UCD GEARY INSTITUTE DISCUSSION PAPER SERIES

Lab Experiments are a Major Source of Knowledge in the Social Sciences

Lab Experiments are a Major Source of Knowledge in the Social Sciences

Armin Falk

University of Bonn

CEPR, CESifo, IZA

James J. Heckman

University of Chicago

University College Dublin

Cowles Foundation, Yale University

Friday, October 30, 2009

Abstract. Laboratory experiments are a widely used methodology for advancing causal knowledge in the physical and life sciences. With the exception of psychology, the adoption of laboratory experiments has been much slower in the social sciences, although during the last two decades, the use of lab experiments has accelerated. Nonetheless, there remains considerable resistance among social scientists who argue that lab experiments lack "realism" and "generalizability". In this article we discuss the advantages and limitations of laboratory social science experiments by comparing them to research based on non-experimental data and to field experiments. We argue that many recent objections against lab experiments are misguided and that even more lab experiments should be conducted.

The social sciences have generally been less willing to use laboratory experiments than the natural sciences, and empirical social science has traditionally been considered as largely non-experimental, that is, based on observations collected in naturally occurring situations. The first lab experiments in economics were not conducted until the late 1940s. Fewer than 10 experimental papers per year were published before 1965, and about 30 per year by 1975 (*1, 2*). Starting from this low level, experimentation in economics greatly increased in the mid 1980s. In three well-known economics journals—the *American Economic Review*, *Econometrica* and the *Quarterly Journal of Economics*—the fraction of laboratory experimental papers in relation to all published papers was between 0.84 percent and 1.58 percent in the 1980s, between 3.06 percent and 3.32 percent in the 1990s and between 3.8 percent and 4.15 percent between 2000 and 2008 (authors' calculations). The percentages were much higher in more specialized economics journals. The first specialty journal, *Experimental Economics*, was founded in 1998. A similar increase in lab experiments has taken place in other social sciences as well, e.g., in political science *(3)*.

Many social scientists are still reluctant to rely on laboratory evidence. Common objections are that student subject pools are unrepresentative and that sample sizes are small. There is also a widespread view that the lab produces "unrealistic" data, which lacks relevance for understanding the "real world". This notion is based on an implicit hierarchy in terms of generating relevant data, with field data being superior to lab data. We argue that this view, despite its intuitive appeal, is based on a misunderstanding of the nature of evidence in science and of the kind of data collected in the lab. We also argue that many of the objections against evidence from the lab suggest the wisdom of conducting *more* lab experiments, not fewer. While most of our examples and topics are taken from economics, the methodological points we discuss can be applied to all social sciences.

The lab provides controlled variation

Controlled variation is the foundation of empirical scientific knowledge. The laboratory allows tight control of decision environments. As an illustration, consider a simple experiment, the gift exchange game, which tests the theory that employment relationships are governed by a gift exchange, i.e., that workers reciprocate "fair" wages with high effort. A positive relationship between wages and effort is the central assumption of efficiency wage theories that have important implications for the functioning of labor markets and that can explain rigid wages and involuntary unemployment *(4)*. Testing this class of theories with field data is notoriously difficult. For example, in firms, worker effort is not easily observed

or measured, and workers are confronted with a mix of different incentives. This makes interpretation of different effort levels difficult. An observed variation in wages may not reflect generosity but may be due to firm size, self-selection of workers, or simply productivity differences. Even if a relationship between wages and effort is detected, it may not necessarily reflect a fair-wage effort relationship; instead it could reflect strategic considerations based on reputation and repeated interactions. In the laboratory, these factors can be varied in a controlled fashion. The experimenter observes effort and wages and can rule out confounding effects such as multiple incentives, selection, productivity differences, or repeated interactions.

