



Artificiality: The tension between internal and external validity in economic experiments

Arthur Schram

To cite this article: Arthur Schram (2005) Artificiality: The tension between internal and external validity in economic experiments, Journal of Economic Methodology, 12:2, 225-237, DOI: [10.1080/13501780500086081](https://doi.org/10.1080/13501780500086081)

To link to this article: <https://doi.org/10.1080/13501780500086081>



Published online: 16 Aug 2006.



Submit your article to this journal [↗](#)



Article views: 734



View related articles [↗](#)



Citing articles: 68 View citing articles [↗](#)

Artificiality: The tension between internal and external validity in economic experiments

Arthur Schram

Abstract The artificiality of a laboratory situation is placed in the context of the tension between external and internal validity. Most economists consider internal validity to be most important. A proper evaluation of the ‘artificiality criticism’ (a lack of external validity) requires distinguishing the various goals experimentalists pursue. External validity is relatively more important for experiments searching for empirical regularities than for theory-testing experiments. As experimental results are being used more often in the development of new theories, a methodological discussion of their external validity is becoming more important.

Keywords: laboratory experiments, artificiality, internal validity, external validity

1 INTRODUCTION

The use of experiments in economics has grown tremendously over the last few decades and is still growing. When a new academic field emerges, it is understandable that methodological discussions are given little attention. This has led to a situation, however, where (experimental) economists lag behind their colleagues in other disciplines in discussing important methodological issues. Whereas methodological issues are a standard part of almost any curriculum in psychology, they are virtually missing in the education of (experimental) economists. For example, the leading textbook on experimental economics (Davis and Holt 1993) has one subsection called ‘Some Procedural and Design Considerations’ in which the ‘standard’ way of running experiments in economics is explained.

In spite of this shortcoming, economists appear to have achieved a widespread consensus on a variety of issues concerning how to run an experiment. As noted by Loewenstein (1999):

(economic) experiments share methodological features to a much greater extent than is true of experiments conducted by psychologists (F26).

Experimental economists agree on important methodological issues, like using ‘salient rewards’ to motivate subjects and a ban on using deception.

However, until recently, there were very few studies critically analyzing this methodology.

Some recent studies on the methodology of experimental economics focus on issues on which many economists agree (though psychologists often have opposite opinions). McDaniel and Starmer (1998) and Ortmann and Hertwig (2002) discuss and support the ban of deception. Camerer and Hogarth (1999) review (with inconclusive results) 74 experiments to study the impact of using financial incentives. Hertwig and Ortmann (2001) find stronger support for using financial incentives.

This note discusses a different methodological issue: the artificiality of the laboratory situation. The question is whether the stylized form of experimental institutions allow for conclusions pertaining to the 'real world'. This is closely related to the discussion about the internal versus external validity of a research project. Therefore, this paper starts with this discussion, followed by a brief overview of how experimental economists have thought about this issue over the years. Subsequently, it discusses the importance of artificiality for experiments with various goals.

2 INTERNAL AND EXTERNAL VALIDITY

The internal validity of an experiment refers to the ability to draw confident causal conclusions from the research (Loewenstein 1999). An internally valid design will yield results that are robust and replicable. External validity refers to the possibility of generalizing the conclusions to situations that prompted the research.¹ There is an obvious tension between the two. Where internal validity often requires abstraction and simplification to make the research more tractable, these concessions are made at the cost of decreasing external validity. Loewenstein (1999) points out that while this tension is a starting point in learning research methods in psychology, the discussion is often completely neglected by economists.

Loewenstein also notes that in psychology, experiments are considered to score highly on the criterion of internal validity and lowly on external validity. The opposite holds for field studies. Field experiments hold a position in between the two. Experimental economists, on the other hand, believe that their methodology is superior to field studies on both dimensions but Loewenstein argues that this is only justified for experiments studying highly structured market settings. Ortmann (2003) notices that the increase in the popularity of field experiments has created an interest in the external validity of laboratory experiments that is long overdue (e.g., Harrison and List 2003).

A major obstacle to the external validity of an experiment is the artificiality of the setting. If the laboratory institutions and incentives do not sufficiently mirror those of the outside-the-laboratory situation they intend to study, the loss of external validity may be significant. Of course,

one needs to interpret the word 'sufficiently'. In my view, this will vary substantially, depending on the type and goal of the experiment.

