JOURNAL OF
ECONOMIC
SURVEYS

doi: 10.1111/j.1467-6419.2010.00676.x

WHAT'S TO KNOW ABOUT LABORATORY EXPERIMENTATION IN ECONOMICS?

Nick Feltovich
University of Aberdeen

Abstract. Experimental economics has grown as a discipline from near non-existence 50 years ago to a full-fledged field within economics in the present. Much of experimental economics research involves experimental methods as a tool, applied to problems in other fields of economics. However, some of this research is inward looking, focusing on questions of the methodology of experimental economics. In this note, I briefly discuss two methodological issues in experimental economics that might benefit from meta-analysis: the pool from which experimental participants are drawn (university undergraduate students versus other populations) and the scale of monetary incentives faced by participants (large, small or hypothetical).

Keywords. Field experiments; Incentives; Laboratory experiments; Meta-analysis; Subject pool

Background

The field of experimental economics has grown rapidly over the past 50 years. As recently as the early 1980s, fewer than 50 experimental economics articles per year were published, but this number had more than quadrupled by 2000 (Holt, 2007), and by 2007, the *American Economic Review* alone had a dozen articles containing either experiments or theoretical models designed to explain experimental results. Experimentation in economics takes many forms, but in what follows, I will concentrate on controlled experiments: primarily laboratory experiments, and to a lesser extent field experiments. Clear-cut definitions of these classes of experiments are difficult to produce, but laboratory experiments tend to have the following features. Participants (also called *subjects*) – usually university undergraduate students – are given instructions describing the decision-making environment. The environment can be any one of interest to economists: negotiating, bidding in auctions, buying and selling assets, contributing to a public good, setting product prices, etc. After the instructions are read by the subjects – and usually aloud by the experimenter – subjects make decisions in the setting described by the instructions, often repeatedly. Subjects' interactions typically take place over computer networks, though pen-and-paper experiments were more common in the past, and are still sometimes used. After all decisions in the experimental session have been made, the subjects are paid. By contrast with most experiments done by other social scientists, it is standard practice in economics to pay subjects based on results (e.g. if subjects play the role of firms in a market, they might be paid based on their profits), and these payments tend to be high compared to alternative earning opportunities such as minimum-wage work.

Lab experiments have several desirable features, which may partly explain their rising popularity. First, the experimenter has substantially more information about the decision-making environment than researchers in the field tend to have. *Induced-value theory* (Smith, 1976) allows the experimenter to endow subjects with desired preferences, in contrast to other approaches which either assume certain preferences or attempt measurement of them (imperfectly, by necessity). Second, and relatedly, the experimenter can vary the parameters of the experiment in a systematic way; achieving the same amount of control outside the lab can range from difficult to impossible, leading to questions of selection bias. Third, the experimenter can collect information that the field researcher typically doesn't have access to (e.g. in studies of markets, the field researcher may have better access to data about transactions that did occur than ones that did not due to failure to negotiate an agreement), lessening issues of truncation or censoring in the data.

But the use of laboratory experiments in economics – and in particular, the standard methodology described above – has attracted some criticism. A growing strand of literature within experimental analysis is inward looking: devoted to experimental analysis of experimental economics methodology. As this literature continues to grow, the potential value of meta-analysis as a complementary tool to individual experiments will grow apace. Below, I briefly discuss two issues of experimental economics methodology that are – or will soon be – ripe for meta-analysis.

Are University Undergraduates People Too?

This first question arises from the fact, mentioned above, that experimental subjects are usually university undergraduates. It is natural that experimental economics developed this way, given that undergraduates represent a large population in close proximity to the experimenter. Friedman and Sunder (1994) list additional advantages of undergraduates as subjects: (1) they have plenty of free time, so are often available to participate; (2) they learn the rules for a new situation relatively quickly and (3) they are cash poor, so incentives of a given size will seem relatively large to them. However, criticism of experimenters' reliance on undergraduates has existed almost as long as experimental economics has (see, e.g. Enis *et al.*, 1972; Cunningham *et al.*, 1974). Harrison and List (2004) present a nice discussion of the main disadvantages. First, undergraduates are not representative of the general population: they tend to be younger and better educated, and with a narrower range of socio-demographic characteristics. This lack of representativeness is exacerbated by the fact that students self-select into university attendance (likely based on unobservable as well as observable traits), leading to potential selection bias problems. Second, while undergraduates may be clever, they usually lack

		Player 2	
		Heads	Tails
Player	Heads	2, 0	0, 3
1	Tails	0, 2	3, 0

Figure 1. A Generalized Matching Pennies Game.

specialized expertise in the experimental environment; they usually have little experience in tasks such as negotiation or asset trading, and likely none in others such as product pricing.

