The Changing Relationship between Theory and Experiment in Economics

Robert Sugden†‡

Until recently, economics was generally understood to be a nonexperimental science with a hypothetico-deductive methodology. This article considers how the methodology of economics has changed with the spread of experimental methods. Initially, most experimental economists saw their work as testing pre-existing theories. However, a method of systematic inductive enquiry in which theory plays a less central role is now evolving. This method is structured around the discovery and progressive refinement of regularities. "Exhibits"—experimental designs that generate significant regularities—are taking over some of the functions formerly performed by theoretical models.

1. Introduction. For most of the twentieth century, reports of experiments were almost unknown in the literature of economics. Economics was generally understood, both by practitioners and by methodologists, to be a nonexperimental science. This understanding is encapsulated in an incidental remark in Milton Friedman's famous essay on the methodology of positive economics: "Unfortunately, we can seldom test particular predictions in the social sciences by experiments explicitly designed to eliminate what are judged to be the most important disturbing influences. Generally, we must rely on evidence cast up by the 'experiments' that happen to occur" (1953, 10). The implication is that the methods of economics are an adaptation to the practical impossibility of controlled experiments. But, from the 1980s onward, there has been an explosive growth in the use of experimental methods in economics. The acceptance of experiments as a respectable part of the discipline was given a ceremonial signal in 2002 with the award of the Nobel Prize in economics to two pioneers of experimental methods, Daniel Kahneman and Vernon

†To contact the author, please write to: School of Economics, University of East Anglia, Norwich NR4 7TJ, UK; e-mail: r.sugden@uea.ac.uk.

‡My work was supported by the Economic and Social Research Council of the United Kingdom (award no. RES 051 27 0146).

Philosophy of Science, 75 (December 2008) pp. 621–632. 0031-8248/2008/7505-0011$10.00
Copyright 2008 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/594509 Published online by Cambridge University Press

Smith. In this article, I consider how the methodological practices of economics are changing as a result of the spread of experimental methods.

When experimental economics was taking off in the 1980s, most practitioners interpreted their work as theory testing. It was presupposed that economic knowledge was encoded in a body of theory and that the function of empirical research was to test preexisting theories. But, I shall argue, experimentally observed regularities are now providing an alternative structure for encoding knowledge: experimental designs are taking on roles formerly played by theory.

2. Economics as a Hypothetico-Deductive Science. According to most accounts of the methodology of economics, explanation is achieved by using deductively based theories. In its most traditional form, the deductive method starts from general "laws" whose truth is claimed to be beyond reasonable doubt. More recent methodologists, such as Daniel Hausman (1992), have characterized economics as a hypothetico-deductive science, in which conclusions are deduced from provisional hypotheses, rather than self-evident axioms or known laws, and are then tested against evidence from the real world. In practice, however, economists have been strongly committed to a core set of assumptions about rationality. Until relatively recently, the idea that those assumptions might need to be revised in the light of regularities found in empirical evidence was not seriously entertained.

In a hypothetico-deductive science, theoretical propositions need not be directly grounded in observations of the world: it is the *conclusions* of theoretical reasoning, not the premises, that are required to be confirmed by evidence. One might say that such a science confronts the world only through the medium of theory. Of course, there is still the question of how theoretical premises should be chosen so that their implications are likely to be confirmed. But the advocates of the hypothetico-deductive method have generally had little to say, either descriptively or normatively, about the methods by which scientists construct hypotheses in the light of observation. For methodologists working in the tradition of Karl Popper ([1934] 1959), the creation of hypotheses lies outside the scope of any logic of scientific method. Thus, for example, Friedman (1953, 42–43) says:

> On [the construction of new hypotheses] there is little to say on a formal level. The construction of hypotheses is a creative act of inspiration, intuition, invention; its essence is a vision of something new in familiar material. The process must be discussed in psychological, not logical categories; studied in autobiographies and biographies, not treatises on scientific method; and promoted by maxim and example, not syllogism or theorem.

https://doi.org/10.1086/594509 Published online by Cambridge University Press

Although Hausman (1992) does not endorse Friedman's position, and he suggests in passing that hypotheses might be generated by some inductive "logic of discovery" (302), he does not pursue the question of what this logic might be, and he has almost nothing to say about how economists arrive at their hypotheses.

