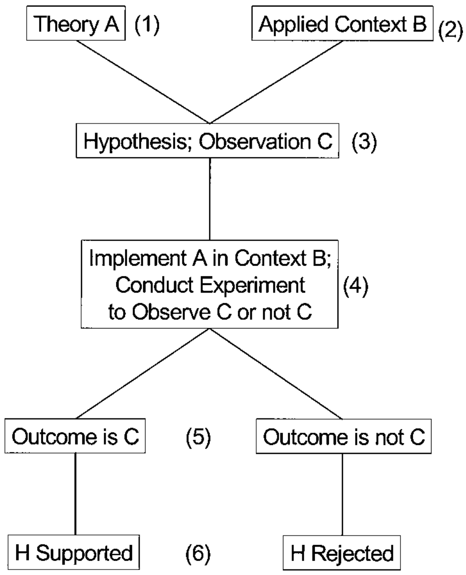

Figure 1. What is the scientist's image of what he does?

what scientists actually do. Furthermore, the rhetoric does not constitute a viable, defensible and coherent methodology. But what we actually do, I believe, is highly defensible and on the whole, constructively impacts what we think we know from experiment. Implicitly, as experimentalists, we understand that every step, (1)–(6), in the above process is subject to judgement, learning from past experiments, our knowledge of protocols and technique, and to error. This is reflected in what we do, if not in what we say about what we do, in the standard scientific paper.

The problem with the above image is well known as the "Duhem-Quine (D-Q) problem": experimental results always present a joint test of the theory (however well articulated, formally) that motivated the test, and all the things you had to do to implement the test.[5] Thus, if theoretical hypothesis H is implemented with context specific auxiliary hypotheses required to make the test operational, $A_1, A_2, \ldots, A_n$; then it is $(H \mid A_1, A_2, \ldots, A_n)$ that implies observation C. If you observe not C, this can be because any of the antecedents $(H; A_1, \ldots, A_n)$ can represent what is falsified. Thus, the interpretation of observations in relation to a theoretical hypothesis is inherently and inescapably ambiguous, contrary to our accustomed thinking and rhetoric.

The reality of what we do, and indeed must do, is implied by the truth, "No theory is or can be killed by an observation. Theories can always be rescued by auxiliary hypotheses" (Lakatos, 1978, p. 34).

4.1. A D-Q example from physics

Here is a historical example from physics: in 1905 Kaufmann (cited in Fölsing, 1997, p. 205), a very accomplished experimentalist, (in 1902 he had showed that the mass of an

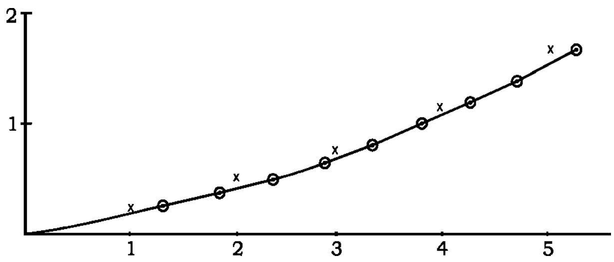

Figure 2. Reconstructed using data. (From A. Beck (translator), The Collected Papers of Albert Einstein, Vol. 2, p. 283. Princeton University Press, Princeton, 1989.)

electron is increased by its velocity!) published a paper "falsifying" Einstein's special theory of relativity the same year in which the latter was published (Einstein, 1905). Subsequently, Einstein (1907) in a review paper reproduced Kaufmann's figure 2, also shown here as figure 2, commenting that

> The little crosses above the (Kaufmann's) curve indicate the curve calculated according to the theory of relativity. In view of the difficulties involved in the experiment one would be inclined to consider the agreement as satisfactory. However, the deviations are systematic and considerably beyond the limits of error of Kaufmann's experiment. That the calculations of Mr. Kaufmann are error-free is shown by the fact that, using another method of calculation, Mr. Planck arrived at results that are in full agreement with those of Mr. Kaufmann. Only after a more diverse body of observations becomes available will it be possible to decide with confidence whether the systematic deviations are due to a not yet recognized source of errors or to the circumstance that the foundations of the theory of relativity do not correspond to the facts (Einstein, 1907).

Kaufmann was testing the hypothesis that $(H \mid A)$ implies C, where H was an implication of special relativity theory, A was the auxiliary hypothesis that his context specific test and enabling experimental apparatus was without a "recognized source of errors," C was the curve indicated by the "little crosses," and not-C was Kaufmann's empirical curve in figure 2.

Einstein's comment says that either of the antecedents $(H \mid A)$ are what is falsified by Kaufmann's results. Others, such as Planck, and Kaufmann himself, however, recognized that the observations might conceivably be in error (Fölsing, 1997, p. 205). In less than a year, Bucherer (see Fölsing, 1997, p. 207) showed that there had been a "problem" with Kaufmann's experiments and obtained results supporting Einstein's theory.