It is not surprising that economists focus more on internal than on external validity. The long tradition of deductive reasoning and modeling in economics relies strongly on internal validity. The way experiments are often applied fits nicely into this tradition. It is noticeable that in response to criticisms concerning external validity, experimental economists have typically resorted to arguments relating the experiments to theory, as will be discussed in the next section.

3 THE POINT OF VIEW OF EXPERIMENTAL ECONOMISTS

Experimental economists are often confronted with the alleged artificiality of their experiments. The most often heard comment on experimental studies may be that the experiment does not reflect the 'real world' and therefore does not teach us much about economics. This reaction seems to have been around for quite a while. The experimental literature of twenty to thirty years ago contains various 'remarks' that appear to try to convince the reader of the external validity of the research.

Two North American pioneers in experimental economics, Charlie Plott and Vernon Smith, explicitly deal with this issue in often-cited papers. In a literature review, Plott (1982) dedicates one section ('Defense of Experiments') to methodological issues. He argues that laboratory markets are 'real' markets and that general principles of economics, if valid, should apply there just as in any other market. The core of his argument is that experiments should primarily be used in conjunction with theory:

Thus theory ... serves importantly to simplify the experimental process. The more that accepted theory can be invoked, the less that the experimental process needs to 'mirror' the natural analog (Plott 1982: 1521).

In other words, Plott takes the position that experiments do not need to be realistic so long as they closely implement the theory being tested. Hence, this approach primarily sees experiments as a method for testing and comparing theories. Plott notes that the proposition that experiments are too artificial cannot be falsified. The only response can be to show that experiments have been useful in theory development and in applied research.

Smith (1976, 1980, 1982, 1985, 1989, 2002) provides a formalization of an economic experiment in an attempt to provide a framework for methodological discussions. Smith (1982) discusses four 'precepts' that are sufficient to allow for a controlled testing of economic theories ('nonsatiation', 'saliency', 'dominance' and 'privacy'). He essentially argues that these precepts ensure the internal validity. A fifth precept, 'parallelism' is sufficient to make experimental results transferable to other environments.

Parallelism is defined as the applicability of propositions that have been tested in the laboratory to 'nonlaboratory microeconomies where similar *ceteris paribus* condition hold'. Hence, parallelism refers to the external validity of experiments. In a response to Cross' (1980) criticism that the differences between experimental situations and outside-the-laboratory markets are too large, Smith (1982) downplays the argument and appears to take the position that it is up to the critics to falsify parallelism of any particular experimental result. However, in Smith (1985) he argues that:

... empirical investigations of all aspects of parallelism between laboratory and field behavior are important. Similarly, our experimental and other investigations should not be *confined* to testing formal theory ... since this objective requires us to impose more structure on free play of decision making than ultimately may be justified (266).

This is a first explicit recognition by a leading experimental economist that we should, perhaps, pay more attention to the external validity of our research. Note, however, that Smith explicitly refers to experiments that do not aim at testing theories. Most experimentalists still primarily see experiments as a way of testing theories and support the 'classic' view that the external validity is not an issue.

The discussion was picked up again in 1999. Starmer (1999a) again focuses on experimentation as a test-bed for economic theory. He dismisses the artificiality argument in a way similar to Plott and Smith, by referring to the link between experiments and theory. He does, however, qualify the classic position by noting that theories can be insulated from refuting experimental evidence by restricting their domain. In other words, some theories may not have the claim to general applicability that allows them to be tested in the laboratory in the way suggested by Plott and Smith. A novel argument (also noted by Roth 1995a) is that experimental results can also be used to successfully adapt theories and develop new ones in light of repeated experimental observations. Hence, he sees an ongoing dialogue between theory and experiments.²

Loewenstein (1999) discusses the external validity of economic experiments.³ He argues that this is high for a number of situations (in particular, highly structured markets) but very limited for most applications. His point is that many features of a typical experiment are not relevant for decision making outside of the laboratory. One feature he elaborates on is repetition, arguing that the stationary replication of decisions commonly imposed in experiments is not characteristic of everyday life. Binmore (1994) takes the opposite point of view, by arguing that experimental environments cannot replicate the experience and familiarity that humans have in everyday decisions.