These shortcomings may be overcome – at least in part – by using non-traditional experimental subjects instead of the usual undergraduates, and there are now a fair number of studies which do this. These studies fall into two groups: those using a broader, more heterogeneous segment of the population and those using people thought to have specialized expertise. It is worth noting that either of these types of study generally addresses only one of the shortcomings of using undergraduates: broader segments of the population may bring with them no more understanding of the experimental environment than undergraduates do, and experts may also be unrepresentative of the larger population. Indeed, the use of experts may even be counterproductive, in cases where the experimental environment is not exactly the same as the environment experienced by these participants. Binmore (2007, pp. 9–10) and others have argued that one reason that behaviour in experiments sometimes differs from theoretical predictions is that people bring their experience from related (but not identical) situations into the lab with them, and only with practice do they gradually understand the true decision making environment – and act accordingly.[1] If so, then using experts might lead to a loss of experimental control over preferences, at least initially.

An illustration of this approach is the literature on experts playing *generalized matching pennies games*, though there have been plenty of studies using other experimental environments and heterogeneous populations.[2] Generalized matching pennies games capture an important idea in economics – the value of being unpredictable to others – and are found in many economic settings. An example is shown in Figure 1. Player 1 and Player 2 simultaneously (i.e. without knowing the other's choice) make their choices (Heads or Tails), which determine one of the four cells of the payoff matrix. Within each cell, the first number is Player 1's payoff and the second is Player 2's payoff. The defining feature of this class of games is that they have a unique Nash equilibrium, which is in mixed strategies. Indeed, the game in Figure 1 has only one Nash equilibrium, where Player 1 chooses Heads with probability 0.4 and Player 2 chooses Heads with probability 0.6.

Experiments based on these games present a severe test of game theory as a prediction of behaviour, because when one player plays according to the Nash equilibrium prediction, the other has no strict incentive to do so, and if one player deviates even a minuscule amount from this prediction, payoff maximization requires the other player to choose a pure strategy instead. In fact, experiments with undergraduate subjects often find two types of discrepancy between their behaviour and the theoretical predictions: (1) actual choice frequencies differ from those implied by the mixed-strategy equilibrium (Ochs, 1995) and (2) subjects' play is exploitable in the sense that their past outcomes can be used to predict their future choices (Brown and Rosenthal, 1990).

These results beg the question of whether people with experience in such games would yield qualitatively different results. Walker and Wooders (2001) argue that professional sports provides a rich supply of real generalized matching pennies games in which highly motivated experts face each other. They analyse a set of men's professional tennis matches, mainly from Grand Slams, and examine whether servers' behaviour (whether to serve to the opponent's left or right) is consistent with mixed-strategy Nash equilibrium. They find that choice frequencies are nearly never significantly different from the equilibrium prediction, whereas their analysis of an earlier lab experiment with undergraduate subjects (O'Neill, 1987) finds substantial differences between observed and predicted frequencies according to the same measures.[3] On the other hand, they find that even the professionals behave in an exploitable way: they were more likely to switch between choices than the theory predicts, though they showed less of this negative auto-correlation than the undergraduates. Hsu *et al.* (2007) report even stronger evidence that professionals play as the theory predicts, using a larger data set containing men's, women's and juniors' tennis matches. Not only do they find choice frequencies consistent with equilibrium, but they find little evidence of negative auto-correlation in their data, in contrast to Walker and Wooders (2001).

Chiappori *et al.* (2002) and Palacios-Huerta (2003) examine another situation in sports where unpredictability is important: penalty kicks in soccer. Chiappori *et al.* (2002) model the penalty kick as a game between kicker and goalkeeper, with simultaneous choices of left, middle and right. Using data from the French and Italian first leagues, they find that behaviour corresponds closely to the theoretical predictions in both aspects: choice frequencies and lack of exploitability. Palacios-Huerta (2003) reaches a similar conclusion using data from multiple European soccer leagues. In a follow-up paper, Palacios-Huerta and Volij (2008) find that Spanish professional soccer players not only conform to the theory on the pitch, but also in the lab – in two abstract games like the one in Figure 1 – while a group of students playing the same games are much less likely to do so.