I shall argue that a method of systematic inductive enquiry is now evolving in experimental economics. I must stress that the method I shall describe does not provide criteria for legitimating inductive inferences. Rather, it organizes the work of a decentralized scientific community in discovering and recording significant regularities in the world and in trying to arrive at sharper, more useful, or more suggestive specifications of those regularities. In empiricist language, progress shows up as a movement from apparently arbitrary and isolated observations to propositions that are more promising candidates for theoretical explanation. A methodological realist might prefer to say that the method is directed toward the discovery of causal mechanisms or 'capacities'. I shall argue that this method differs from the traditional strategies of investigation in economics by being much less dependent on theory. In my limited reading of the philosophy of science—I write as a practitioner of experimental economics—the account of scientific method that comes closest to what I have in mind is that of Deborah Mayo (1996).

3. Exhibits. The first sustained programs of experimental economics—those in which Kahneman and Smith worked—were presented as tests of core components of received theories. Kahneman's program tested economists' assumptions about rational choice, while Smith's tested the hypothesis that, in markets with many buyers and sellers, prices are determined by the equilibrium of supply and demand. In broad terms, the experimental strategy was as follows: take a preexisting theory from the core of economics; configure the laboratory environment so that in most respects it reproduces the abstractions of the theory but allows subjects some of the freedom of action that agents have in the field; then investigate whether the theory's predictions are confirmed. Skepticism about the external validity of experiments was countered by a strategy of argument that has come to be called 'Blame the theory', perhaps first diagnosed by Chris Starmer (1999). This is to point out that any unrealism of the laboratory environment mirrors the unrealism of the theory being tested. The traditional instrumental answer to this—that realism is not required of theoretical assumptions—became less credible as, over the 1950s, 1960s, and 1970s, economists made progressively more ambitious claims about the domain of applicability of their theories. If economic theory could be expected to explain the behavior of voters in elections, donors to charities, and even people choosing marriage partners, why not also that of students

https://doi.org/10.1086/594509 Published online by Cambridge University Press

in controlled experiments with real money payoffs? Thus, a niche for theory-testing experiments came into existence.

Experimental economics began to move beyond theory testing as a result of the accumulation of what I have called *exhibits*—replicable experimental designs that reliably produce interesting results—and through exhibits becoming objects of interest in their own right (Sugden 2005). I have space only for sketches of a few examples of exhibits. The *common ratio effect* is observed in an experimental design involving two tasks, each requiring a choice between a 'riskier' lottery (R_1 in one task, R_2 in the other) and a 'safer' lottery (S_1 or S_2). The lotteries are configured so that, according to expected utility theory (the received theory of choice under risk), whatever preference a person has for R_1 relative to S_1 will be replicated for R_2 relative to S_2. However, the probabilities of winning prizes in R_2 and S_2 are proportionately smaller than in the corresponding R_1 and S_1. It turns out that, contrary to expected utility theory, people tend to be less risk averse when the probabilities of winning are smaller: this is the common ratio effect (Allais 1953). *Focal points* are revealed in an experimental design in which each of two isolated subjects has to choose from the same set of options with the object of choosing the same option as the other. People are remarkably successful at coordinating their choices, even when the set of options is extremely large (Schelling 1960). The *trust game* is a game with two players, the 'sender' and the 'responder'. The sender is allocated a sum of money. He chooses how much of this to keep for himself. The remainder is tripled and transferred to the responder. The responder then chooses how much of this sum to keep for herself; the remainder is returned to the sender. On the assumption that players are self-interested, conventional game theory predicts that no transfers will be made. In fact, senders typically transfer about half of their endowments to responders, and responders return about a third of what they receive (Berg, Dickhaut, and McCabe 1995).

The claim that an exhibit produces a genuine regularity is typically supported by classical statistical tests. Take the common ratio effect. In one standard design, each subject makes two binary choices, between R_1 and S_1 and between R_2 and S_2. The null hypothesis is that subjects act according to expected utility theory but with random error. A subject who makes no error will make either the 'riskier' choices (R_1, R_2) or the 'safer' choices (S_1, S_2). Random error is interpreted to imply that mistakes are equally likely in each choice task, so that, under the null hypothesis, (R_1, S_2) and (S_1, R_2) occur with equal probability. In most experiments, the second of these choice patterns is more frequent than the first to a degree that would be very improbable were the null hypothesis true. The null hypothesis is therefore rejected in favor of the alternative hypothesis that there is a common ratio effect.