There is an important lesson in this example if we develop it a little more. Suppose Bucherer's experiments had not changed Kaufmann's results enough to change the conclusion. Einstein could still have argued that there may be "a not yet recognized source of errors." If so, the implication is that H is not falsifiable, for the same argument can be made after each new falsifying test! Recall that the deviations were alleged to be "considerably beyond the limits of error of Kaufmann's experiment." We can go still further in explicating the problem of testing H conditional on any A. The key is to note in this example the strong dependence of any test outcome on the state of experimental technique: Bucherer found a way to "improve" Kaufmann's experimental technique so as to rescue Einstein's "prediction." But the predictive content of H (and therefore of the special theory) was inextricably bound up with A. Einstein's theory did not embrace any of the elements of Kaufmann's (or Bucherer's) apparatus. The theory could not be so extended because relativity had nothing to do with the apparatus, and only was concerned with what the apparatus could reveal about the behavior of particles manipulated by the apparatus; yet, as always, those auxiliary elements affect the interpretation of the results.

5. A proposition and an economics example

All experimental economists have encountered a common objection to economics experiments: the payoffs are too small. It is the kind of objection that arises after the experiment is finished, and someone is more skeptical of the test than of the theory being tested. This is one of the principal issues in a target article by Hertwig and Ortmann (2001), with comments by 34 experimental psychologists and economists. This objection sometimes is packaged with an elaboration to the effect that economic (game) theory is about situations that involve large stakes, and you "can't" study these in the laboratory (Actually, of course, you can, but funding is not readily available except for poor countries where modest sums in U.S. dollars are a sizable proportion of the subjects' monthly income).

Suppose, therefore that we have the following:

H (from theory): subjects will choose the equilibrium outcome (e.g. Nash or subgame perfect).

A (auxiliary hypothesis): payoffs are adequate to motivate subjects.

Proposition 2. *Suppose a specific rigorous test rejects* $(H \mid A)$, *and someone* (*say* T), *protests that what must be rejected is A not H. Let E replicate the original experiment with an n-fold increase in payoffs. There are only two outcomes and corresponding interpretations, neither of which is comforting to the above rhetorical image of science*:

1. *The test outcome is negative. Then* T *can imagine a still larger increase in payoffs,* $N > n$, *and argue for rejecting* A *not* H. *But this implies that* H *cannot be falsified.*
2. *After further increases in payoffs, the test outcome is positive for some* $N \geqq n^*$. *Then* H *has no predictive content.* E, *with no guidelines from the theory, discovers an empirical payoff multiple,* n^*, *that is consistent with the theory, but* n^* *is an extra theoretical property of things outside* H *and the theory.*

Proposition 2 holds independent of any of the following considerations:

1. however well articulated, rigorous or formal is the theory; game theory in no sense constitutes an exception;
2. however effective is the experimental design in reducing the number of auxiliary hypotheses—it only takes one to create the problem.
3. the character or nature of the auxiliary hypothesis—A can be anything not contained in the theory.

In experimental economics, reward adequacy is just one of a standard litany of objections sometimes to experiments in general, and sometimes to experiments in particular. Here are three additional categories in this litany:

Subject sophistication; the standard claim is that undergraduates are not sophisticated enough.[6] Subjects are not out there participating in the "real world." (Actually they are when we use subjects from industry). In the "real world" where the stakes are large, such as in the FCC spectrum rights auctions, bidders use game theorists as consultants (Banks et al., 2001).

Subjects need an opportunity to learn; this is a common response from both experimentalists and theorists when you report the results of single play games in which "too many" people cooperate. The usual proposed "fix" is to do repeat single protocols in which distinct pairs play on each trial, and apply a model of learning to play noncooperatively.[7] But there are many unanswered questions implicit in this auxiliary hypothesis: since repeat single protocols require large groups of subjects (20 subjects to produce a 10-trial sequence), have any of these games been run long enough to measure adequately the extent of learning? In single play two-person anonymous trust games data have been reported showing that group size matters; i.e. it makes a difference whether you are running 12 subjects (6 pairs) or 24 subjects (12 pairs) simultaneously in a session (Burnham et al., 2000). In large groups, pairs are less trusting than in small groups—perhaps not too surprising. But in repeat single, larger groups are needed for longer trial sequences. Hence, learning and group size as auxiliary hypotheses lose independence, and we have knotty new problems of complex joint hypothesis testing. Yes, Virginia, these problems also arise in the laboratory. The techniques we fashion for solving such problems are among the sources of our experimental knowledge.

Instructions are adequate (or decisions are robust with respect to instructions, etc.); What does it mean for instructions to be adequate? Clear? If so, clear about what? What does it mean to say that subjects understand the instructions? Usually this is interpreted to mean that they can perform the task, which is judged by how they answer questions about what they are to do. This means that the subject understands when her answers to the questions conform to what the experimentalist wants her to understand, thereby appearing to think like the experimentalist. In two-person interactions, instructions often matter so much that they must be considered a (powerful) treatment.[8] Instructions can be important because they define context, and context matters because memory is auto biographical. Ultimatum and dictator game experiments yield statistically and economically significant differences in results due to differing instructions and protocols.[9]

6. In view of Proposition 2, what are experimentalists and theorists to do?

Consider first the example in which we have a possible failure of A: rewards are adequate to motivate subjects. Experimentalists should do what comes naturally, namely, do new experiments that increase rewards and/or lower decision cost by simplifying experiment procedures. The literature is full of examples (for surveys and interpretations see Smith and Walker, 1993a, 1993b; Camerer and Hogarth, 1999; also a new comparison of several payoff levels by Holt and Laury, 2001). Anyone who doubts that payoffs can and do matter has not looked at the evidence. What is not predictable by any theory is what situations will be sensitive at what payoff levels and what situations will not be sensitive at the levels commonly used.

