Exp Econ
DOI 10.1007/s10683-008-9196-2

Peter Bohm: Father of field experiments

Martin Dufwenberg · Glenn W. Harrison

1 Introduction

Peter Bohm (1935–2005) was the father of modern field experiments. He clearly understood and stated the differences between laboratory experiments and experiments with field counterparts. His research was a clear precursor to the methodology that is now becoming widely accepted as complementary to laboratory experiments.

M. Dufwenberg (✉)
Department of Economics and Institute for Behavioral Economics, University of Arizona, Tucson, AZ 85721-0108, USA
e-mail: martind@eller.arizona.edu

G.W. Harrison
Department of Economics, College of Business Administration, University of Central Florida, Orlando, FL 32816-1400, USA
e-mail: gharrison@research.bus.ucf.edu

2 Personal background

Peter completed his undergraduate education and the equivalent of a doctoral degree at the University of Stockholm, and was there for most of his professional life. In 1964 and 1965 he visited Berkeley and Harvard, in 1978 visited Maryland and RFF, and in 1990 visited the World Bank. He was editor of what is now known as the *Scandinavian Journal of Economics* between 1969 and 1973, and was an Associate Editor of the *Journal of Public Economics* between 1985 and 1997.

Peter was a member of the Expert Group on Public Finance of the Swedish Ministry of Finance between 1985 and 1997. This independent think tank, under the Swedish Ministry of Finance, tried to formulate answers to economic questions which were deemed important to society. One of Peter's final publications was a report to this group (Bohm and Dufwenberg 2003). The topic: how to use experiments to inform and improve economic policy.

One position that Peter enjoyed was as a long-standing Judge in the Swedish Market Court between 1971 and 1994. This specialized court handles cases related to the Competition Act as well as cases involving the Marketing Act and other consumer and marketing legislation. Peter's involvement ranged from a major antitrust cartel suit to considerations of rules for regulating vacuum cleaner salesmen (his distaste for pressure sales tactics, applied to little old ladies, was intense).

3 General research interests

Peter was generally interested in public economics, with special concerns for cost-benefit analysis and environmental economics. His compilation of readings with Kneese (Bohm and Kneese 1971) was a staple of many undergraduate environmental economics classes around the world. He was also widely known for his concise introduction to welfare economics (Bohm 1974). Both books have gone through numerous reprintings, and even new editions.

One can often learn a lot about a major scholar from his the first academic publication, and Bohm (1967) is no exception. This mathematical examination of second-best theory results in some down-to-earth and non-mathematical recommendations. Peter's conclusions call into question the use of abstract "marginal rate" conditions as a general guide to optimal policy, and urges a more concrete focus on the political tools available (e.g. p. 314). The article anticipates what later emerged as hallmarks of the research Peter did and liked best: a desire to do improve welfare, and a keen sensitivity to the need to achieve such ends pragmatically.

4 Field experiments

Peter was drawn to conduct field experiments long before laboratory experiments had become a staple in the methodological arsenal of economists. Just as some experimentalists do not comprehend why one would ask questions with no real economic consequences, or care too much about the responses to such questions, Peter

began doing field experiments simply because they answered the questions he was interested in. He did not come to field experiments because of any frustration with lab experiments, or from any long methodological angst about laboratory experiments: it was just obvious to him that experiments needed field referents to be interesting. He later became interested in the methodological differences between laboratory and field experiments, well after his own pioneering contributions to the later had been published (Bohm 1972, 1984a). His later work deliberately examined the sources of differences one might find between the lab and field (Bohm and Lind 1993; Bohm 1994a, 1994b), and he thought explicitly about methodological differences between the two (Bohm 2002). Finally, he always had a clear "eye on the policy prize," in the sense that he saw experiments as a way to answer real policy questions (Bohm 2003).

While Peter cared about methodological differences between various forms of experiments his distinctions were practical in nature, and he would probably feel stifled by the taxonomy suggested by Harrison and List (2004) to structure the burgeoning literature. If a laboratory design seemed far removed from the naturally occurring circumstance that economists would care about, then Peter often would view it as an uninteresting or irrelevant experiment. He was also well aware of the vast literature on social experiments. He viewed the range of experiments one could conduct on a continuum, where some were too abstract to be interesting, and others were too complicated to provide insight into theory or even policy. He enjoyed talking about lab experiments much as a theorist should enjoy talking about how one might make a theoretical proposition operationally meaningful. Lab experiments forced one to be explicit about things, and that was a virtue. They also helped spell out the essential steps needed when one went to the field. And, unlike social experiments, they were designed to help test propositions of theory, not just to "estimate behavioral parameters" within the context of a particular social circumstance. This virtue Peter wanted to see extended in good field experiments, which thus would combine the best of both worlds. Thus he viewed "field experiments" as an attractive intersection of several tools of research that had been around for many decades.