Loewenstein also criticizes the lack of attention for the context in which laboratory decisions are placed. He advocates a context that is as close as

possible to that in which economic agents actually operate. His bottom line is that experimental economists:

... have not been able to avoid the problem of low external validity that is the Achilles heel of all laboratory experimentation (F33).

Note that there is no reference to the relationship between theory and experiments in this analysis. It is this relationship (the internal validity) that is at the core of the arguments of the other studies mentioned.

Starmer (1999b) gives the most comprehensive discussion of artificiality. He asserts that arguments against experiments based on artificiality are typically arguments against specific applications (e.g., the subject pools, the level of financial incentives or the design of the laboratory institutions) and do not reflect any 'deep philosophical objection to experimenting on humans'. Although he acknowledges that limitations exist in the extent to which naturally occurring economic environments can be replicated in the laboratory, he does not expect these to be 'especially binding'. Starmer limits his discussion to experiments that are meant to test economic theories, however.

The main problem when testing any theory is the Duhem-Quine thesis common to most empirical research. This involves the central tenet that one can never test a theory (or hypothesis) in isolation. One is always simultaneously testing various auxiliary hypotheses. In an experiment, these auxiliary hypotheses include the assumption that subjects are motivated and understand the instructions, and that the theory is applicable to the experimental design. Starmer concludes that a firm outcome in this debate is not possible, because the 'dedicated sceptic' cannot be proven wrong by logical arguments. The question is simply whether or not the scepticism is reasonable for any specific experiment.

Starmer sees Smith's (1982) precepts as an elaboration of the auxiliary hypotheses. One does not know whether the precepts hold. He posits that Smith and Plott used their methodological writings more to 'sell' experimental economics to a general audience than to rigorously evaluate the methodology. As a consequence, he asserts, an 'official rhetoric' has emerged amongst experimentalists that buries the difficulties at hand. On the other hand, theorists appear to take positions attempting to make them immune to experimental testing. Starmer advocates a more open approach from both sides. He sees great value in experimentation, if experimentalists are willing to make stronger efforts to further explore the external validity of their work.

Starmer's conclusion brings him close to Smith's 1985 point of view: it is time for experimentalists to explicitly focus on the external validity. Neither author gives suggestion on how this endeavor should be undertaken, however. In my view, the answer to this question depends strongly on the goal one has with the experiment.

4 ARTIFICIALITY AND THE TYPE OF EXPERIMENT

Experimentalists distinguish various types of experiments. Davis and Holt (1993) distinguish (i) tests of behavioral hypotheses; (ii) theory stress tests; and (iii) searching for empirical regularities. Many add (iv) advising policy makers to this list. Similarly, Roth (1995a) distinguishes (i) 'speaking to theorists'; (ii) 'searching for facts'; and (iii) 'whispering to the ears of princes'. A proper evaluation of the artificiality criticism is not possible without distinguishing between the goals experimentalists pursue. Therefore, I will continue the discussion for each type of experiments separately.

4.1 Testing theory

Given the rhetoric used by, especially, Plott and Smith, one would start to believe that they see this as the only goal of experimentation. Though this is not the case, I do estimate that this is the category in which most experiments fall.⁴ The bottom line in much of the literature discussed in the previous section is that the internal validity of the experimental research (in relation to the theory tested) is considered to be much more important than the external validity. Despite calls by Smith and Starmer for a stronger focus on the external validity, there appears to be a consensus that much can be learned from the interaction between experiments and theory.

The main issue at hand is the Duhem-Quine thesis discussed above. This may limit the internal validity of experiments aimed at testing a theory. However, in many experiments comparative static predictions of a theory are tested under *ceteris paribus* conditions. It seems to me that most (if not all) of the auxiliary hypotheses may confidently be expected to fall within these *ceteris paribus* conditions. Take, for example, an experiment where one observes distinct behavior in two treatments that differ only in one parameter. A claim that subjects understood the instructions differently in one treatment than in the other would seem unreasonable. At the very least, the person making the claim should provide (logical and/or testable) arguments why this is the case. The same holds for any claim that a causal relationship observed in the laboratory depends on a third variable that may be different inside and outside of the laboratory (as discussed by Guala 2002, for example); or claims that consciously participating in an experiment conditions responses in a way that affects the causal relationship observed (Starmer 1999b). In all of these cases, the burden of the proof lies with the claimant.