This miniature survey suggests that there are indeed systematic differences between experts and undergraduates in generalized matching pennies games. Meta-analysis could further increase our understanding of the effects of using experts in these and other environments. In which environments do experts behave differently from undergraduates? How does this effect vary with the situation, and how does it compare to the effects of other treatment variables? (Some treatment of publication bias might be needed to accurately answer these questions.) Similarly, meta-analysis can help us understand when the use of more heterogeneous populations should be expected to make a difference – and what kind of difference to expect.

Are Experimental Subjects Paid the Right Amount?

Our second question actually has two components, only somewhat related: (1) are typical payments to experimental subjects sufficient to induce appropriate preferences and (2) would smaller – or even hypothetical – payments also suffice? Smith (1982) sets out three conditions for payments to achieve control over preferences: *monotonicity* (subjects prefer more of the payment medium to less), *salience* (payments correspond to outcomes in the experiment, such as profits) and *dominance* (payments are large enough to overwhelm other aspects of outcomes that subjects might care about). Any one of these conditions seems to rule out hypothetical payments, and dominance implies that payments should be as large as feasible (taking into account the experimenter's budget constraint). The profession – both within and outside experimental economics – is nearly unanimous regarding the use of substantial monetary payments. Aumann (1990) criticizes the use of experiments in which subjects 'are paid off by pittances' and argues that 'when . . . the payoff is not five dollars but a successful career . . . the predictions of game theory fare quite well' (p. xi). More recently, Friedman and Cassar (2004) claim that '[e]xperiments with no salient rewards are not experimental economics. . .' (p. 28). Indeed, for a researcher using hypothetical payments, the burden of proof is on her to conduct additional paid sessions and show that behaviour is not fundamentally different; likewise, the only way to address claims that monetary incentives are too small is to add treatments with larger stakes.

Both of these augmentations to an experimental design obviously can add substantially to the cost of running experiments. Rubinstein (2001) criticizes the proscription on hypothetical payments: 'Good experiments which demonstrate commonly used [patterns] of reasoning 'work' even when we use cheap procedures that do not require paying any monetary rewards. . .' (p. 626) – also pointing out that the costs of experiments with monetary payments can serve as a barrier to entry into the field. Camerer and Hogarth (1999) disparage 'seminar participants [who] invariably criticize experimental evidence of violations of rationality principles by conjecturing that if enough incentive were offered . . . violations would disappear, ignorant of the fact that this conjecture has generally proved false' (p. 31).

Camerer and Hogarth (1999) themselves conducted a quasi-meta-analysis of incentives in experiments. They examine a sample of 74 studies involving high, low or hypothetical payments, and classify each according to the effect of raising incentives (either from none to low, or low to high). Their main result is that using financial incentives, or increasing their size, often has little benefit (though it can serve to decrease the variance of choices), and when it does have an effect, its size is comparable to that of other treatment variables (and indeed can interact with these other variables). As Camerer and Hogarth acknowledge, though, their study used a fairly small, non-random sample and noted that the reader could regard their paper as 'an informed essay or collection of conjectures, which may or may not prove true after a more careful meta-analysis. . .' (p. 11). Given the volume of literature since their article was published, their recommendation is increasingly apt.

Discussion

Controlled experimentation has grown increasingly common as a tool for analysis of problems in economics. In order for conclusions based on experimental results to be taken seriously, however, it is necessary for experimental economists to convince the rest of the discipline that their methodology yields results generalizable to the outside world. A small but significant portion of experimental economics research is devoted to examining methodology; however, there has been little systematic study of this strand of the literature. It seems ripe for meta-analysis. Two topics that might benefit from meta-analysis – the pool from which subject are drawn (university students versus other populations) and the scale of monetary incentives faced by participants (large, small or hypothetical) – are discussed above.

On the other hand, there is also plenty of scope for meta-analysis addressed to the other main strand of experimental research: using experiments to address questions from other parts of economics. While the potential for systematic study of multiple experiments to provide more powerful results than a single experiment is well known (for example, Roth, 1995, argues that discussion of experimental results should be organized around the series of experiments, rather than the individual experiment), there have been relatively few such studies. One of the first was by Sally (1995), who performs a meta-analysis of prisoners'-dilemma experiments up to 1992. Two of his most noteworthy findings were that (verbal) conversation is substantially more effective at increasing cooperation than written messages, and that cooperation is improved when the experimenter frames the communication as 'promises'. Many prisoners'-dilemma experiments have been conducted since 1992, and a meta-analysis that includes these more recent experiments could serve not only to confirm Sally's results, but also to take advantage of this larger sample to examine still more factors affecting cooperation. (For example, one might distinguish experiments involving written messages according to whether they use highly structured messages – such as a binary 'I will cooperate/defect' choice – or free-form messages such as in a computer chat room, to see whether there is more to messages than a claim of intended action.)