To see how fundamental these contributions were, consider where they do fit into the taxonomy of field experiments proposed by Harrison and List (2004). They propose the following terminology: a *conventional lab experiment* is one that employs a standard subject pool of students, an abstract framing, and an imposed set of rules; an *artefactual field experiment* is the same as a conventional lab experiment but with a non-standard subject pool; a *framed field experiment* is the same as an artefactual field experiment but with field context in either the commodity, task, or information set that the subjects can use; and a *natural field experiment* is the same as a framed field experiment but where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are in an experiment.

Bohm (1972, 1979, 1984a, 1984b, 1994a, 1994b) repeatedly stressed the importance of recruiting subjects who have some field experience with the task or who have an interest in the particular task. His experiments generally involved imposing institutions on subjects who were not familiar with the institution, since the objective of the early experiments was to study new ways of overcoming free-rider bias. But his choice of commodity was usually driven by a desire to confront subjects with stakes

and consequences that are natural to them. In other words, his experiments illustrate how one can seek out subject pools for whom certain stakes are meaningful, in the sense that they had some experience and naturally-occurring, non-induced interest in valuing them. Thus the bulk of his research falls into the categories of artefactual or framed field experiments.

To illustrate, again with his first major publication in the area, consider Bohm (1972). This was a landmark study, actually conducted in the late 1960s, that had a great impact on many researchers in the areas of field public good valuation and experimentation on the extent of free-riding. The commodity was a closed-circuit broadcast of a new Swedish TV program. At the time (November 1969), Sweden only had one television station, and this was a much-anticipated program by two well-known comedians. Six elicitation procedures were used. In each case except one the good was produced, and the group was able to see the program, if aggregate WTP equaled or exceeded a known total cost. Every subject received SEK 50 upon arrival at the experiment, broken down into standard denominations.

Bohm (1972) employed five basic procedures for valuing his commodity.[1] In Procedure I the subject pays according to his stated willingness to pay (WTP). In Procedure II the subject pays some fraction of stated WTP, with the fraction determined equally for all in the group such that total costs are just covered (and the fraction is not greater than one). In Procedure III the payment scheme is unknown to subjects at the time of their bid. In Procedure IV each subject pays a fixed amount. In Procedure V the subject pays nothing.

No formal theory was provided to generate free-riding hypotheses for these procedures. Procedure I was deemed the most likely to generate strategic *under*-bidding (p. 113), and Procedure V the most likely to generate strategic *over*-bidding. The other procedures were thought to lie somewhere in between these two extremes. Explicit cheap talk admonitions *against* strategic bidding were given to subjects in Procedures I, II, IV and V, in order "for the test to be relevant for situations of real referenda" where "there would definitely be discussions about the advantages to the individual of understating his willingness to pay as well about 'moral obligations' to tell the truth, etc." (p. 117; cf. pp. 115, 119, 128–129).

The major conclusion from Bohm (1972) was that bids were virtually identical for all institutions, averaging between SEK 7.29 and SEK 10.33. These results have been critically evaluated elsewhere (Cummings and Harrison 1994), and that evaluation characteristically challenged in some respects by Bohm (1994b). The main qualitative conclusion was that free-riding was not the behavioral absolute that theory had led economists to simply assume, and that result comes through irrespective of debates over the quantitative significance of differences in the levels of contributions across institutions.

The use of multiple evaluation instruments in Bohm (1972) became later formalized in what is now known as the Bohm Interval Method for evaluating public goods (Bohm 1984a). The idea is simple: use a simple, transparent method that *a priori* generates over-valuation, and another simple, transparent method that *a priori* generates

[1] A sixth procedure was also employed, and is discussed in Cummings and Harrison (1994).

under-valuation. If the two methods yield valuations that exceed some cost of undertaking the project, you can go ahead and say that benefits exceed costs, even if you do not know the true level benefits. Or if both methods yield valuations that fell below the cost, you can also go ahead and say that the project should not be implemented. If the benefit estimates straddled the true costs, you needed more precise valuation methods or other criteria to decide if the project should go ahead. Peter was well aware of the theoretical research by Groves and Ledyard into incentive-compatible mechanisms for eliciting valuations for public goods, as well as the initial forays of Vernon Smith (1979) into the use of neo-Wicksellian mechanisms for providing public goods in experiments. But he openly stressed the need for "practicable" mechanisms in field applications (Bohm 1984b).

The pragmatic insight of the Bohm Interval Method remains strikingly relevant to continuing debates over the validity of contingent valuation methods. As stressed by Peter and many others, these methods generate hypothetical responses which will typically be significant over-statements of true valuations (Bohm 1994b). But if those hypothetical responses are well below expected costs, you have the answer you need to make a reliable policy conclusion.