The first experimental test of the existence of gift exchanges in the framework of a formal game-theoretic model was designed to mimic an employment relationship. Participants assumed the roles of workers and firms (*5*). Firms made (binding) wage offers, which workers could accept. If a worker accepted, he or she then had to choose a costly effort level. Labor contracts are generally incomplete contracts, i.e., effort is not fully contractually enforceable. In the experiment this is reflected by the fact that workers were free to choose any effort level above the contractually enforceable level. In this framework, it is possible to test the gift-exchange hypothesis against the self-interest assumption commonly made in economics that a self-interested worker would always choose the lowest possible effort, since effort is costly and there is no punishment for minimal effort. Anticipating this, the firm has no incentive to pay an above-minimum wage, because self-interested workers work no harder if given a higher than minimum wage. Nevertheless, the results of numerous gift-exchange experiments in the lab reveal that higher wages induce workers to provide higher effort levels. The experiment is a good example of the many experiments that have challenged the assumption of a universally selfish and rational *homo economicus*. Systematic lab evidence shows that people are boundedly rational, prone to behavior such as loss aversion, present-bias, or judgment biases (*6*). Phenomena such as reciprocity or social approval, which have been largely neglected by mainstream economics, have been shown to be important in affecting economic outcomes in bargaining and market interactions (*7,8*).

These experiments illustrate that the lab offers possibilities to control decision environments in ways that are hard to duplicate using naturally occurring settings. In the laboratory, the experimenter knows and controls the material payoffs, the order in which the different parties can act, and the information they possess when they make choices and whether the game is repeated or one-shot. This control allows for the testing of precise predictions derived from game-theoretic models. Subjects are randomly assigned and decisions are rewarded. Payment ensures that subjects take their decisions seriously. For

example, if a firm pays a higher wage or a subject provides higher effort, costs are higher and final earnings are lower. In this sense behavior in the laboratory is reliable and "real": subjects in the lab are human beings who perceive their behavior as relevant, experience real emotions, and take decisions with real economic consequences *(6, 9-12)*. Lab experiments can be used for testing theories and to study institutions at relatively low cost (*9*). This is particularly interesting for policy questions where the proposed program intervention has no counterpart in reality, and where constructing the counterfactual states of interest may be done more easily in the lab than in the field. Moreover, while existing institutions are adopted endogenously, rendering causal inferences about their effects difficult, the lab allows exogenous changes in institutions. Lab experiments have turned out to be valuable in solving practical problems that arise in implementing matching markets (*13*), government regulation (*14*) or airport time slot allocation (*15*). "Economic engineering", a combination of theory and experiments, has improved design and functioning of many markets *(16)*.

Lab or field is not the choice

Resistance concerning laboratory evidence, often centers on an appeal to "realism". This skepticism has recently manifested itself in a lively debate in economics about field versus lab experiments. While some scholars have argued passionately in favor of laboratory experiments where controlled manipulations of conditions on carefully documented populations are performed (*17-19*), others have argued in favor of field experiments where conditions are "more realistic" though perhaps less tightly controlled and where more realism also implies greater relevance to policy *(3, 20-22)*. These controversies appear throughout the social sciences *(3, 23-26)*.

The casual reader may mistakenly interpret arguments about "realism" as an effective critique against the lab, potentially discouraging lab experimentation, and slowing down the production of knowledge in economics and other social sciences. The issue of "realism", however, is not a distinctive feature of lab vs. field data. The real issue is determining the best way to isolate the causal effect of interest. To illustrate this point and to structure the debate about field and lab data, we suggest the following simple model.

Consider an outcome of interest Y (e.g., effort supplied by a worker) and a list of determinants of Y, $(X_1,..., X_N)$. For specificity, suppose that

$$Y = f(X_1,...,X_N), \tag{1}$$

which is sometimes called an "all causes model" (*27,28*) because it captures all possible causes of Y in $(X_1,...,X_N)$. The causal effect of X_1 on Y is the effect of varying X_1 holding

fixed $\tilde{X} = (X_2, \ldots, X_N)$. In the pioneering field experiments ([20]), X_1 was the tax rate on wages. In the laboratory gift-exchange experiment, the values of X_1 are the different wage levels paid by firms. Unless f is separable in X_1, so that

$$Y = \phi(X_1) + g(\tilde{X}), \tag{2}$$

the level of Y response to X_1 will depend on the level of $\tilde{X}$. Even in the separable case, unless $\phi(X_1)$ is a linear function of X_1, the causal effect of X_1 depends on the level of X_1 and the size of the variation of X_1. These problems appear in both field and lab experiments, and in any estimation of the causal effect of X_1.