Two kinds of experiments fall within this category. First, experiments can be used to test theories about how institutions work by recreating these institutions in the laboratory. Second, experiments can be used to test the behavioral assumptions underlying a theory. These two do not always coincide.

Consider a theory comparing the outcome of a first price auction to that of a second price auction. This can be tested in the laboratory. The outcome tells us something about the institutions but not necessarily about the underlying hypotheses. For example, if the theory assumes that an individual's bid is linear in her value then support of the predicted (comparative static) outcome does not necessarily imply that bids are indeed linear in values. This hypothesis should be tested directly using data on values and bids. On the other hand, support of the linearity assumption does not necessarily imply support of the hypothesis at the institutional level. Both research approaches are valid. The experimental literature seems to have achieved a fine balance between the two.

I find the consensus on the role of experiments in testing theories well founded. A large number of economic theories have been tested in the laboratory and results have led to a critical reassessment of many, especially theories on individual behavior (like models of individual choice under uncertainty, Starmer 1999a). Other theories, notably general theories about economic institutions, find support in a large number of experiments (many theories in industrial organization, especially about the functioning of markets, Holt 1995). In the latter case, this has led to an exploration of the limits of the theory as well as to policy applications. Both are discussed below.

Finally, the consensus should not be interpreted as implying that all economic theories are appropriate for testing with laboratory data nor that laboratory data will tell the whole story. For example, theories that involve evolutionary processes or norms that cannot be controlled in the laboratory would be difficult to test (although in many cases elements of the behavioral hypotheses underlying these theories could be tested). Moreover, experimental evidence in support of a theory should not be interpreted as conclusive evidence in favor of the thesis. This is the stage where a shift from internal to external validity is desirable, for example by moving to stress tests.

4.2 Theory stress tests

In the consensus on experiments as test beds for theories, the focus is almost entirely on what it means if a theory is rejected. The bottom line is that there is no reason to believe that a general theory that is rejected in the laboratory would work well in the world outside of the laboratory.

There is much less discussion about the implications of supporting a theory. A theory that passes the laboratory test is not necessarily applicable to the outside world. Experimentation is only one way to increase our understanding of economic behavior. If a theory is supported, one can move further along various lines, including computer simulations, empirical analysis of field data and the application of field experiments. These serve to

further test the external validity of the theory (and experiments). Compare this to the development of a new airplane. After a theoretical design, a test in a wind tunnel is the stage of laboratory experimentation. If it does not 'crash' in this experiment, the plane is not immediately used for the transport of passengers, however. One will typically conduct further tests in the wind tunnel under extreme circumstances. In addition, further testing including 'real' flights without passengers will be conducted.

Examples of stress tests are often found in market experiments. Ever since the early experiments by Smith (1962), it has been regularly observed that the equilibrium model of supply and demand finds large-scale support in double auction laboratory markets. These markets involve buyers submitting bids and sellers placing price offers in a centralized market, where a trade takes place every time a bidder and seller agree on a price.⁵ Competitive equilibrium theory is easily confirmed in these markets if the number of buyers and sellers is large and aggregate demand (supply) is decreasing (increasing). One stress test that has been conducted with respect to competitive equilibrium theory is limiting the number of sellers. The question asked is how many sellers are necessary for the double auction to generate a competitive equilibrium? Smith and Williams (1990) observe this outcome in many double auction markets with only one or two sellers. A second stress test of the theory is the variation of shapes and stability of the demand and supply curves. Holt *et al.* (1986) study various combinations of horizontal demand and supply curves. They find that the competitive equilibrium prediction is robust to such variations.