https://doi.org/10.1086/594509 Published online by Cambridge University Press

Interpreted as a test of received theory, this is a classic exercise in hypothetico-deductive methodology: we take a well-defined theory, we deduce a specific conclusion, we test this, and the theory is disconfirmed. But this experiment has come to be construed not so much as a test of expected utility theory but as positive evidence of a particular regularity in behavior, namely, the common ratio effect. Expected utility theory merely provides a background against which the effect can be observed: the effect is characterized as a particular kind of deviation from that theory (roughly as probabilities of winning decrease, so does risk aversion). Crucially, we can describe the effect and establish its existence without deducing it as the implication of any well-specified theory. And, in fact, the effects for which exhibits are used as evidence are *not* always theoretical predictions. More commonly, the experiments that discover these effects are motivated only by untheorized hunches.

For example, consider how Berg et al. (1995) present the trust game. They begin with the claim: "A fundamental assumption in economics is that individuals act in their own self interest." They contrast this assumption with the hypothesis that "trust is an economic primitive" (122–123); that is, that economic agents have some direct motivation to trust others and to repay other agents' trust in them. They refer in passing to a number of different theoretical explanations of trust that have been developed by economists, but they do not apply these theories to the trust game itself, and they do not interpret their results as confirming any specific theoretical prediction. Summarizing their findings, they point to a well-defined negative conclusion—the rejection of the self-interest hypothesis. They are much less specific about the nature of the regularity they *have* found, describing their experiment as providing "evidence that people are . . . willing to reward appropriate behavior" and "strong support to . . . efforts to integrate reciprocity into standard game theory" (Berg et al. 1995, 138–139). It seems that Berg et al. see their experiment as establishing the existence of a particular regularity or 'effect' and not merely as a test of a received theory. They interpret that regularity as evidence that people are motivated by trust, but the concept of 'trust' is intuitive and untheorized.

For economists trained in hypothetico-deductive methods, it may be tempting to think that exhibits of this kind are valuable only in prompting searches for explanatory theories. (Berg et al.'s remarks about the need to integrate reciprocity into standard game theory hint at just this idea.) And it is undoubtedly true that many recent developments in economic theory have been motivated by a desire to explain otherwise anomalous exhibits. Still, it is important to recognize that exhibits can be useful as explanatory devices in their own right: they can *substitute for* theory.

Schelling's (1960) study of focal points is an example. Schelling presents

https://doi.org/10.1086/594509 Published online by Cambridge University Press

the idea of focal points through a discussion of a wide range of coordination games. These games are clearly intended as simple models of coordination and bargaining problems that are faced in the real world. However, they are not theoretical models in the usual sense of formally described worlds in which abstract agents act according to pre-specified principles. Rather, they are well-specified games to be played by *real* individuals: they are experimental designs. Using a mixture of conjectures and experimental results, Schelling argues that ordinary human players can solve coordination problems by finding focal points. As a result of Schelling's work, the concept of a focal point has become part of the standard toolkit of game theory, being used to resolve otherwise intractable problems of equilibrium selection. However, there is still nothing approaching a received *theory* of focal points. Game theorists work with an intuitive rather than a formal understanding of focal points, one grounded in experimental rather than theoretical models. If asked to explain what a focal point is, most game theorists would simply refer to Schelling's exhibits.

4. Refining Exhibits. Exhibits typically arouse interest because they reveal regularities that received theories cannot explain. The obvious hypothetico-deductive response is to look for ways of revising the hypotheses on which the relevant theory is based so as to encompass the newly discovered regularities while retaining as much as possible of that theory's existing predictive power. As I have already pointed out, many economists have responded to anomalous exhibits in just this way. However, I am concerned with a different response, one that uses experimental rather than theoretical methods. The aim here is to *refine* exhibits, so as to uncover regularities that are (as I have said) "sharper" or more "useful" or "suggestive" than the originals. These are vague terms, which I will try to elucidate by examples.

One broad category of refinement is to test exhibits for *robustness*. To say that a regularity observed in one experimental design is robust is to say that similar regularities are found in similar designs: minor differences in parameter values, instructions, incentives, subject pools, and so on do not make major differences to the results. After an exhibit has been discovered, its robustness is typically tested by running variants of the original design. If it proves robust, we can feel confident in describing the exhibit in generic terms, without referring to every detail of the original design. This is 'refinement' in the sense of simplification or purification.