Among the $\tilde{X}$ in the gift exchange experiments described above are concrete details of market institutions such as the number of firms and workers, the order of moves, the choice set, payoff functions, information available, whether or not interactions are one-shot, and whether or not they are anonymous. More generally, $\tilde{X}$ could be demographic characteristics of the participants, the level of observation of actions by third parties, individual preference parameters (e.g., morality, persistence, self-control, social and peer influences) and other aspects of environments.

Many laboratory experiments like the gift-exchange experiment have provided evidence of gift-exchange and social preferences in lab settings for certain values of $\tilde{X}$, usually, but not always, using populations of undergraduates, and different bargaining and market institutions ([7, 30]). The relevance of these findings has been questioned in recent field experiments analyzing behavior in a population of sports card traders in a "natural setting", i.e., for another set of conditions $\tilde{X}'$ including, e.g., different institutional details, payoffs, and a different participant population ([29]). In this particular market, the evidence for social preferences is weaker. If one is interested in the effect of social preferences under a third condition ($\tilde{X}''$), neither the undergraduate nor the sports cards field study may identify the effect of interest. It is not obvious whether the lab $\tilde{X}$ or the field $\tilde{X}'$ is more informative for the third condition unless a more tightly specified economic model is postulated or a more precisely formulated policy problem is specified. Recent evidence suggests that empirical support for the existence of social preferences such as reciprocity or gift-exchange is not a matter of lab or field, but of the prevailing conditions such as the details of agent interactions ([31-33]). When the exact question being addressed and the population being studied are mirrored in an experiment, the information from it can be clear and informative. Otherwise, to transport experimental findings to new populations or new environments requires a model ([34]).

Field methods are able to obtain a universally defined causal effect only if the special functional form (2) is specified and the response of Y to X_1 is linear. If this is the case, however, lab experiments are equally able to obtain accurate inferences about universal effects. Observing behavior in the field is in general not more interesting or informative than observing behavior in the lab. The general quest for running experiments in the field to obtain more realistic data is therefore misguided. In fact, the key issue is what is the best way to isolate the effect of X_1 holding constant $\tilde{X}$. No greater reality is conveyed by one set of $\tilde{X}$ than another unless the proposed use of an estimate, as well as target populations and settings are carefully specified. The pioneering field experiments (20) defined the target population and the questions sought to be answered very precisely.

The usefulness of particular methods and data is ultimately a matter of the underlying research question. Lab experiments are very powerful whenever tight control of $\tilde{X}$ is essential. This applies in particular for testing (game theoretic) models and behavioral assumptions. The lab can also easily implement many different values of $\tilde{X}$, e.g., the number of buyers or sellers in market experiments, and in this way explore the issue of robustness of an estimated effect. Tight control of $\tilde{X}$ also allows replicability of results, which is generally more difficult with field data. The field on the other hand offers a large range of variations in $\tilde{X}$, which are potentially relevant but hard to implement in the lab. This way field experiments can provide important complementary insights to lab findings, e.g., in the area of development economics, or by addressing specific policy questions such as testing antipoverty programs that are targeted to, and need to be evaluated on, populations in poverty (20).

Other objections

In addition to the "lack of realism" critique, other objections concerning lab evidence have been put forward. Ironically, most objections raise questions that can be very well analyzed with lab experiments, suggesting the wisdom of conducting more lab experiments, not fewer. One common objection is that lab experiments with students do not produce "representative" evidence. For the purpose of testing theories this is not a problem because most economic models derive predictions that are independent of assumptions concerning subject pools. On some aspects of behavior, however, students are not representative of the general population or a target population of interest and we would agree that a richer variation in context, populations and environments $\tilde{X}$ should be used in future lab experiments. In this vein, the gift-exchange game has been run on non-student samples (35) yielding results similar to

6

those obtained using samples of students. Other studies have shown the relevance of social preferences for CEOs (*36*) for professional financial traders (*37,38*) or for the general population (*39,40*).

It has also been noted that (i) stakes in experiments (money paid for decisions taken) are trivial, (ii) the number of subjects or observations is too small, (iii) subjects are inexperienced, (iv) Hawthorne effects may distort experiments, or (v) self-selection into experiments may bias results.