Such stress tests allow one to test the domain of applicability of a theory. This is typically wider than the domain in which the theory's assumptions are literally true. In fact, some economic theories are based on assumptions that are obviously not true in a literal sense (e.g., any assumption involving an infinite number of individuals or an infinite horizon). Other theories are based on assumptions that most people would agree to be often untrue if taken literally (e.g., unlimited computational abilities of decision makers or firms as single-person decision makers). Stress tests in the laboratory are a way to explore the applicability of a theory while relaxing some of the assumptions in a controlled way. In other words, it allows one to test the external validity of a theory step by step. This reduces the need to focus on the external validity of the original experiment testing the theory. It does make artificiality more of a problem for stress test experiments than for theory testing experiments, however.

4.3 Searching for empirical regularities

In many cases, no economic theory exists or the theories that do exist are rejected in the laboratory. For this type of situation, experiments are often used to establish and document 'stylized facts', which may be either

observed phenomena (e.g., groups of individuals are better at solving optimization problems than individuals are) or observed causal effects (e.g., training in economics leads to more selfish behavior). It is definitely an advantage of laboratory experiments that stylized facts can be collected under carefully controlled circumstances. However, the external validity often deserves much more attention here than it gets.

Consider bargaining experiments. The observation that subjects give money to strangers in a dictator game has been widely observed (Roth 1995b).⁶ It is also well documented that the amount one is willing to pay is sensitive to the context (e.g., anonymity plays an important role). However, it appears to me that the artificiality of these experiments is very high.⁷ Therefore it is not clear what the value is of this documentation of empirical regularities. If the aim is to make any claims about other regarding preferences in the world at large, the high artificiality appears to render the experimental results useless. If, in contrast, the aim is to document robust, causal laboratory effects to confront theorists with (cf. Bob Sugden's contribution to this issue), artificiality is less of a problem. This does have the danger of theorists and experimentalists creating their own world, however, as will be discussed below.

For gathering empirical regularities that aim at telling is something about behavior outside of the laboratory, it is important that the artificiality of the situation is considered. Much more than in case of testing theories, the data must be relevant for the situation one is interested in. Without theory to guide the experimenter, she must find guidance in the economic world she is trying to study. This might be a reason why external validity has received much more attention in psychology than in economics. To a large extent, psychological research is inductive and based on observed empirical regularities.

4.4 Advising policy makers

Experimental economists are very active in advising policy makers. A well-known example is the advice on auction designs to allocate spectrum across the world (Crampton 2002). At first sight, one would think that the external validity is of major importance for this type of experiment. However, many of these experiments are guided by theory. This is notably the case for auction experiments. In these cases, some of the arguments mentioned above for category 4.1 hold here. In fact, experimental evidence in support of auction theories has increased the role of economic theory in the design of auctions organized by governments.

However, many questions cannot be answered by theory. For example, in 2002 the Dutch government was planning to use an auction to allocate airwaves to be used by commercial FM radio stations. This auction would involve heterogeneous licenses, preferably to be auctioned simultaneously.

Moreover, because of a market structure involving a set of strong incumbents with a few weak potential newcomers, there are theoretical reasons to prefer an auction with a first-price element (Jehiel and Moldovanu 2002). However, a first-price simultaneous auction of heterogeneous goods is intractable from a theoretical point of view. The Dutch government therefore hired experimentalists to design the auction based on laboratory tests. In these tests, external validity was aimed at by incorporating essential features of the market concerned (such as a distinction between strong incumbents and weak newcomers) in the experimental design. For more details, see Goeree *et al.* (2005).

Artificiality is something that should obviously be avoided as much as possible, in this category of experiments. My impression is that this is indeed done. However, it is very difficult to judge *a priori* whether or not the external validity is high enough. Aside from introspection, one indication of its relevance might be in the demand for such research. In fact, policy-makers appear to find experimental research useful as is evident from the Dutch example. This external demand for laboratory experiments serves as a positive signal of their external validity.

5 CONCLUDING DISCUSSION

In spite of the fact that methodological discussions have only just started to take place amongst experimental economists, the state of affairs is quite positive. The widespread consensus on major issues has led to a situation where most experiments are used to test theories. The common opinion that the internal validity is more important than the external validity for these experiments is quite defensible. Where other types of experiments require more care with respect to their artificiality, this care is often taken.