Theorists should ask if the theory is extendable to include A, so that the effect of payoffs is explicitly modeled. It is something of a minor scandal when economists, whose theoretical models predict the optimal outcome independently of payoff levels, and however gently rounded is the payoff in the neighborhood of the optimum, object to an experiment because the payoffs are inadequate. What is adequate? Why is inadequacy of payoffs a complaint rather than a spur to modeling payoffs? A modest step toward modelling both H and A is provided by Smith and Szidarovzsky (2001). Economic intuition tells us that payoffs should matter. But if they do, it must mean that some cost, which is impeding equilibrium, is being overcome at the margin by higher payoffs. The natural psychological assumption is that there are cognitive (or other) costs of getting the best outcome, and more reward enables higher cognitive cost to be incurred to increase net return.

Generally, both theorists and experimentalists must be aware that for any H, any A and any experiment, one can say that:

1. If the outcome of the experiment rejects $(H \mid A)$, then it should be assumed that either H or A may be false, which is an obvious corollary to Proposition 2. This was Einstein's position concerning the Kaufmann results, and was the correct position, whatever later tests might show. After every test, either the theory (leading to H) is in error or the circumstances of the test, A, is in error.
2. If the outcome fails to reject $(H \mid A)$, the experimentalist should escalate the severity of the test. At some point does H (the theory) fail? This identifies the limits of the theory's validity, and gives the theorist clues for modifying the theory.

7. Experimental knowledge drives our methods

Philosophers have written eloquently, and argued intently, over the implications of D-Q, and related issues for the interpretation of science. Popper tried to demarcate science from pseudoscience with the basic rule requiring the scientist to specify in advance the conditions under which he will give up his theory (see note 1). This is a variation on the idea of a so-called "crucial experiment," a slippery concept (Lakatos, 1978, pp. 3–4, 146–148), as is evident from our Proposition 2. Would that such simple rules could make life easier for experimental science!

The failure of all programs attempting to articulate a defensible rational science of scientific method has bred postmodern negative reactions to science and scientists. These exercises and controversies make fascinating reading, and provide a window on the social context of science, but I believe they miss much of what is most important in the working lives of all practitioners.

The point I want to emphasize is that experimentalists do not have to know anything about D-Q and statements like Proposition 2 to appreciate that the results of an experiment nearly always suggest new questions precisely because the interpretation of results in terms of the theory are ambiguous. This ambiguity is reflected in the discussion whenever new results are presented at seminars and conferences. Without ambiguity there would be nothing to discuss. What is the response to this conversation? Invariably, if it is a matter of consequence, experimentalists design new experiments with the intention of confronting the issues in the controversy, and in the conflicting views that have arisen in interpreting the previous results. This leads to new experimental knowledge of how results are influenced, or not, by changes in procedures, context, instructions and control protocols. The new knowledge may include new techniques that have application to areas other than the initiating circumstance. This process is driven by the D-Q problem, but practitioners need have no knowledge of the philosophy of science literature to take the right next local steps in the laboratory. Myopia here is not a handicap.

This is because the theory or primary model that motivates the questions tells you nothing definitive, or even very useful about precisely how to construct tests. Tests are based on extra theoretical intuition, conjectures, and experiential knowledge of procedures. The context, subjects, instructions, parameterization, etc. are determined outside the theory, and their evolution constitutes the experimental knowledge that defines our methodology. The forms taken by the totality of individual research testing programs cannot be accurately described in terms of the falsificationist rhetoric, no matter how much we speak of the need for theories to be falsifiable, the tests discriminating or "crucial," and the results robust.

Whenever negative experimental results threaten perceived important new theoretical tests, the resulting controversies galvanize experimentalists into a search for different, or better, tests, which examine the robustness of the original results. Hence, Kaufmann's experimental apparatus was greatly improved by Bucherer, although there was no question about Kaufmann's skill and competence in laboratory technique. The point is that with escalated personal incentives, and a fresh perspective, scientists found improved techniques. This scenario is common as new challenges bring forth renewed effort. This process generates the constantly changing body of experimental knowledge whose insights in solving one problem often carry over to applications in many others.