(i) Most of what we know about the level of stakes on outcomes is derived from controlled lab experiments. The effects of varying stake size are mixed and seem to depend on concrete experimental contexts (*41*). Reciprocity does not vanish if subjects in the gift-exchange experiment reported above are paid an equivalent of three months income (*42*). Even if stake effects are relevant, however, it is not obvious what the "right" level of stakes should be, that is, what are the right levels of X_1 and $\tilde{X}$? We would ask in reply how often do people make decisions involving monthly incomes, and how representative would such high-stake experiments be for the many decisions people make on a daily basis, which involve relatively small stakes? In any case, if one is seriously interested in how stakes affect behavior, one can run experiments with varying stake sizes.

(ii) The issue of sample size is a red herring. Effective methods have been developed for analyzing small sample experiments (*43-45*). Moreover, many experiments nowadays are run with samples of several hundred participants, sometimes with more than 1000 participants (*46*).

(iii) Experiments do not typically distinguish between experienced and inexperienced subjects. It is an empirically interesting question how experience, learning etc. affect behaviors. Failure to account for experience in some experiments is not an intrinsic weakness of the experimental method. In fact, it is common to run experiments with experienced subjects and to study learning effects. Good examples are a study on ratchet effects and incentives with Chinese managers (*47*) or a study comparing the behavior of workers and students (*48*). In a recent field experiment on gift-exchange, it was shown, for instance, that experienced donors, that is, donors who frequently donate for a particular charitable organization, reciprocally respond to gifts by donating more frequently (*32*).

(iv) Another concern often raised is scrutiny, that is, the possibility that subjects in the lab behave differently because they perceive that they are observed. This is one version of a Hawthorne effect. (Parenthetically, reanalysis of the original Hawthorne data shows that no Hawthorne effect was present in the Hawthorne study. See (*49*).) It is a minor problem in many experiments, especially if the decision environment is interactive and "rich", such as in

sequential bargaining or market experiments. Moreover, being observed is not an exclusive feature of the laboratory: many decisions outside the lab are observed. Even on the Internet agents can be observed. In the lab, observers can be added to the experimental protocol. The lab allows the analyst to study the relevance of scrutiny by varying the degree of anonymity, e.g., contrasting video experiments where subjects are explicitly observed, with single anonymous (subject-subject anonymity) and double anonymous (full anonymity between subjects and experimenter) procedures *(50-52)*.

(v) Many scholars have expressed concerns about self-selection of particular participants into experiments. Self-selection is not necessarily a scourge. It can be a source of information on agent preferences *(27,28,34)*. In the lab, one can collect detailed data on the backgrounds and personality traits of participants to control for selection or to explicitly study selection in a controlled way *(53,54)*. Selection is a feature of both field and social experiments *(55)* and is not a problem unique to lab experiments. Indeed problems of noncompliance, attrition and randomization bias plague many field experiments *(27, 56)*.

Exploiting complementarities

Experiments can be productive in complementing the information obtained from other empirical methods. One can combine lab and field experiments to better understand the mechanisms observed in the field. For example, this can be done by eliciting preferences and relating these preferences to observed behavior in the field *(57-58)*. Another example of exploiting complementarities is the experimental validation of survey instruments *(59)*. While surveys can generate large and representative data sets that provide statistical power, experiments allow the elicitation of preferences and attitudes in a controlled and incentive compatible way, as participants have to make choices with real money at stake. Such evidence is particularly important in securing better understanding of preference heterogeneity *(60-63)*. The evidence that people are different clashes sharply with the widely used "representative agent" model that assumes that agents are homogenous or can be represented as if they are homogenous. Accounting for heterogeneity in preference parameters enables macroeconomists to calibrate economic models in an empirically founded way *(61)*.

We conclude by restating our argument. Causal knowledge requires controlled variation. In recent years, social scientists have hotly debated which form of controlled variation is most informative. This discussion is fruitful and will continue. In this context it is important to acknowledge that empirical methods and data sources are complements, not

substitutes. Field data, survey data, and experiments, both lab and field, as well as standard econometric methods can all improve the state of knowledge in the social sciences. There is no hierarchy among these methods and the issue of generalizability of results is universal to all of them.