Robustness tests are often prompted by skepticism about the external validity of particular exhibits. Many economists have doubted whether the behavior of student subjects facing "artificial" tasks without prior experience and with only small (and sometimes no) financial incentives is

https://doi.org/10.1086/594509 Published online by Cambridge University Press

representative of the behavior of "real" economic agents. This skepticism has been expressed in more specific questions about the robustness of exhibits to such design variations as increases in incentives, changes in incentive mechanisms, changes in instructions, repetition of tasks, and the use of different subject pools. There have been many investigations of the effects of rerunning classic experiments with larger incentives and unusual subject pools. In a famous early example of this strategy, Sarah Lichtenstein and Paul Slovic (1971) first reported the 'preference reversal' exhibit (an apparent inconsistency between preferences revealed in choices and preferences revealed in monetary valuations), citing evidence from experiments that had used student subjects and payoffs that were either hypothetical (in two cases) or small (in the third). Having found a strong effect, they replicated their design with larger payoffs and with participants recruited at a Las Vegas casino, and they found similar results (Lichtenstein and Slovic 1973). Another standard robustness test is to investigate whether the regularities found in an exhibit persist if subjects face the same task repeatedly. For example, several studies have investigated whether the disparity between (low) willingness-to-pay valuations and (high) willingness-to-accept valuations of goods—another famous exhibit—decays if valuations are elicited in repeated markets. The jury is still out on this issue, but the balance of evidence suggests that disparities do decay under these circumstances (Loomes, Starmer, and Sugden 2003).

In most cases, robustness tests are not prompted by preexisting theories that predict particular effects of incentives or experience on the prevalence of anomalies. Rather, they start from curiosity or skepticism about the robustness of particular exhibits, often supported by a background belief that received theories work reasonably well in the domains in which they are usually applied. The hypotheses that people would be less likely to violate received theory if their decisions had more substantial consequences or if they had more experience of relevantly similar problems are conjectures or hunches rather than theoretical predictions. Still, these hypotheses can be and are tested experimentally. Experimental designs are substituting for theoretical models as vehicles for firming up hunches and appraising their credibility.

A different way in which the regularity captured by an exhibit can be refined is to break it down into finer-grained components. For example, consider the ultimatum game, in which one player (the 'proposer') proposes a division of a fixed amount of money between himself and his coplayer (the 'respondent'); the respondent either agrees to the division or she does not; if she does not agree, neither gets anything. As far as the behavior of proposers is concerned, the surprising observation is that, of the stock of money available for division, most proposers offer relatively large shares (typically around 40%) to their coplayers. One possible ex-

https://doi.org/10.1086/594509 Published online by Cambridge University Press

planation is that proposers have a 'taste for fairness': they *want* their coplayers to receive relatively large shares. An alternative explanation is that proposers act out of self-interest in the belief that responders will reject small offers. One way of trying to distinguish between these explanations is to investigate what happens if the rules of the game are changed so that responders are not able to reject proposers' offers. The resulting game is the *dictator game*. The idea behind this game, first studied by Kahneman, Knetsch, and Thaler (1986), is that proposers who act on self-interest have no reason to offer anything, while a taste for fairness still provides a motivation for positive offers. Thus, by comparing behavior in ultimatum and dictator games, we can break down the positive offers of the ultimatum game into two components—a component that persists in the dictator game and the residual. The former seems to be most naturally interpreted as resulting from some form of nonselfish motivation, the latter as the effect of self-interest. Kahneman et al.'s experiment allowed proposers to choose between only two alternative divisions (giving either 10% or 50% to the coplayer); a large majority chose the equal division, suggesting that a taste for fairness might be the main factor at work in ultimatum games. However, subsequent dictator game experiments in which proposers' choices have been unconstrained have found that the modal offer to coplayers is roughly half the value observed in ultimatum games, implying that both factors are significant (e.g., Forsythe et al. 1994).