Just as often experimental knowledge is generated from curiosity about the properties of phenomena that we observe long before there exists a body of theory that deals specifically with the phenomenon at issue. An example in experimental economics is the continuous double auction trading mechanism, used in financial and commodity markets for at least 200 years, and studied intensively in the laboratory for only 45 years. Its equilibrating properties are well known by experimentalists who have learned much about alternative implementations, robustness, and its performance characteristics in markets much different than those accounting for its origin. Yet constructive models of the double auction are still not available, although abstract theoretical tools have helped to enlighten and explain its performance.[10]

An example from physics is Brownian motion, discovered by the botanist, Robert Brown in 1827, who first observed the unpredictable motion of small particles suspended in a fluid, which motion keeps them from sinking under gravity. This was 78 years before Einstein's famous paper (one of five in 1905) developed the molecular kinetic theory that was able to account for it, although he did not know whether the long familiar observations of Brownian motion represented the phenomena his theory predicted (see Mayo, 1996, Chapter 7 for references and details). The long "inquiry into the cause of Brownian motion has been a story of hundreds of experiments . . . (testing hypotheses attributing the motion). . . either to the nature of the particle studied or to the various factors external to the liquid medium . . . "

(Mayo, 1996, pp. 217–218). The essential point is "that these early experiments on the possible cause of Brownian motion were not testing any full-fledged theories. Indeed it was not yet known whether Brownian motion would turn out to be a problem in chemistry, biology, physics, or something else. Nevertheless, a lot of information was turned up and put to good use by those later researchers who studied their Brownian motion experimental kits" (Mayo, 1996, p. 240). The problem was finally solved by drawing on the extensive bag of experimental tricks, tools and past mistakes that constitute "a log of the extant experimental knowledge of the phenomena in question" (Mayo, 1996, p. 240).

Again, . . . "the infinitely many alternatives really fall into a few categories. Experimental methods (for answering new questions) coupled with experimental knowledge (for using techniques and information already learned) enable local questions to be split off and answered" (Mayo, 1996, p. 242).

The bottom line is that good-enough solutions emerge to the baffling infinity of possibilities, as new measuring systems emerge, experimental tool kits are updated, and understanding is sharpened.

This bottom line also goes far to writing the history of experimental economics and its many detailed encounters with data, and the inevitable ambiguity of subsequent interpretation. And in most cases the jury remains in session on whether we are dealing with a problem in psychology (perception), economics (opportunity cost, and strategy), social psychology (equality, equity or reciprocity), or neuroscience (functional imaging and brain modeling).

8. The machine builders

Mayo's (1996) discussion and examples of experimental knowledge leave unexamined the question of how technology impacts the experimentalist's tool kit. The heroes of science are neither the theorists or the experimentalists, but the unsung tinkerers, mechanics, inventors and engineers that create the new generation of machines that make obsolete yesterday's observations, and heated arguments over whether it is H or A that has been falsified. Scientists, of course, are sometimes a part of this creative destruction, but what is remembered is the new scientific knowledge they created, not the instruments they or their lab technicians invented that made possible the new knowledge. Michael Faraday, "one of the greatest physicists of all time" (Segrè, 1984, p. 134) had no formal scientific education. He was a bookbinder, who had the habit of trying to read the books that he bound. He was preeminently a tinkerer for whom "some pieces of wood, some wire and some pieces of iron seemed to suffice him for making the greatest discoveries" (quoted from a letter by Helmholz, in Segrè, 1984, p. 140). Yet he revolutionized how physicists thought about electromagnetic phenomena, invented the concept of lines of force (fields), and inspired Maxwell's theoretical contributions. "He writes many times that he must experience new phenomena by repeating the experiments, and that reading is totally insufficient for him" (Segrè, 1984, p. 141). This is what I mean, herein, when I use the term "experimental knowledge." I think every experimentalist can relate to what Faraday was talking about. Its what N. R. Hanson (1969) called, understanding the "go" of things. (For a discussion of the tools-to-theories heuristics of discovery in psychology, see Gigerenger, 1991.)

8.1. Technology and science

With the first moon landing, contingent theories of the origin and composition of our lunar satellite were upstaged by direct observation; the first Saturn probe sent theorists back to their desks and computers to reevaluate her mysterious rings, whose layered richness had not been anticipated. Similar experiences have followed the results of ice core sampling in the Artic, and instrumentation for mapping the genome of any species. Galileo's primitive telescope opened a startling window on the solar system, as does Roger Angel's multiple mirror light gathering machines (created under the University of Arizona football stadium) that opens the mind to a view of the structure of the universe.[11] The technique of tree ring dating, invented by an Arizona astronomer has revolutionized the interpretation of archeological data from the last 5000 years; DNA analysis now holds the hope of resolving the long running controversy as to whether the Neanderthal people are one of our modern human ancestors, or represent a failed adaptation that became an extinct branch of the human bush.

Yesterday's reductionists, shunned by mainstream "practical" scientists, create the demand for new subsurface observations, and hence for the machines that can deliver answers to entirely new questions. Each new machine—microscope, telescope, Teletype, computer, the Internet—changes the way teams of users think about their science. The host of auxiliary hypotheses needed to give meaning to theory in the context of remote and indirect observations (inferring the structure of Saturn's ring from earth-based telescopes), are suddenly made irrelevant by deep and more direct observations of the underlying phenomena (fly-by computer enhanced photos). It's the machines that drive the new theory, hypotheses, and testing programs that take you from atoms, to protons, to quarks. Yet with each new advance comes a blizzard of auxiliary hypotheses, all handcuffed to new theory, giving expression to new controversies seeking to rescue T and reject A, or to accept A and reject T.