References

1. C.A. Holt, *Markets, Games, and Strategic Behavior* (Pearson Addison Wesley, 2006).

2. *An illustration of the view that economics is a non-experimental science is a quote from Samuelson and Nordhaus who in their famous economics textbook stated that economists "cannot perform the controlled experiments of chemists or biologists because they cannot easily control other important factors. Like astronomers or meteorologists, they generally must be content largely to observe"* (P.A. Samuelson, W.D. Nordhaus, *Economics,* McGraw Hill, 1985, 12. ed., p. 8).

3. R. B. Morton, K. C. Williams, *The Oxford Handbook of Political Methodology* (Oxford University Press, 2006).

4. G.A. Akerlof, J.L. Yellen, *Quarterly Journal of Economics* **105**, 255 (1990).

5. E. Fehr, G. Kirchsteiger, A. Riedl, *Quarterly Journal of Economics* **108**, 437 (1993).

6. J. H. Kagel, A. E. Roth, *The Handbook of Experimental Economics*. (Princeton University Press, Princeton, NJ, 1995).

7. E. Fehr, S. Gächter, *Journal of Economic Perspectives* **14**, 159 (2000).

8. *A new development in uncovering human motivation is the field of experimental neuroeconomics. The goal is to provide insights into the biological foundations of behavior in order to improve economic modeling*. (C. Camerer, G. Loewenstein, D. Prelec, *Journal of Economic Literature* **43**, 9 (2005)).

9. D. Friedman, S. Sunder, *Experimental Methods: A Primer for Economists*. (Cambridge University Press, Cambridge, MA, 1994).

10. E. Reuben, F. van Winden, *Journal of Public Economics* **92**, 34 (2008).

11. D.J.F. De Quervain *et al.*, *Science* **305**, 1254 (2004).

12. E. Xiao, D. Houser, *Proceedings of the National Academy of Sciences* **102**, 7398 (2005).

13. J. H. Kagel, A. E. Roth, *Quarterly Journal of Economics* **115**, 201 (2000).

14. C. Plott, in *Laboratory Experimentation in Economics*, A. E. Roth, Ed. (Cambridge University Press, Cambridge, 1987), pp. 193-219.

15. S.J. Rassenti, V.L. Smith, R.L. Bulfin, *The Bell Journal of Economics* **13**, 402 (1982).

16. A.E. Roth, *Econometrica* **70**, 1341 (2002).

17. V.L. Smith, *Journal of Economic Perspectives* **8,** 113 (1994).

18. C. Plott, *Science* **232**, 4751 (1986).

19. V. L. Smith, *American Economic Review* **93**, 465 (2003).

20. G. G. Cain, H. W. Watts, *Income Maintenance and Labor Supply: Econometric Studies*. (Rand McNally College Publishing Company, Chicago, 1973), pp. 1-367.

21. S. D. Levitt, J. A. List, *Journal of Economic Perspectives* **21**, 153 (2007).

22. S.D. Levitt, J.A. List, *European Economic Review* **53**, 1 (2009).

23. A. Aronson, T.D. Wilson, M.B. Brewer, in *The handbook of social psychology*, D. Gilbert, S. T. Fiske, G. Lindzey, Eds. (McGraw-Hill, Boston) **1**, 99 (1998).

24. M. Webster, J. Sell, *Laboratory Experiments in the Social Sciences* (Academic Press, Oxford) (2007).

25. D. G. Mook, *American Psychologist* **38**, 4 (1983).

26. E. Ostrom, *Journal of Economic Behavior & Organization* **61**, 149 (2006).

27. J. J. Heckman, E. J. Vytlacil, in *Handbook of Econometrics,* J. Heckman, E. Leamer, Eds. (Elsevier, Amsterdam, 2007), vol. 6B, pp. 4779-4874.

28. J. J. Heckman, *International Statistical Review* **76**, 1 (2008).

29. J. A. List, *Journal of Political Economy* **114**, 1 (2006).

30. E. Fehr, U. Fischbacher, *The Economic Journal* **112**, C1 (2002).

31. S. Kube, C. Puppe, M.A. Maréchal, "The Currency of Reciprocity - Cash, Perks and Workers' Motivation in the Field", *Working Paper,* Max Planck Institute for Research on Collective Goods, Bonn.

32. A. Falk, *Econometrica* **75,** 1501, (2007).

33. S. DellaVigna, *Journal of Economic Literature* **47**, 315 (2009).

34. J.J. Heckman, in *Evaluating Welfare and Training Programs*, C. Manski, I. Garfinkel, Eds. (Harvard University Press, Cambridge, MA, 1992), pp. 201-230.