Nevertheless, it is important that experimentalists pay more attention to such methodological issues. As the discipline develops, the relationship between experiments and theory will become more complicated. Increasingly, experimental results are being used to develop new theories. For example, the recent surge in theories on equity, fairness and reciprocity (e.g., Charness and Rabin 2002; Fehr and Fischbacher 2003) is highly motivated by experimental results documenting non-selfish behavior.

However, if the role of experiments shifts from testing theories to motivating the development of new theories, the 'mutual internal validity' of theory and experimental test has the danger of creating its own world. Theories aimed at describing laboratory behavior are becoming increasingly capable of doing so. Because their direct aim is no longer to describe behavior outside of the laboratory, and because the experiments are structured according to the theory, the connection between these two on the one hand and the outside-the-laboratory world on the other may be lost. Examples where this danger exists include the connection between recent

theories of reciprocity and gift exchange experiments⁸ and that between non-expected utility theories and experiments on individual choices between gambles.

When there is a repeated feedback between theory and experiments, reflecting on the external validity seems an important way to keep contact with the outside world. There are limits to what we can learn from running new experiments to test theories that were developed on the basis of experimental results. In this respect, the increased interest in methodology that experimentalists have shown over the past few years is very timely. What is still missing, however, is a careful analysis of possible new experimental methods that will help increase the external validity. A search for ways to modify laboratory experiments seems a fertile research area now that experiments are starting to be used for much more than just testing theories.

Arthur Schram
University of Amsterdam
Schram@uva.nl

ACKNOWLEDGEMENTS

I would like to thank Nick Bardsley, Klarita Gërxhani, Andreas Ortmann, Joep Sonnemans, Bob Sugden and Frans van Winden for useful comments.

NOTES

- 1 Guala (2002) defines: ‘an experimental result is *internally* valid, if the experimenter attributes the production of an effect *B* to the factor ... *A*, and *A* really is the ... cause of *B* in the experimental set-up *E* ... it is *externally* valid ... if *A* causes *B* not only in *E*, but also in a set of other circumstances of interest *F*, *G*, *H*, etc’.
- 2 Starmer refers mainly to experiments on individual decision-making, where alternatives to the standard theories have been developed, based on experimental results.
- 3 Even though he has experience with experiments, Loewenstein explicitly posits himself as an outsider (behavioral economist) looking at the field from the sideline. Implicitly, he defines ‘experimental economics’ as something narrower than ‘using experiments to study economics’. This distinction is not common. For example, leading behavioral economists serve on the editorial board of *Experimental Economics* and are on the board (and president) of the Economic Science Association.
- 4 A rough categorization of the 69 papers published in *Experimental Economics* up to and including volume 6, issue 1 gives 14 papers in a category ‘methodology and surveys’; 33 papers in ‘testing theories’; 14 papers in ‘stress tests of theories’; 7 papers in ‘searching for empirical regularities’ and 1 paper in ‘policy advice’.

- 5 For a detailed introductory description of double auction markets and typical results, see Davis and Holt (1993, ch. 3).
- 6 In a dictator game, players are in pairs. One player is given an endowment by the experimenter and is free to offer any amount of this to the other, which completes the game.
- 7 As one of my colleagues likes to point out: in reality people continuously have an opportunity to give money to strangers, but very few do. Especially the endowment given by the experimenter finds little parallel outside of the laboratory.
- 8 In the classic gift exchange experiment (Berg *et al.* 1995), subjects are paired. The first subject has an endowment that she can split between herself and player 2. The amount offered to player 2 is tripled by the experimenter, after which player 2 can return any amount to player 1. A typical result is that player 1 offers everything and player 2 at least returns 1's original endowment.