8.2. Experimental economics and computer/communication technology

When Arlington Williams (1980) created the first electronic double auction software program for market experiments, all of us involved in that enterprise thought we were simply making it easier to run experiments, collect more accurate data, observe longer time series, facilitate data analysis, etc. What were computerized were the procedures, recording and accounting that here-to-fore had been done manually. No one was anticipating how this tool might impact and change the way we thought about the nature of doing experiments. But with each new electronic experiment we were "learning" (impacted by, but without conscious awareness) the fact that traders could be matched at essentially zero cost, that the set of feasible rules that could be considered was no longer restricted by costly forms of implementation and monitoring, that vastly larger message spaces could be accommodated, and that optimization algorithms could now be applied to the messages to define new electronic market forms for trading energy, air emission permits, water and rights in other network industries. In short, the transactions cost of running experimental markets became miniscule in comparison with the pre electronic days, and this opened up new directions that previously had been unthinkable.

The concept of smart computer assisted markets appeared in the early 1980s (Rassenti, 1981; Rassenti et al., 1982), extended conceptually to electric power and gas pipelines in the late 1980s (Rassenti and Smith, 1986; McCabe et al., 1989), with practical applications to electric power networks and the trading of emission permits across time and regions in the 1990s (Rassenti et al., 2001). These developments continue as a major new effort in which the laboratory is used as a test bed for measuring, modifying, and further testing the performance characteristics of new institutional forms.

What is called e-commerce has spawned a rush to reproduce on the Internet the auction, retailing and other trading systems people know about from past experience. But the new experience of being able to match traders at practically zero cost is sure to change how people think about trade and commerce, and ultimately this will change the very nature of trading institutions. In the short run, of course, efforts to protect existing institutions will spawn efforts to shield them from entry by deliberately introducing cost barriers but in the long-run these efforts will be increasingly uneconomical, and untenable

Similarly, neuroscience will revolutionize the experimental study of individual, two-person interactive, and indeed all other, decision making. The neural correlates of decision making, how it is affected by rewards, cognitive constraints, working memory, repeat experience and a host of factors that in the past we could neither control or observe, can in the future be expected to become an integral part of the way we think about and model decision making. Models of decision, now driven by game and utility theory, and based on trivial, patently false, models of mind, must take account of new models of cognitive, calculation and memory properties of mental function that are accessible to more direct observational interpretation. Game theoretic models assume consciously calculating, rational mental processes. Models of mind include non self-aware processes just as accessible to neural brain imaging as the conscious. For the first time we may be able to give rigorous observational content to the vague idea of "bounded rationality" (see McCabe et al., 2001).

In principle the D-Q problem is a barrier to any defensible notion of a rational science that selects theories by a logical process of confrontation with scientific evidence. This is cause for joy not despair. Think how dull would be a life of science if, once we were trained, all we had to do was to turn on the threshing machine of science, feed it the facts and send its output to the printer. In practice the D-Q problem is not a barrier to resolving ambiguity in interpreting test results. The action is always in imaginative new tests and the conversation it stimulates.

My personal experience as an experimental economist since 1956, resonates well with Mayo's critique of Lakatos:

> Lakatos, recall, gives up on justifying control; at best we decide—by appeal to convention—that the experiment is controlled. . . . I reject Lakatos and others' apprehension about experimental control. Happily, the image of experimental testing that gives these philosophers cold feet bears little resemblance to actual experimental learning. Literal control is not needed to correctly attribute experimental results (whether to affirm or deny a hypothesis). Enough experimental knowledge will do. Nor need it be assured that the various factors in the experimental context have no influence on the result in question—far from it. A more typical strategy is to learn enough about the type and extent of their influences and then estimate their likely effects in the given experiment.

Acknowledgments

I want to thank Charles Holt and two anonymous referees for many detailed comments on an earlier version of this paper that led to a substantial revision. It is written out of my personal experience as an experimental economist, but the issues I address are generic and not restricted to any audience.