More recent meta-analyses in experimental economics include those of Zelmer (2003), who looks at factors affecting investment in linear public goods experiments; Oosterbeek *et al.* (2004), who search for cultural differences among responders in ultimatum game experiments; Johnson and Mislin (2009), who examine trusting and trustworthy behaviour in trust games and Engel (2010), who studies offers in dictator games. However, there are other areas of the literature where multiple experiments have been conducted, making meta-analysis possible. One promising candidate for future meta-analysis is beauty-contest games, where a group of players simultaneously 'guess' numbers in an interval, and win a prize if their guess is closest to a fraction (usually less than one) times an order statistic calculated from the guesses (usually the mean). As with generalized matching pennies games, sharp differences exist between equilibrium predictions and observed behaviour in beauty-contest games, likely due to breakdown of the assumption of common knowledge of rationality (see, e.g. Nagel, 1995). Experiments have found that behaviour can be sensitive to many features of the experimental environment, including the number of players in the game (Ho *et al.*, 1998), the order statistic used (Duffy and Nagel, 1997), the type of feedback given to subjects (Weber, 2003) and whether 'players' were individual subjects or three-person teams (Kocher and Sutter, 2005). Meta-analysis could provide a welcome characterization of the effects of these and other manipulations.

Acknowledgement

I thank Chris Doucouliagos and Tom Stanley for helpful suggestions and comments.

Notes

1. A classic example is the ultimatum bargaining game, where Player 1 makes a take-it-or-leave-it proposal to Player 2 to divide a known amount of money (say, \$10), which is implemented if Player 2 accepts, while both players get nothing if Player 2 rejects. If players care only about their money earnings, then Player 1's proposal should offer Player 2 either nothing or nearly nothing, knowing that Player 2 should accept any positive offer, and is even indifferent toward accepting a zero offer! This outcome is nearly never seen in experiments; rather, subjects in the role of Player 1 typically offer nearly half of the money, and Player 2s frequently reject positive (let alone zero) offers. Binmore's (2007) argument is that subjects bring their experience of real bargaining – where one side rarely has the degree of bargaining power seen in the ultimatum game – into the lab with them. An argument along the same lines was made by Smith (1976), who argues in favour of phrasing experiment instructions in abstract terms rather than in real-world context, so as to avoid this loss of control: 'it may be preferable *not* [emphasis in original] to embellish the instructions with well-intentioned attempts at "realism." Let the explicit reward structure be the singular source of valuation. . .'. (p. 278)
2. As examples of using experts in other experimental environments, see Cooper *et al.*'s (1999) comparison of managers of Chinese state enterprises and Chinese undergraduates in a 'ratchet effect' game, or Potters and van Winden's (2000) comparison of Dutch students and professional lobbyists in a signalling game meant to capture aspects of lobbying. As examples of experiments using more heterogeneous populations, see Bosch-Domenèch *et al.*'s (2002) comparison of beauty-contest games conducted via newspaper contests with those conducted in the lab, or Cummings *et al.*'s (1995) use of subjects recruited from churches for a valuation task.
3. Of course, it is impossible to construct the entire payoff matrix in the 'game' between two tennis professionals, so Walker and Wooders could not directly compare choice frequencies to their predicted values. Instead, they examined an implication of equilibrium play: whether the server was equally likely to win the point by serving to either side.

References

Aumann, R.H. (1990) Foreword. In A.E. Roth and M.A. Oliveira Sotomayor (eds), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis* (p. xi). Cambridge, UK: Cambridge University Press.

Binmore, K. (2007) *Does Game Theory Work? The Bargaining Challenge*. Cambridge, MA: MIT Press.

Bosch-Domenèch, A., Montalvo, J.G, Nagel, R.C. and Satorra, A. (2002) One, two, (three), infinity, . . . : Newspaper and lab beauty-contest experiments. *American Economic Review* 92: 1687–1701.

Brown, J.N. and Rosenthal, R.W. (1990) Testing the minimax hypothesis: A re-examination of O'Neill's game experiment. *Econometrica* 58: 1065–1081.

