Journal of Economic Behavior & Organization 91 (2013) 93–100

Contents lists available at SciVerse ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Experimental methods: Extra-laboratory experiments-extending the reach of experimental economics

Gary Charness[a,*], Uri Gneezy[b,c], Michael A. Kuhn[b]

[a] University of California, Santa Barbara, United States
[b] University of California, San Diego, United States
[c] CREED, University of Amsterdam, Netherlands

ARTICLE INFO

Article history:
Received 28 December 2012
Received in revised form 26 March 2013
Accepted 1 April 2013
Available online 17 April 2013

JEL classification:
C91
C93
C99

Keywords:
Extra-laboratory experiments
Laboratory experiments
Field experiments
Experimental methodology

ABSTRACT

We propose a new organizing scheme for classifying types of experiments. In addition to the standard categories of laboratory and field experiments, we suggest a new category: "extra-laboratory experiments." These are experiments that have the same spirit as laboratory experiments, but are conducted in a non-standard manner. We also suggest some organizing principles to distinguish amongst the three types of experiments.

1. Introduction

We propose that experiments that are conducted outside the typical on-campus lab and with participant knowledge should be considered *extra-laboratory* experiments. These are akin to classic lab experiments in every way except for the venue and the subject pool. Given the recent increase in popularity of online experimentation, targeting non-student populations and use of experiments in developing countries, it is important that these studies are anchored to their lab precursors as a means of comparison.

Extra-lab experiments differ from lab experiments along a number of dimensions. They typically entail a different subject pool and/or a different venue. They incorporate a wide variety of selection processes: mandatory participation in the classroom, voluntary participation from a select population online, targeted sampling in rural villages and many others. Extra-lab experiments have provided observations of behavior under a broad spectrum of stakes, from pennies online to months of income in developing countries. The administrators of the experiment are often very different: online interfaces remove all human interaction while local translators and community members have a closer connection to the subjects. Whether or not these differences are desirable depends primarily on the question being asked; often the direct effects of these differences are objects of economic interest.

* Corresponding author. Tel.: +1 8056881408.
E-mail addresses: charness@econ.ucsb.edu (G. Charness), ugneezy@ucsd.edu (U. Gneezy), mkuhn@ucsd.edu (M.A. Kuhn).

0167-2681/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jebo.2013.04.002

Our paper relates to the taxonomy of Harrison and List (2004), which expanded the definition of field experiments to include experiments conducted outside of the lab and with non-student participant pools. In contrast with this taxonomy, we propose that extra-lab experiments are different than field experiments, and offer some clear organizing principles that define extra-laboratory experiments as separate from both laboratory experiments and field experiments.

In our classification, a sufficient, but not necessary, condition for an experiment to be classified as a field experiment is that the participants are not aware of the existence of an experiment. Participant awareness of being involved in an experiment does not, by itself, disqualify that experiment from being a field experiment, nor does its location: by our classification, a field experiment could be conducted in the laboratory. Gneezy et al. (2013) conduct an experiment in the laboratory that involves overpaying participants from a previous, unrelated experimental task and observing whether they report the overpayment. The overpayment is designed to look like a casual mistake: instead of ten $1 bills for the correct payment of $10, subjects receive nine $1 bills with a $5 bill mixed somewhere in the middle. By mimicking an experience that could happen anywhere, the experimenters bring the field to the lab.[1] More generally, while the venue for an experiment is important, we do not feel that this factor is what distinguishes a field experiment from a lab experiment.[2]

Instead, one primary factor for determining whether an experiment is a field experiment is whether the activity of interest would have happened but for the actions of the experimenter. If this is the case, the experimenter is taking advantage of an existing natural environment. Examples include Lucking-Reiley (1999), an experiment that involves auctions of collectible cards on the Internet, List (2001), which experiments in an in-person sports-card market, and Garratt et al. (2012), which involves E-bay auctions for goods common to the site. However, on this basis, neither the public-goods experiments conducted in these venues nor the experiments in rural third-world settings with standard laboratory games (e.g., Gneezy et al., 2009; Henrich et al., 2001) would be classified as field experiments.

A similar approach to defining what constitutes a natural environment is to say that when the data is non-observational, the experiment is likely not a field experiment. In other words, when the subjects report their decisions or opinions directly to the experimenter, more likely than not, the natural environment has been compromised in some way.