The policy relevance of this insight is even more general than the validity of using hypothetical surveys, and extends to any setting in which there is considerable uncertainty about the true estimates of benefits and costs. Peter advocated a simple approach towards cost-benefit analysis when there is some uncertainty about costs, benefits or both: see if one can still place bounds on each such that you make the right binary conclusion as to whether net benefits are positive or negative, and that is all that matters. In the most general form, one would end up with statements about the *probability* of net benefits being positive or negative, but one could sensibly just look at extreme bounds or, say, 95% confidence intervals. Peter's point is that such binary inferences are still extremely valuable to public decision-makers, even if the analyst was numerically imprecise about the size of the net benefit. Harrison and Kriström (1998) illustrate this approach in an evaluation of the benefits and costs of Sweden imposing *unilateral* carbon taxes to mitigate global warming, concluding that:

> "The conclusion is clear. The benefits of increasing the carbon tax in Sweden are a tiny fraction of the 'price tag' which Swedes must pay in the form of higher prices and reduced incomes. Although we do not put much credence in these gross benefit numbers, they do serve to highlight the basis of our conclusion that carbon tax increases are not currently justifiable in Sweden. They also serve to focus the debate on the net benefits of further carbon taxes onto the question of estimating gross benefits for Swedes. *If these numbers are correct*, then advocates of carbon tax increases are telling the average Swede that he or she must pay a lot more for some environmental good than that Swede appears to derive as a benefit. This might be because the *advocate* derives significant enough benefits and would be willing to pay the price tag, but that does not justify foisting the price on others. (p. 83)"

This was a policy application that Peter was directly involved in, as an advisor to a Swedish Government Tax Exchange Commission that commissioned this evaluation. It was also one that he cared passionately about, as his last major area of research

focused on the design of field experiments to evaluate how one might come to an international agreement to mitigate global warming (Bohm and Carlén 2002). He would have been delighted to have seen the application of his Interval method in the overall cost-benefit logic of the *Stern Review of the Economics of Global Warming* (Stern 2006).

One final example of Peter's contribution to field experimental methodology came about by accident, in an experiment he regarded as a failure. But what it shows is critical for the central difference between field and lab experiments, the natural context of the field. It shows dramatically what can happen when the subjects perceive a meta-game beyond the experiment itself.

In 1980 he undertook a field experiment for a local government in Stockholm that was considering expanding a bus route to a major hospital and a factory. The experiment was to elicit valuations from people who were naturally affected by this route, and to test whether their aggregate contributions would make it worthwhile to provide the service. A key feature of the experiment was that the subjects would have to be willing to pay for the public good if it was to be provided for a trial period of 6 months. Everyone who was likely to contribute was given information on the experiment, but when it came time for the experiment virtually nobody turned up! The reason was that the local trade unions had decided to boycott the experiment, since it represented a threat to the current way in which such services were provided. The union leaders expressed their concerns, summarized by Bohm (1984b, p. 136) as follows:

> "They reported that they had held meetings of their own and had decided (1) that they did not accept the local government's decision not to provide them with regular bus service on regular terms; (2) that they did not accept the idea of having to pay in a way that differs from the way that "everybody else" pays (bus service is subsidized in the area)—the implication being that they would rather go without this bus service, even if their members felt it would be worth the costs; (3) that they would not like to help in realizing an arrangement that might reduce the level of public services provided free or at low costs. It was argued that such an arrangement, if accepted here, could spread to other parts of the public sector; and (4) on these grounds, they advised their union members to abstain from participating in the project."

This fascinating outcome is actually more relevant for experimental economics in general than it might seem. When certain institutions are imposed on subjects, and certain outcomes tabulated, it does not follow that the outcomes of interest for the experimenter are the ones that are of interest to the subject. And, most critically, running field experiments forces one to be aware of the manner in which subjects select themselves into tasks based on their beliefs about the outcomes. This process might be a direct social choice over institutions or rules, it might be Tiebout-like migration, it might be a literal or behavioral rejection of the task, it might be literal or behavioral attrition once the task is understood, it might be the evolution of social norms to resolve implicit coordination problems, or it might be some combination of these. This is an active and exciting area of research in laboratory experiments now (e.g., the literature cited in Botelho et al. 2005), and one that draws on insights

from field experiments such as those conducted by Peter. The point is that we design better lab experiments when we worry about what one just cannot ignore in the field experiment.