Camerer, C.F. and Hogarth, R.M. (1999) The effects of financial incentives in experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 19: 7–42.

Chiappori, P.-A., Levitt, S. and Groseclose, T. (2002) Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review* 92: 1138–1151.

Cooper, D.J., Kage, J.H., Lo, W. and Gu, Q.L. (1999) Gaming against managers in incentive systems: Experimental results with Chinese students and Chinese managers. *American Economic Review* 89: 781–804.

Cummings, R.G., Harrison, G.W. and Rutström, E.E. (1995) Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive compatible? *American Economic Review* 85: 260–266.

Cunningham, W.H., Anderson, T. and Murphy, J.H. (1974) Are students real people? *Journal of Business* 47: 399–409.

Duffy, J. and Nagel, R.C. (1997) On the robustness of behaviour in experimental 'beauty contest' games. *Economic Journal* 107: 1684–1700.

Engel, C. (2010) Dictator games: A meta study. Working paper, Max Planck Institute for Research on Collective Goods.

Enis, B.E., Cox, K. and Stafford, J. (1972) Students as subjects in consumer behavior experiments. *Journal of Marketing Research* 9: 72–74.

Friedman, D. and Cassar, A. (2004) First principles: Induced value theory. In D. Friedman and A. Cassar (eds), *Economics Lab: An Intensive Course in Experimental Economics* (pp. 25–31). London, UK: Routledge.

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

Harrison, G.W. and List, J.A. (2004) Field experiments. *Journal of Economic Literature* 42: 1009–1055.

Ho, T.-H., Camerer, C. and Weigelt, K. (1998) Iterated dominance and iterated best response in experimental 'p-beauty contests'. *American Economic Review* 88: 947–969.

Holt, C.A. (2007) *Markets, Games, & Strategic Behavior*. Boston, MA: Pearson.

Hsu, S.-H., Huang, C.-Y. and Tang, C.-T. (2007) Minimax play at Wimbledon: Comment. *American Economic Review* 97: 517–523.

Johnson, N.D. and Mislin, A. (2009) Cultures of kindness: A meta-analysis of trust game experiments. Working paper, George Mason University.

Kocher, M.G. and Sutter, M. (2005) The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *Economic Journal* 115: 200–223

Levitt, S.D. and List, J.A. (2007) What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21: 153–174.

Nagel, R.C. (1995) Unraveling in guessing games: An experimental study. *American Economic Review* 85: 1313–1326.

Ochs, J. (1995) Games with unique, mixed strategy equilibria: An experimental study. *Games and Economic Behavior* 10: 202–217.

O'Neill, B. (1987) Nonmetric test of the minimax theory of two-person zerosum games. *Proceedings of the National Academy of Sciences* 84: 2106–2109.

Oosterbeek, H., Sloof, R. and van de Kuilen, G. (2004) Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. *Experimental Economics* 7: 171–188.

Palacios-Huerta, I. (2003) Professionals play minimax. *Review of Economic Studies* 70: 395–415.

Palacios-Huerta, I. and Volu, O. (2008) Experientia docet: Professionals play minimax in laboratory experiments. *Econometrica* 76: 71–115.

Potters, J. and van Winden, F. (2000) Professionals and students in a lobbying experiment: Professional rules of conduct and subject surrogacy. *Journal of Economic Behavior & Organization* 43: 499–522.

Roth, A.E. (1995) Introduction to experimental economics. In J.H. Kagel and A.E. Roth (eds), *The Handbook of Experimental Economics* (pp. 3–109). Princeton, NJ: Princeton University Press.

Rubinstein, A. (2001) A theorist's view of experiments. *European Economic Review* 45: 615–628.

Sally, D. (1995) Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society* 7: 58–92.

Siegel, S. and Goldstein, D.A. (1959) Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology* 57: 37–42.

Smith, V.L. (1976) Experimental economics: Induced value theory. *American Economic Review* 66: 274–279.

Smith, V.L. (1982) Microeconomic systems as an experimental science. *American Economic Review* 72: 923–955.

Walker, M and Wooders, J. (2001) Minimax play at Wimbledon. *American Economic Review* 91: 1521–1538.

Weber, R.A. (2003) 'Learning' with no feedback in a competitive guessing game. *Games and Economic Behavior* 44: 134–144.

Zelmer, J. (2003) Linear public goods experiments: A meta-analysis. *Experimental Economics* 6: 299–310.