Of course, this leaves some gray areas, and generates exceptions. Consider Gneezy et al. (2004), where the experimenters vary how a restaurant bill is split amongst the participants (and the experimenter). The restaurant is clearly a natural environment and choosing how to split the bill is a natural feature, but the diners are aware that the experimenter has altered the costs and benefits of splitting the check in certain ways. Even though all other aspects of the experiment meet the "but-for" definition provided earlier, the experimental manipulation of costs does not. Hence, we would classify it as an extra-lab experiment and not a field experiment. Another case in point is Charness and Gneezy (2009), in which participants are paid to go to the gym over time. While it is natural for students to go to the gym, paying them to do so is an unusual intervention.

Importantly, experiments in education or health interventions (e.g., Angrist and Lavy, 2009; Loewenstein et al., 2012) in which participants sign consent forms and are paid for performance improvements could fit into either category. As an analogy, consider an experiment that incentivizes students to get a high GPA. The students are engaged in an activity they would do regardless of the study, and the experimenter just adds incentives.

What distinguishes a standard laboratory experiment from an extra-lab experiment? Once again, the venue is not at the heart of the matter. Running standard lab experiments with special populations such as soccer players (Levitt et al., 2010), older and younger workers on location at French firms (Charness and Villeval, 2009) is, in our classification, an extra-lab experiment. Some primary factors in separating standard laboratory experiments from extra-lab experiments include having access to a broader range of age, culture, and experience than is otherwise available in student populations, having a population more attuned to the research question, and permitting activities to be carried out continuously over a longer period of time than is feasible in the laboratory.

Importantly, we do not argue for the any general superiority of any of these methodological approaches over any other. We simply feel that one should use the best tool available for the task at hand. In this respect, in many cases it is particularly useful to take the lab out to the field and conduct tests using standard stylized tasks, but with special subject pools and special locations.

The paper is organized as follows: Section 2 highlights motivations for taking experiments out of the lab, Section 3 discusses the methodology of doing so, Section 4 presents some of the differences in results between the two paradigms and we conclude in Section 5.

2. Motivations for extra-laboratory experiments

2.1. Changing the subject pool

Beyond looking for the differential importance of a particular theory, a primary motivation for extra-lab experiments is that the variation in individuals found on campus is simply insufficient to answer some questions. Socio-economic and

[1] Another example comes from psychology: Bargh et al. (1996) had people create sentences from available words that contained references to negative side effects of aging. The researchers were not interested in the sentences created, but whether the primed individuals later walked more slowly down the hallway while exiting the experiment. They perform similar experiments using racial and emotional primes (and different outcome measures) as well.

[2] We recognize that to some degree this is an abuse of language, but prefer to stick with the terms commonly used by economists rather than add entirely new nomenclature.

demographic differences are likely to be important in determining the scope and magnitude of many phenomena that we commonly study in the lab. Consider research into whether or not gender differences in competitiveness are a result of nature or nurture. Gneezy et al. (2009) conduct extra-lab experiments in India and Tanzania that compare behavior across matrilineal and patriarchal societies to identify the effect that relative social positioning has on competition. Even if individuals from these societies were present in large numbers on campus, it is likely that the shared collegiate experience between the groups and selective processes that brought them there would dilute the culture effect substantially. In order to test the hypothesis of the research (that culture affects behavior), one must go to where the specific cultures are in full force. Nevertheless, there are often little to no protocol differences between the in-lab and extra-lab versions of these studies: a similarity that our classification highlights.

Similar motivations are behind the extra-lab experiments of Sutter and Rützler (2010) which studies the emergence of gender differences in children as young as 3 years old, Giordano et al. (2002) which finds that hyperbolic discounting accurately describes the behavior of opiate addicts over both time and heroin, and Haigh and List (2005) which shows that professional commodity traders exhibit more myopic loss aversion than do undergraduate students. All of these works are clear examples in which a change of subject pools corroborated a behavioral theory more thoroughly than would have been possible on campus

2.2. Dealing with selection

The students that come into economics laboratories are a highly selected group of individuals for many reasons. First, they are motivated, high ability individuals. Second, students who show up for lab experiments have exerted effort in search of the opportunity to participate. This often means registering for an online database, altering their daily schedule and showing up at the location of the experiment. The students who show up in our labs are likely a non-representative sample of a group that may be non-representative of the general population. For example, some evidence (e.g., Burger et al., 2011) suggests that students who participate in experiments are better students.