REFERENCES

- Berg, J., Dickhaut, J. and McCabe, K. (1995) 'Trust, reciprocity and social history', *Games and Economic Behavior* 10: 122–42.
- Binmore, K. (1994) *Playing Fair*, Cambridge, MA: MIT Press.
- Camerer, C.F. and Hogarth, R.M. (1999) 'The effect of financial incentives in experiments: a review and capital-labor-production framework', *Journal of Risk and Uncertainty* 19: 7–42.
- Charness, G. and Rabin, M. (2002) 'Understanding social preferences with simple tests', *The Quarterly Journal of Economics* 117: 817–69.
- Crampton, P. (2002) 'Spectrum auctions', in M. Cave, S. Majumdar and I. Vogelsang (eds) *Handbook of Telecommunications Economics*, Amsterdam: Elsevier Science, pp. 606–39.
- Cross, J. (1980) 'Some comments on the papers by Kagel and Battalio and by Smith', in J. Kmenta and J. Ramsey (eds) *Evaluation of Econometric Models*, New York: NYU Press, pp. 403–06.
- Davis, D.D. and Holt, C.A. (1993) *Experimental Economics*, Princeton, NJ: Princeton University Press.
- Fehr, E. and Fischbacher, U. (2003) 'The nature of human altruism', *Nature* 425: 785–91.
- Goeree, J.K., Offerman, T.J.S. and Schram, A.J.H.C. (2005) 'Using first-price auctions to sell heterogeneous licenses', Working Paper, University of Amsterdam.
- Guala, F. (2002) 'On the scope of experiments in economics: comments on Siakantaris', *Cambridge Journal of Economics* 26: 261–67.
- Harrison, G.W. and List, J.A. (2003) 'What constitutes a field experiment in economics?' Mimeo, University of Central Florida.
- Hertwig, R. and Ortmann, A. (2001) 'Experimental practices in economics; a challenge for psychologists?', *Behavioral and Brain Sciences* 24: 383–403.
- Holt, C.A. (1995) 'Industrial organization: a survey of laboratory research', in J. Kagel and A. Roth (eds) *The Handbook of Experimental Economics*, Princeton, NJ: Princeton University, pp. 349–444.
- Holt, C.A., Langan, L. and Villamil, A.P. (1986) 'Market power in oral double auctions', *Economic Inquiry* 24: 107–23.
- Jehiel, Ph. and Moldovanu, B. (2002) 'An economic perspective on auctions' Working Paper, Mannheim University.

- Loewenstein, G. (1999) 'Experimental economics from the vantage-point of behavioural economics', *The Economic Journal* 109: F25–F34.
- McDaniel, T. and Starmer, C. (1998) 'Experimental economics and deception: a comment', *Journal of Economic Psychology* 19: 403–09.
- Ortmann, A. (2003) 'Field experiments in economics; some methodological caveats' Mimeo, CERGE.
- Ortmann, A. and Hertwig, R. (2002) 'The costs of deception: evidence from psychology', *Experimental Economics* 5: 111–32.
- Plott, C.R. (1982) 'Industrial organization theory and experimental economics', *Journal of Economic Literature* 20: 1485–527.
- Roth, A.E. (1995a) 'Introduction to experimental economics', in J. Kagel and A. Roth (eds) *The Handbook of Experimental Economics*, Princeton, NJ: Princeton University, pp. 3–110.
- Roth, A.E. (1995b) 'Bargaining experiments', in J. Kagel and A. Roth (eds) *The Handbook of Experimental Economics*, Princeton, NJ: Princeton University, pp. 253–348.
- Smith, V.L. (1962) 'An experimental study of competitive market behavior', *Journal of Political Economy* 70: 111–37.
- Smith, V.L. (1976) 'Experimental economics: induced value theory', *American Economic Review* 66: 274–79.
- 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: NYU Press, pp. 345–77.
- Smith, V.L. (1982) 'Microeconomic systems as an experimental science', *American Economic Review* 72: 923–55.
- Smith, V.L. (1985) 'Experimental economics: reply', *American Economic Review* 75: 265–72.
- Smith, V.L. (1989) 'Theory, experiment and economics', *Journal of Economic Perspectives* 3: 151–69.
- Smith, V.L. (2002) 'Method in experiment: rhetoric and reality', *Experimental Economics* 5: 91–110.
- Smith, V.L. and Williams, A.W. (1990) 'The boundaries of competitive price theory: convergence, expectations and transaction costs', in L. Green and J. Kagel (eds) *Advances in Behavioral Economics*, vol. 2, Norwood, NJ: Ablex Publishing, pp. 31–53.
- Starmer, C. (1999a) 'Experimental economics: hard science or wasteful tinkering?', *The Economic Journal* 109: F5–F15.
- Starmer, C. (1999b) 'Experiments in economics: should we trust the dismal scientists in white coats?', *Journal of Economic Methodology* 6: 1–30.