Notes

1. Such a disagreement led to a $1000 wager between the ecologist, Paul Ehrlich (and two colleagues), and the economist, Julian Simon, concerning the change in the real price of 5 metals after 10 years from September 29, 1980. Simon won the bet, but admitted that he had been lucky in the sense that any particular 10 year period could buck this overall trend. Thus, Ehrlich's loss did not change Ehrlich's views on resource availability, and if the bet had gone the other way it would not have changed Simon's (Simon, 1993, pp. 359–380). Mellers et al. (2001) provides an exercise in adversarial collaboration between Hertwig and Kahneman and arbitrated by Mellers. The most striking finding was not anticipated by any of the three authors. Was there convergence in views? Yes, some, but Hertwig and Kahneman wrote separate summary statements of their different views. Such examples of confrontations are rare, but the image reflected in Robinson's rhetoric is commonplace.
2. Cox and Oaxaca (1990, 1999) use this methodology, not to test equilibrium hypotheses, but to identify supply and demand functions induced by the experimenter on individual subject agents in double auction trading environments known to have rapid equilibrating properties. Comparing the identified functions, measured by alternative econometric estimation procedures applied to the market realizations, with the true induced functions, they derive a test of the efficacy of alternative econometric estimators. They discovered that just as economists do not think of market equilibrium theory as something you can test, so econometricians do not think of an estimation procedure as something that is testable. This is why the Cox/Oacxca "experimentrics" study took so long to be published, and appeared in a journal of the American Statistical Association.
3. Stigler (1957) provides a historical treatment of the concept of perfect competition. Why do we tend to believe or assert that economic theories assume complete information by each agent on the circumstances (tastes, costs) of all other agents? It cannot be because there is a theorem about behavior dealing with information in a constructive process of equilibrium formation, stating that when all agents have such information the equilibrium is attainable, and when they do not have such information the equilibrium fails. The answer is that we confuse the cognitive requirements of analysis used to deduce the properties of end states, with interactive decision processes that are as consciously inaccessible to the agent as to the scientist. Modeling mechanisms is not the same thing as modeling interactive minds. In the former theorist's find it difficult to imagine that a noncooperative equilibrium obtains except under a Cartesian conceptual regime in which each agent calculates (exactly as does the theorist) his or her strategy given the information on all other agents needed to make the calculation. It isn't that Hayek (1945) is wrong; he is irrelevant to game theory in which the concept of a strategy requires one to have information on adversaries—enough information (perfect or probabilistic) to calculate mutually compatible optimal strategies. It's that, or its nothing. Hence, observations are inadmissible outside this conceptual framework, and therefore are not separable from that framework.
4. "Are there any conditions under which double-auction markets do not generate competitive outcomes? The only known exception is an experiment with a 'market power' design reported by Holt et al. (1986) and replicated by Davis and Williams (1991)" (Davis and Holt, 1993, p. 154). The example reported in this exception used a constant excess supply in the market of only one unit, which has inherently weak Walrasian equilibrating properties. Actually, there were three earlier reported exceptions: (1) one in which information about private circumstances is known by all (the alternative to the Hayek Hypothesis as stated above), (2) an example in which the excess supply in the market was only two units, and (3) an example in which monetary rewards were not used. Exception (1) was reported in Smith (1980, 1976, reprinted 1991, Chapter 6, pp. 104–105); (2) in Smith (1965, reprinted, 1991, Chapter 4, p. 67); and (3) in Smith (1962, reprinted, 1991,

Chapter 1, p. 16). Many others have been reported: e.g. when a competitive equilibrium does not exist, the double auction performs poorly (Van Boening and Wilcox, 1996); it also performs poorly when contracts are incomplete causing a property right failure (Fehr and Falk, 1999). Both of the latter provide exceptions that prove the rule: competitive institutions work if competitive equilibria exist and agents have property rights in the resources they use; otherwise it fails.

5. See Soberg (2000) for an excellent summary and discussion of the D-Q thesis, its relevance for experimental economics, and results showing how the process of replication can be used, in the limit, to inductively eliminate clusters of alternative hypotheses and lend increasing weight to the conclusion that the theory itself is in doubt. Perhaps better known to economists than D-Q is the methodological perspective associated with Friedman (1953), which fails to provide an adequate foundation for experimental (field and laboratory) science, but which influenced economists for decades and still has much currency: the proposition is that the truth value of a theory should be judged by its testable and tested predictions not by its assumptions. This proposition is deficient for several reasons: (1) If a theory fails a test, we should ask why, not always just move on to seek funding for a different new project; obviously, one or more of its assumptions may be wrong, and it behooves us to design experiments that will probe conjectures about which assumptions failed. Thus, if first price auction theory fails a test is it a consequence of the failure of one of the axioms of expected utility theory, e.g., the compound lottery axiom? If a subgame perfect equilibrium prediction fails, does the theory fail because the subjects do not satisfy the assumption that agents will choose dominant strategies? Or, was this assumption true but not common knowledge? Or did the subjects fail to use backward induction? Or was it none of the above because the theory was irrelevant to how some motivated agents solve such problems in a world of bilateral reciprocity in social exchange. When a theory fails there is no more important question than to ask what it is about the theory that has failed. (2) Theories may have the if-and-only-if property that one (or more) of the "assumptions" can be derived as implication(s) from one (or more) of its "results." (3) If a theory passes one or more tests, this need not be because its assumptions are correct (Proposition 1). A subject may choose the subgame perfect equilibrium because she believes she is paired with a person that is not trustworthy, and not because she always chooses dominant strategies, assumes that others always so choose, or that this is common knowledge. This is why you are not done when a theory is corroborated by a test. You have only examined one point in the parameter space of payoffs, subjects, tree structure, etc. Your results may be a freak accident of nature, be due to a complex of suitabilities, or in any case may have other explanations.
6. For an investigation of this hypotheses, see McCabe and Smith (2000) who report a comparison of undergraduate and graduate students, and these with economics faculty, in a two-person trust game. The first two groups were indistinguishable, and both earned more money (hardly irrational) than the faculty because with greater frequency they risked defection in offering to cooperate, as against opting for the subgame perfect outcome.
7. See McCabe et al. (1996; reprinted in Smith, 2000, Chapter 8, pp. 162–163, 168) for a trust game with the option of punishing defection in which support for the cooperative outcome does not decrease in repeat single relative to single play across trials, and therefore subjects do not "learn" to play non-cooperatively.
8. Thus, Hertwig and Ortmann (2001, Section 2) argue that scripts (instructions) are important for replication, and that "ad-libbing" should be avoided. Also see Mellars et al. (1999) where words used in instructions were at the heart of the controversy, and in the results that surprised all the authors. He who has not varied instructions will do much to insulate himself from surprises (Hoffman et al., 2000).
9. The results are different if you imbed the ultimatum game in an exchange between a buyer and a seller, or let subjects "earn" the right to be the first mover by winning a pregame contest; dictator game experiments yield different results under a double blind protocol in which neither the experimenter or the subjects know who made what decision (Hoffman et al., 1994). Dictator game results are also influenced by whether you recruit the subjects (volunteers) to the lab for an experiment, or use the students in your class during class (Eckel and Grossman, 2000). The latter are more generous. This can be interpreted as the teacher-is-watching effect.
10. Friedman (1984) and Wilson (1987) account for most of the sparse theory of the double auction that exists. Both limit their analysis to the case in which each trader (buyer or seller) has only a single unit to trade.
11. See Angel (2001) for a brief summary of past, current, and likely future impacts on astronomy of rapid change in optical and infrared (terrestrial and space) telescopes.