Changing the subject pool even within a campus can affect results. For example, Eckel and Grossman (2000) test dictator game behavior of traditional lab volunteers, students recruited to show up for a specific session, and pseudo-volunteers, students in a regularly-scheduled lecture asked to participate in the same experiment. They find across all of their specifications that pseudo-volunteers are more generous than volunteers and that personal characteristics affect contributions less for volunteers than for pseudo-volunteers.

A likely explanation of these results is that the volunteer subjects are more motivated by financial incentives, as one might expect given that lab subject recruitment is done most frequently by advertising the possibility to earn money. Thus, if a researcher wants to capture an effect of intrinsic incentives, they could consider shortcutting the selection process associated with the laboratory and running their experiment in a classroom or in some other venue. Support for this view comes from Cleave et al. (2011), who take a similar classroom versus volunteer approach for the elicitation of time and risk preferences, a domain in which intrinsic motivation is not expected to influence behavior, and find only minimal differences.

Extra-lab experiments have also been used to examine selection on observables. Charness and Villeval (2009) administer cooperation, competition and risk experiments on groups of elderly and young individuals in both the laboratory and on-site with the employees of large firms. The human-resources departments at these companies recruited employees, while researchers recruited students and retirees living near the laboratory to show up to the lab. This design allows considerations of the effects of age in isolation from employment selection. Without this aspect, observations of retiree cooperativeness and competitiveness would be highly suspect as general age-specific effects that could be applicable to workplace conduct.

Harrison et al. (2002) avoid much of the selection problem of the lab using what we classify as an extra-lab experiment. Through an arm of the Danish government, the researchers draw a random sample from all individuals aged 19–75 that had participated in a previous economic survey. By setting up facilities all over the country, they were able to generate a sample that was nationally representative in terms of the population distribution from which to elicit individual discount rates. The estimates of discounting generated from this nationally representative extra-lab experiment are likely more dependable for forecasting the impacts of policy interventions than estimates generated from a single on-campus lab with students in one city.

2.3. Controlling costs

Leaving the on-campus lab environment allows the researcher to exploit the stakes of play for various purposes. Most notably, researchers can observe behavior under incentives that are very large for the subjects. Ariely et al. (2009) examine the productivity effects of performance-based pay in India and the United States. People in some treatments in rural India could earn $45, which is almost half of the mean yearly expenditure in their village, if they achieved a high performance level on all their tasks. The portion of the experiment in the US had maximum earnings of $300. The strength of the incentive in India is much greater than is feasible in Western nations. For similar reasons, the ultimatum game has been well studied by researchers seeking to maximize the incentive power of their available budget (Slonim and Roth, 1998; Cameron, 1999; Andersen et al., 2011).

Other extra-laboratory environments allow researchers to greatly reduce the total costs of experiments. Many have started using relatively inexpensive online experimentation for part or even all of an experiment.[3] The Internet has the advantage of generally lower costs (such as not requiring show-up fees), and having to compensate individuals less for their convenience costs. In addition, it allows for the rapid recruitment of larger samples of participants than are usually available on campuses.

A problem with online experiments is that they replace the problematic selection into the laboratory with a problematic selection into online experiments. For example, users accessing sites like Amazon's Mechanical Turk are also selected for extrinsic motivation, sometimes to a greater extent than laboratory participants, with compensation as low as 5 cents per participant.[4] While labs recruit individuals from universities, online subjects initiate the search for earnings opportunities.

2.4. Extending the time frame

One potential advantage of extra-lab experiments is that they can be used to generate long time-series data. Recalling subjects to the lab every day, week or month is a difficult undertaking. Dealing statistically with attrition can be time consuming. Online, in-classroom or at-work, extra-lab experiments can potentially overcome this hurdle by making repeated participation convenient.

In labor economics and public finance, access to better panel data in recent years has greatly expanded the scope and power of empirical economics. It should be no different in experimental economics. However, the panels we create during the course of hour-long laboratory sessions are problematic for obvious reasons: The classic 10–20 period repetition of the public-goods game may not be a good analog for the expression of preference for tax policy that citizens make when they vote every couple of years. Extra-lab experiments could generate panels that reflect real-world scenarios of interest more closely.