5 Conclusion

Peter learned to take the long view when it came to influencing the way economists did experiments, and how they viewed experiments in the broader context of actual policy evaluations. His work on field experiments did not generate a flurry of work when it first came out, but was instead regarded as a striking contribution to the debates over the external validity of laboratory experiments.[2] He later grew concerned at the manner in which behaviorists were using results from narrow, artefactual laboratory experiments to make global pronouncements about the health of mainstream economic theory, ignoring what we now see as the ecological validity of focusing on behavior in field experiments. He did not succeed in "marketing" his approach, perhaps due to an impatience with what he viewed as overly parametric theorizing, as well as a lusty taste for picking fights. But his legacy is clear from published work, it is now entrenched in the practices of field experimenters, and need not be forgotten.

Acknowledgement We are grateful to Astri Muren for comments.

References

Bohm, P. (1967). On the theory of second best. *Review of Economic Studies*, *34*, 301–314.
Bohm, P. (1972). Estimating the demand for public goods: an experiment. *European Economic Review*, *3*, 111–130.
Bohm, P. (1974). *Social efficiency: a concise introduction to welfare economics*. London: Macmillan & Co.
Bohm, P. (1979). Estimating willingness to pay: why and how? *Scandinavian Journal of Economics*, *81*, 142–153.
Bohm, P. (1984a). Revealing demand for an actual public good. *Journal of Public Economics*, *24*, 135–151.
Bohm, P. (1984b). Are there practicable demand-revealing mechanisms? In H. Hanusch (Ed.), *Public finance and the quest for efficiency*. Detroit: Wayne State University Press.
Bohm, P. (1994a). Behavior under uncertainty without preference reversal: a field experiment. In J. Hey (Ed.), *Experimental economics*. Heidelberg: Physica.

[2] Peter never liked this way of characterizing his work, and was justifiably concerned with the casual manner in which commentators used the expression "external validity." The term normally refers to the validity of applying the inferences drawn from the lab to field behavior. Nevertheless, from a scientific viewpoint, the validity of experimental inferences must first and foremost depend on the theoretical framework that is being used to draw inferences from the observed behavior in the experiment. If we have a theory that (implicitly) says that hair color does not affect behavior, then any experiment that ignores hair color is valid from the perspective of that theory. On the other hand, if our experiment confirms casual field observations that hair color does matter for behavior, but we have no theory that logically predicts any causation, how can we claim that the experiment has external validity? In order to identify what factors make an experiment valid we have to possess some priors from a theoretical framework, which is crossing into the turf of "internal validity." The critical role of non-parametric priors from theory, in his Interval Method, illustrates the point perfectly. Thus, he did not see external and internal validity as independent.

Bohm, P. (1994b). CVM spells responses to hypothetical questions. *Natural Resources Journal*, *34*, 37–50.
Bohm, P. (2002). Pitfalls in experimental economics. In F. Andersson & H. Holm (Eds.), *Experimental economics: financial markets, auctions, and decision making*. Dordrecht: Kluwer Academic.
Bohm, P. (2003). Experimental evaluations of policy instruments. In K. G. Mäler & J. R. Vincent (Eds.), *Handbook of environmental economics* (Vol. 1, pp. 438–460). Amsterdam: Elsevier.
Bohm, P., & Carlén, B. (2002). Cost-effective approaches to attracting low-income countries to international emissions trading: theory and experiments. *Environmental and Resource Economics*, *23*, 187–211.
Bohm, P., & Dufwenberg, M. (2003). *Politik på prov—en ESO-rapport om experimentell ekonomi*. Ds 2003:31, Swedish Ministry of Finance.
Bohm, P., & Kneese, A. V. (Eds.). (1971). *The economics of environment*. London: Macmillan & Co.
Bohm, P., & Lind, H. (1993). Preference reversal, real-world lotteries, and lottery-interested subjects. *Journal of Economic Behavior and Organization*, *22*, 327–348.
Botelho, A., Harrison, G. W., Pinto, L. M. C., & Rutström, E. E. (2005). *Social norms and social choice* (Working Paper 05-23). Department of Economics, College of Business Administration, University of Central Florida.
Cummings, R. G., & Harrison, G. W. (1994). Was the *Ohio* court well informed in their assessment of the accuracy of the contingent valuation method? *Natural Resources Journal*, *34*, 1–36.
Harrison, G. W., & Kriström, B. (1998). General equilibrium effects of increasing carbon taxes in Sweden. In R. Brännlund & I. Gren (Eds.), *Green taxes: theory and practice*. London: Elgar.
Harrison, G. W., & List, J. A. (2004). Field experiments. *Journal of Economic Literature*, *42*, 1013–1059.
Smith, V. L. (1979). Incentive compatible experimental processes for the provision of public goods. In V. L. Smith (Ed.), *Research in experimental economics* (pp. 59–168). Greenwich, CT: JAI Press.
Stern, N. (2006). *The Stern review of the economics of global warming*. Available from web site at http://www.hm-treasury.gov.uk/independent_reviews/stern_review_economics_climate_change/stern_review_report.cfm.