35. E. Fehr, E. Kirchler, A. Weichbold, S. Gächter, *Journal of Labor Economics* **16**, 324 (1998).

36. E. Fehr, J. A. List, *Journal of the European Economic Association* **2**, 743 (2004).

37. J. H. Kagel, D. Levin, *American Economic Review* **76**, 894 (1986).

38. V.L. Smith, G.L. Suchanek, A.W. Williams, *Econometrica* **56**, 1119 (1988).

39. T. Dohmen, A. Falk, D. Huffman, U. Sunde *Econ. J.* **119**, 592 (2009).

40. C. Bellemare, S. Kroeger, A. van Soest, *Econometrica* **76**, 815 (2008).

41. C.F. Camerer, R. M. Hogarth, *Journal of Risk and Uncertainty* **19**, 7 (1999).

42.   E. Fehr, U. Fischbacher, E. Tougareva, "Do High Stakes and Competition Undermine Fairness? Evidence from Russia", *Working Paper No.120*, University of Zurich.

43.   J. J. Heckman, R. Pinto, A. M. Shaikh, A. Yavitz, "Compromised Randomization and Uncertainty of Treatment Assignments in Social Experiments: The Case of the Perry Preschool Program", *Working Paper*, University of Chicago (2009).

44.   J. J. Heckman, S. H. Moon, R. Pinto, P. A. Savelyev, A. Yavitz, "A reanalysis of the HighScope Perry Preschool Program", *Working Paper*, University of Chicago (2009).

45.   S. Siegel and N. J. Castellan, *Nonparametric Statistics for The Behavioral Sciences*, (McGraw-Hill, London, 1988).

46.   B. Herrmann, C. Thöni, S. Gächter, *Science* **319**, 1362 (2008).

47.   D.J. Cooper, J.H. Kagel, W. Lo, Q.L. Gu, *American Economic Review* **89**, 781 (1999).

48.   J. Carpenter, S. Burks, E. Verhoogen, in *Field Experiments in Economics, Research in Experimental Economics*, J. Carpenter, G. Harrison, J. List, Eds. (Elsevier, Greenwich, Conn. and London), 261 (2005).

49.   S. R. G. Jones *American Journal of Sociology* **98,** 451 (1992).

50.   G. E. Bolton, E. Katok, R. Zwick, *International Journal of Game Theory* **27**, 269 (1998).

51.   S. K. Laury, J. M. Walker, A. W. Williams, *Journal of Economic Behavior and Organization* **27**, 365 (1995).

52.   E. Hoffman, et al., *Games and Economic Behavior* **7**, 346 (1994).

53.   C. Camerer, D. Lovallo, *American Economic Review* **89**, 306 (1999).

54.   T. Dohmen, A. Falk, "Performance Pay and Multidimensional Sorting: Productivity, Preferences and Gender", *IZA Discussion Paper 2001* (2006).

55.   J. J. Heckman, R. J. LaLonde, J. A. Smith, in *Handbook of Labor Economics,* O. Ashenfelter, D. Card, Eds. (North-Holland, New York), **3A**, 1865 (1999).

56.   J. J. Heckman, J. A. Smith, *Journal of Economic Perspectives* **9**, 85 (Spring, 1995).

57.   D. S. Karlan, *American Economic Review* **95**, 1688 (2005).

58.   P. Todd, K. I. Wolpin, *American Economic Review*, **96**, 1384 (2006).

59.   T. Dohmen *et al.*, "Individual Risk Attitudes: New Evidence from a Large, Representative, Experimentally-Validated Survey", forthcoming in *Journal of the European Economic Association*.

60.   U. Fischbacher, S. Gächter, forthcoming in *American Economic Review*.

61.   M. Browning, L. P. Hansen, J. J. Heckman, in *Handbook of Macroeconomics,* J. B. Taylor, M. Woodford, Eds. (Elsevier, Amsterdam), **1A**, 543 (1999).

62. J. J. Heckman, *Journal of Political Economy* **82**, S136 (1974).

63. D. Houser, M. Keane, K. McCabe, *Econometrica* **72**, 781 (2004).