To our knowledge, this is a relatively unexploited use for extra-lab experiments that study behavior in economic games. Burger et al. (2011) use an online experiment to examine procrastination over a two-day period. Subjects are paid to solve economics problems over either a Tuesday–Wednesday or a Friday–Saturday period. The weekday–weekend variation manipulates the time-costs of effort. It would be difficult to design an hour-long laboratory experiment that captured the motivations at issue here while maintaining external validity. Additionally, this paper reports an experiment that uses an online study diary to track subjects over a five-week horizon. Similar use of the Internet for record-keeping during an extra-lab experiment has been used for incentivized exercise studies (Charness and Gneezy, 2009; Babcock and Hartman, 2010; Acland and Levy, 2013; Babcock et al., 2011).

2.5. Why not do a field experiment?

The motivations for extra-laboratory experiments listed above were presented as if the alternative were on-campus lab experiments. In many cases, field experiments are the obvious alternative. For example, when one is interested in a sensitive topic like discrimination, going out to the field and testing the reaction of individuals in a real market who do not know that they are taking part in a study and being observed is the best alternative (List, 2004a).

In general, field experiments have many theoretical advantages, but serious practical drawbacks. In some cases setting up a field experiment in which there is sufficient variation, true randomization, and the possibility of cloaking the experiment in a shield of invisibility may be rather infeasible or even impossible. Another important drawback of field experiments is the loss of experimental control relative to the laboratory. Related to this is the difficulty in replicating field experiments. In most cases laboratory experiments are easier to replicate, which is an important property for any empirical research. Thus, there is a panoply of reasons why an extra-lab experiment might be more suitable than a field experiment.

Once again, there are advantages and disadvantages to each category of experiment. We do not believe that there is a hierarchy of methods between laboratory, extra-laboratory and field experiments. Different methods are better suited for different environments. At the end of the day, the convergence of evidence from all methods should be the researcher's goal.

3. Methodology

Laboratory experiments set the gold standard for environmental control. Any uncontrolled complexities or confounds in the environment push researchers in the direction of the laboratory, since it is easier to exercise strong control over an experiment with (by design) fewer moving parts. Nevertheless, one might wish to perform an extra-lab experiment with a highly specialized group of experts at work for the purpose of implementing a very complicated protocol. Such versatility is one of the advantages of extra-lab experiments compared to both field and traditional laboratory studies.

[3] A relatively embryonic experimental methodology involves Second Life, an online computer game in which online avatars make choices (see, for example, Fiedler et al., 2011). Aside from the other benefits and drawbacks of this approach, the units of pay in the Second Life are quite small in terms of actual currency.

[4] However, when it comes to external validity, Paolacci et al. (2010) assert that Mechanical Turk is "at least as representative of the U.S. population as traditional subject pools."

3.1. Classroom experiments

Classroom experiments can deliver a substantial number of observations very quickly. But to ensure high-quality data, experimenters must solve the problems of distributing and collecting materials, typically confidentially, garnering the close attention of the students and preventing spillovers between neighbors, all in a very short time frame so as to avoid using too much of the instructor's time. Given these conditions, simplicity should be the organizing principle behind experiments designed for classrooms.

The Beauty Contest game is an example of a simple, classroom-ready experiment. It is commonly used in many undergraduate microeconomics courses to introduce game-theoretic concepts and thinking. In fact, Rosemarie Nagel, the author of many well-known articles (e.g., Nagel, 1995) on the Beauty Contest, learned about it during a game-theory course, where it was used as a demonstration. All one needs to do is announce the framework of the game, pass out small sheets of paper and collect those sheets in a controlled fashion. More importantly for time considerations, the experimenter only needs to pay one individual in the case of the game (barring ties). Bosch-Domènech et al. (2002) report the results of two Beauty Contest experiments conducted in the classroom, in which results are similar to those obtained in the laboratory, but with samples that are 2–3 times larger per session. Interestingly, a number of extra-lab Beauty Contest experiments were run using newspapers, where participation is solicited based on mail-in replies (Thaler, 1997; Bosch-Domènech and Nagel, 1997; Selten and Nagel, 1997).

A downside of classroom experiments is that while the observation count is larger, so is the variability. Across classes, times of day, days of the week and weeks of the semester, attendance relative to enrollment varies substantially. For this reason, sampling from similar classes in similar time slots at similar points in the schedule is important, especially if the classroom is the level of treatment variation. Care should be taken to avoid any unintended priming in the different classes.