References

Angel, R. (2001). "Future Optical and Infrared Telescopes." *Nature Insights*. 409, 427–430.

Banks, J., Olson, M., Porter, D., Rassenti, S., and Smith, V. (2002). "Theory, Experiment and the FCC Spectrum Auctions." *Journal of Economic Behavior and Organization*.

Burnham, T., McCabe, K., and Smith, V. (2000). "Friend-or-Foe Intentionality Priming In An Extensive Form Trust Game." *Journal of Economic Behavior and Organization*. 93(1), 57–73.

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

Cox, J.C. and Oaxaca, R. (1990). "Using Laboratory Market Experiments to Evaluate Econometric Estimators of Structural Models." Working Paper, University of Arizona.

Cox, J.C. and Oaxaca, R. (1999). "Can Supply and Demand Parameters be Recovered from Data Generated by Market Institutions?" *Journal of Business and Economic Statistics*.

Davis, D. and Holt, C. (1993). *Experimental Economics*. Princeton University Press, Princeton.

Davis, D. and Williams, A. (1991). "The Hayek Hypothesis in Experimental Auctions." *Economic Inquiry.* 29, 261–274.

Eckel, C. and Grossman, P. (2000). "Volunteers and Pseudo-Volunteers; The Effect of Recruitment Method in Dictator Experiments." *Experimental Economics*. 2, 107–120.

Einstein, A. (1905/1989). "On the Electrodynamics of Moving Bodies." English Translation by Anna Beck, in The Collected Papers of Albert Einstein. Vol. 2, Princeton University Press, Princeton.

Einstein, A. (1907/1989). "On the Relativity Principle and the Conclusions Drawn From It." English Translation by Anna Beck, in The Collected Papers of Albert Einstein. Vol. 2, Princeton University Press, Princeton, pp. 252–311.

Einstein, A. (1934). *The World As I See It*. Covici Friede Publishers, New York.

Fehr, E. and Falk, A. (1999). "Wage Rigidity in a Competitive Incomplete Contract Market." *Journal of Political Economy*. 107, 106–134.

Fölsing, A. (1997). *Albert Einstein*. Viking, New York.

Friedman, M. (1953). *Essays in Positive Economics*. University of Chicago Press, Chicago.

Friedman, D. (1984). "On the Efficiency of Experimental Double Auction Markets." *American Economic Review*. 74, 60–72.

Gigerenzer, G. (1991). "From Tools to Theories: A heuristic of discovery in cognitive psychology." *Psychological Review*. 98, 254–267.

Hanson, N.R. (1969). *Perception and Discovery*. Freeman, Cooper and Co., San Francisco.

Hayek, F.A. (1945). "The Use of Knowledge in Society." *American Economic Review*. 35, 519–530.

Hayek, F.A. (1978/1984). *Competition as a Discovery Procedure*. Reprinted in Chapter 13, The Essence of Hayek. Hoover Institution Press, Stanford.

Hertwig, R. and Ortmann, A. (2001). "Experimental Practices in Economics." *Behavioral and Brain Sciences*. 24, 383–451.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). "Preferences, Property Rights and Anonymity in Bargaining Games." *Games and Economic Behavior*. 7, 346–380.

Hoffman, E., McCabe, K., and Smith, V. (2000). "The Impact of Exchange Context on the Activation of Equity in Ultimatum Games." *Experimental Economics*. 3, 5–9.

Holt, C., Langan, L., and Villamil, A. (1986). "Market Power in Oral Double Auctions." *Economic Inquiry*. 24, 107–123.