A regularity can be decomposed in an informative way even if, for some or all of the mechanisms that are being separated out, there is no well-developed theoretical explanation. Take the case of the first dictator game experiment. The rhetoric of the opening paragraphs of Kahneman et al.'s paper strongly suggests that their prior expectation was that ultimatum game behavior is explained by a taste for fairness. (Conventional economists are teased about their 'true belief in nonfairness' and their 'embarrassment' in the face of supposedly transparent evidence to the contrary.) Yet Kahneman et al. do not propose any particular theory or definition of fairness; they simply use the findings of ultimatum game experiments as prima facie evidence that people are motivated by "preferences . . . for treating others fairly." Nor do they present or cite any theory that would predict positive offers in ultimatum games and zero offers in dictator games, presumably because (as far as they knew, and as far as I know) no one had ever proposed such a theory. So the dictator game is neither a test of a theory of fairness nor a test of a theory of self-interest. Kahneman et al. present the experiment as a way of answering the question: "How prevalent is "fair" behavior in the absence of enforcement?" The implication is that 'enforcement' (working through the power of respondents to reject unfair offers) might be expected to *increase*

https://doi.org/10.1086/594509 Published online by Cambridge University Press

the frequency of fair offers; the experiment is a way of finding out *by how much* that frequency is increased.

5. Exhibits as Constructs. So far, I have considered exhibits as explanatory devices (as in the use of focal points in game theory) and as displaying anomalies that need to be explained. But there are other methodological strategies that, while investigating the regularities captured by exhibits, are not primarily directed at explanation. The essential idea behind these strategies is to investigate how the strength or prevalence of a regularity is affected by factors that are not interpreted as potential causal mechanisms but are of interest in the field. One such strategy is to investigate whether the prevalence of a regularity varies according to the demographic characteristics of the subject pool. Thus, one might ask whether, after controlling for other factors, the probability of observing a particular anomaly varies according to subjects' age, gender, educational attainment, social class, or ethnicity. In the early years of experimental economics, there was little interest in such investigations, but they are now an accepted genre.

I suggest that this change in attitudes to demographic variables is another sign of a shift of emphasis from theory testing to inductive generalization. When an experiment is being used to test a general theory of decision-making behavior, it does not much matter whether the subject pool is representative of a wider population. Take the case of expected utility theory. That theory assumes that every individual has preferences that satisfy certain general principles; in other respects, such as the degree of risk aversion, preferences are allowed to differ among individuals in ways that the theory does not attempt to explain. By virtue of the general assumptions made about preferences, *every* individual's behavior is predicted to have certain general properties in common, including the absence of the common ratio effect. Thus, the theory can be tested by taking *any* group of individuals and investigating the frequency of the common ratio effect in their behavior. The idea that it is useful to investigate how demographic variables have an impact on exhibits depends on seeing exhibits not in relation to theories but in relation to behavior in the field. If the regularity captured by an exhibit is interpreted as a representation of some feature of field behavior, we may try to learn about the latter by manipulating the exhibit so as to represent different field conditions or by using subject pools that represent different field populations.

This strategy is illustrated by the ways in which the trust game has been used since it was first exhibited by Berg et al. (1995). As I have already described, Berg et al. present this exhibit as a test of the hypothesis of self-interest and as evidence that people are motivated by trust. Eight years later, however, in a review of experimental investigations of games,

https://doi.org/10.1086/594509 Published online by Cambridge University Press

Colin Camerer (2003, 85) describes the trust game very differently, as "a beautiful simple game to measure trust." An experiment reported by Nancy Buchan, Rachel Croson, and Robyn Dawes (2002) provides an example of this new use of the experimental design. In this experiment, the trust game is played by student subjects in four countries (China, Japan, Korea, and the United States). Buchan et al. do not set out to test any specific hypotheses about cross-country differences in trust; they merely *measure* the extent of trust in each of the four subject groups. However, this is presented as an attempt to learn about differences in trust *in the field*. While Buchan et al.'s measures are descriptions of the behavior of student subjects in particular experimental designs, they are interpreted as indicators of trust and reciprocation in a country as a whole, understood as factors that might have implications for "national prosperity, economic growth, or stability" (201).

In fact, no clear patterns emerge from Buchan et al.'s international comparisons. In any case, extreme caution is required in interpreting comparisons among different convenience samples (as opposed to different groups formed by random subsampling from a given convenience sample). Still, the methodological strategy is significant. 'Trust' as measured in the trust game is being treated as what in psychology is called a *construct*—a concept that is defined in terms of a method of measurement and that is intended to represent a real phenomenon that might otherwise be only vaguely defined. Or, to put this another way, the experimental trust game is being used as a *model* of trust in the field. The idea that experiments and theoretical models play similar roles in scientific enquiry has recently been argued by a number of philosophers of science, including three specialists in the methodology of economics: Francesco Guala (1998), Uskali Mäki (2005), and Mary Morgan (2005).