Holt, C. and Laury, S. (2001). "Varying the Scale of Financial Incentives Under Real and Hypothetical Conditions." *Behavioral and Brain Sciences*. 24, 417–418.

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

Lakatos, I. (1978). *The Methodology of Scientific Research Programmers*. Vol. 1/2, Cambridge University Press, Cambridge.

Mayo, D. (1996). *Error and the Growth of Experimental Knowledge*. University of Chicago Press, Chicago.

McCabe, K. and Smith, V. (2000). "A Comparison of Naïve and Sophisticated Subject Behavior with Game Theoretic Prediction." *Proceedings National Academy of Sciences*. 97, 3777–3781.

McCabe, K., Rassenti, S., and Smith, V. (1989). "Designing Smart Computer Assisted Markets for Gas Networks." *European Journal of Political Economy*. 5, 259–283.

McCabe, K., Rassenti, S., and Smith, V. (1998). "Reciprocity, Trust and Payoff Privacy in Extensive Form Bargaining." *Games and Economic Behavior*. 24, 10–24.

McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, R. (2001). "A Functional Imaging Study of Cooperation in Two-Person Reciprocal Exchange." Cognition and Neuroimaging Laboratories, University of Arizona, Tucson.

Mellers, B., Hertwig, R., and Kahneman, D. (2001). "Do Frequency Representations Eliminate Conjunction Effects?" *American Psychological Society*. 12, 264–275.

Nash, J.F. (1950). "The Bargaining Problem." *Econometrica*. 18, 315–335.

Northrup, F.S.C. (1969). "Einstein's Conception of Science." In P.A. Schilpp (ed.), *Albert Einstein Philosopher–Scientist*. LaSalle, IL. Open Court.

Pirsig, R.M. (1981). *Zen and the Art of Motorcycle Maintenance*. Banton Books, New York.

Rassenti, S. (1981). "O-1 Decision Problems with Multiple Resource Constrains: Algorithms and Applications." Ph.D. Thesis, University of Arizona.

Rassenti, S., Smith, V., and Bulfin, R. (1982). "A Combinatorial Auction Mechanism for Airport Time Slot Allocation." *Bell Journal of Economics*, Autumn.

Rassenti, S. and Smith, V. (1986). "Electric Utility Deregulation." In *Pricing Electric Gas and Telecommunication Services*. The Institute for the Study of Regulation, December.

Rassenti, S., Smith, V., and Wilson, B. (2001). "Controlling Market Power and Price Spikes in Electricity Networks." Available at www.ices-gmu.org

Robinson, J. (1979). "What Are The Questions?" *Journal of Economic Literature*. 15, 1318–1339.

Samuelson, P. and Nordhaus, W. (1985). *Economics*. McGraw-Hill, New York.

Segrè, E. (1984). *From Falling Bodies to Radio Waves*. W.H. Freeman, New York.

Shubik, M. (1959). *Strategy and Market Structure*. Wiley, New York.

Simon, J. (1993). *Population Matters*. Transaction Publishers, London.

Smith, V.L. (1980). "Relevance of Laboratory Experiments to Testing Resource Allocation Theory." In J. Kmenta and J. Ramsey (eds.), *Evaluation of Econometric Models*. New York: Academic Press.

Smith, V.L. (1991). *Papers in Experimental Economics* (Collected works). Cambridge University Press, New York.

Smith, V.L. (2000). *Bargaining and Market Behavior*. Essays in Experimental Economics (Collected works). Cambridge University Press, New York.

Smith, V., McCabe, K., and Rassenti, S. (1991). "Lakatos and Experimental Economics." In N. de Marchi and M. Blaug (eds.), *Appraising Economic Theories*. Edward Elgar, London.

Smith, V. and Szidarovzsky, F. (2001). "Monetary Rewards and Decision Cost in Strategic Interactions." *Papers in Honor of Herb Simon*. Herb Simon Memorial Volume by MIT Press.

Smith, V. and Walker, J. (1993a). "Monetary Rewards and Decision Cost in Experimental Economics." *Economic Inquiry*. Vol. 31, April, 245–261. Western Economic Association, Best Article Award, 1993.

Smith, V. and Walker, J. (1993b). "Rewards, Experience and Decision Costs in First Price Auctions." *Economic Inquiry*. Vol. 31, April, 237–244.

Soberg, M. (2000). "The Duhem-Quine Thesis and Experimental Economics." Research Department Paper, University of Oslo.

Stigler, G. (1957). "Perfect Competition, Historically Contemplated. *Journal of Political Economy*. 65, 1–17.

Suppes, P. (1969). "Models of Data." In *Studies in the Methodology and Foundations of Science*. Dordrecht: Reidel.

Van Boening, M. and Wilcox, N. (1996). "Avoidable Cost: Ride a Double Roller Coaster." *American Economic Review*. 86, 461–477.

Williams, A. (1980). "Computerized Double-Auction Markets: Some Initial Experimental Results." *Journal of Business*. 53, 235–257.

Wilson, R. (1987). "On Equilibria in Bid-Ask Markets." In G. Feiwel (ed.), *Arrow and the Ascent of Modern Economic Theory*. MacMillan, Houndmills, UK.