Experimental models may be more useful than theoretical ones when the modeler has an intuitive understanding of the 'target system' to be modeled but is not sure how the mechanisms at work in that system are best described. This seems to be true of trust. It seems clear enough that a sender in the trust game who transfers money to a responder is 'showing trust' and that a responder who transfers money back to the sender is 'returning trust', using these expressions in their everyday senses. However, we may be quite unsure about what *motivates* trust—both in the game and in the field. (Is it motivated by far-sighted self interest? By altruism? By reciprocity?) If we are unsure of this, it will be difficult to construct *mathematical* models of trust that we can be confident will respond to our manipulations in ways that correspond with behavior in the field. One way of getting round this problem is to use *experimental* models, in which the actors are real human beings. The intention is that the experimental environment will activate in the subjects the same mo-

https://doi.org/10.1086/594509 Published online by Cambridge University Press

tivations as determine their behavior in trust problems in the field—whatever those motivations may be. A person's propensity for trust can then be represented by constructs that are defined in terms of an experimental design rather than by concepts that are defined within a theory.

6. Conclusion. Traditionally, economists have prided themselves on what they have seen as the rigor and unity of their theory, grounded in a common core of formal principles of rational choice. They have been predisposed to see the world through the lens of this theory. However, the expansion of experimental methods in economics seems to be inducing a profound shift in prevailing understandings of the role of theory. At the heart of this shift is the idea that experiments can be understood in relation to behavior in the field and not solely or primarily in relation to preexisting theories. Once this idea is accepted, it becomes possible to organize our knowledge of the world in terms of experimentally observed regularities rather than in terms of overarching theoretical systems.

REFERENCES

Allais, Maurice (1953), "Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine" [Rational behavior under risk: A critique of the postulates and axioms of the American School], *Econometrica* 21: 503–546.

Berg, Joyce, John Dickhaut, and Kevin McCabe (1995), "Trust, Reciprocity, and Social History", *Games and Economic Behavior* 10: 122–142.

Buchan, Nancy, Rachel Croson, and Robyn Dawes (2002), "Swift Neighbors and Persistent Strangers: A Cross-Cultural Investigation of Trust and Reciprocity in Social Exchange", *American Journal of Sociology* 108: 168–206.

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

Forsythe, Robert, Joel Horowitz, N. E. Savin, and Martin Sefton (1994), "Fairness in Simple Bargaining Experiments", *Games and Economic Behavior* 6: 347–369.

Friedman, Milton (1953), "The Methodology of Positive Economics", in Milton Friedman, *Essays in Positive Economics*. Chicago: University of Chicago Press, 3–43.

Guala, Francesco (1998), "Experiments as Mediators in the Non-laboratory Sciences", *Philosophica* 62: 57–75.

Hausman, Daniel (1992), *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Kahneman, Daniel, Jack Knetsch, and Richard Thaler (1986), "The Behavioral Foundations of Economic Theory", *Journal of Business* 59: S285–S300.

Lichtenstein, Sarah, and Paul Slovic (1971), "Reversal of Preferences between Bids and Choices in Gambling Decisions", *Journal of Experimental Psychology* 89: 46–55.

——— (1973), "Response-Induced Reversals of Preference in Gambling: An Extended Replication in Las Vegas", *Journal of Experimental Psychology* 101: 16–20.

Loomes, Graham, Chris Starmer, and Robert Sugden (2003), "Do Anomalies Disappear in Repeated Markets?", *Economic Journal* 113: C153–C166.

Mäki, Uskali (2005), "Models Are Experiments, Experiments Are Models", *Journal of Economic Methodology* 12: 303–315.

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

Morgan, Mary (2005), "Experiments versus Models: New Phenomena, Inference and Surprise", *Journal of Economic Methodology* 12: 317–329.

https://doi.org/10.1086/594509 Published online by Cambridge University Press

Popper, Karl ([1934] 1959), *The Logic of Scientific Discovery*. Reprint. Originally published as *Logik der Forschung* (Vienna: Springer). London: Hutchinson.
Schelling, Thomas (1960), *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Starmer, Chris (1999), "Experiments in Economics: Should We Trust the Dismal Scientists in White Coats?", *Journal of Economic Methodology* 6: 1–30.
Sugden, Robert (2005), "Experiments as Models and Experiments as Tests", *Journal of Economic Methodology* 12: 291–302.

https://doi.org/10.1086/594509 Published online by Cambridge University Press