3.2. Online experiments

The domain in which maintaining the degree of control that is feasible in the laboratory might be the most difficult is on the Internet. Amazon's Mechanical Turk crowdsourcing site is the most notable example. Paolacci et al. (2010) point out that even the basic nomenclature of the service differentiates it from the lab in some respects. Subjects are called "workers" in this environment; they can view and select jobs with whichever wage and task profiles they prefer. Imagine posting your entire list of projects on your lab's website and allowing subjects to choose which ones to participate in. Experimenters using this platform need to carefully consider what message the task description broadcasts, and how this description and the offered wage will interact with the other postings in the subject's environment in order to minimize or manipulate selection from the worker pool.

Horton and Chilton (2010) estimate that the willingness to work on Mechanical Turk is about $1.38 per hour. When subjects encounter offered jobs that pay substantially more than this, what is the effect on their mindset? Researchers studying labor supply might want their posting to blend in with others in order to take advantage of the employee-employer structure that already exists on the site. On the other hand, researchers studying social interaction or social preferences might want their posting to stick out and generate a sense of windfall. This example outlines how online experimentation can easily offer opportunities for extra-lab experiments and field experiments in a nearly identical framework.

Another worry that researchers have about online experimentation is that respondents to surveys will answer as quickly as possible, without regards for accuracy. This critique applies to surveys more generally, as detailed by Krosnick et al. (1996). While concerns about selection into online surveys may exacerbate the satisficing problem, the format allows for a number of remedies. Using a trick question technique, Kapelner and Chandler (2010) find that forcing subjects to spend a requisite amount of time on survey questions, and sequentially highlighting the question text both have a significant positive effect on response accuracy. Asking subjects explicitly for their "careful consideration" does not. Concerns about the additional time such remedial procedures might take are mitigated by another result from Horton and Chilton (2010), that while workers on Mechanical Turk do respond to offered wages, they do not avoid more time-consuming tasks.

3.3. In the wild

Traveling to distant locations (either physical or financial) to perform extra-laboratory experiments often necessitates a reframing of the common tasks used in the lab and an attempt to impose some of the structure of the lab on the environment. Harbaugh and Krause (2000) adapt a linear public-goods game for elementary school students by replacing cash with tokens redeemable for toys, using physical actions (like placing tokens in a cup for keeping or an envelope for sharing) to reinforce the meaning of decisions and acting out different scenarios for the children to observe and consider before playing. Their goal was to turn the activity into a physical, rather than cerebral experience.

Gneezy et al. (2009) use a ball-toss game to replicate laboratory experiments on piece-rate versus tournament pay. Standard lab experiments in this area usually involve solving mazes or math questions. In general, translating strategies on paper into physical actions is a common theme of many extra-lab experiments that travel to remote, underdeveloped locations, or study specialized populations. An interesting aspect of this approach is that it avoids the standard ultra-reduced-form approach that is so common in lab studies, which contributes to the external validity of these results. These

physically-oriented versions of lab studies may in some ways be closer to the behaviors that the games and models we study are meant to describe.

This account of the importance of structure in extra-lab experiments may appear counter-intuitive. List (2006) argues that "representativeness of the environment, rather than representative[ness] of the sample populations, is the most crucial variable in determining generalizability of results . . ." While we interpret the term "environment" more broadly and consider the issues we mentioned earlier, such as awareness and observational or collected data, we only partially agree with this statement. We feel that the representativeness of the population is a very important consideration.

4. Do laboratory and extra-laboratory results differ?

We consider a number of canonical games and tasks to compare across the lab and extra-lab paradigms.

4.1. Ultimatum game

This simple bargaining framework in which a proposer makes an offer for how to split a pool of money and the responder can accept or reject this offer has become a workhorse of modern experimental economics ever since its introduction to the laboratory by Güth et al. (1982). The finding that responders often reject small offers at a cost to themselves (and that proposers, anticipating this, appear to make non-trivial offers) was a substantial motivating influence behind the explosion of research into social preferences.

Since then, researchers have been using lab experiments to examine the robustness of the results to an increase of stakes. Roth et al. (1991), Straub and Murnighan (1995), Hoffman et al. (1996) and Slonim and Roth (1998) all use laboratory experiments to test whether or not proposer generosity and responder spite decrease as stakes are increased. Some of these efforts involve moving the experiments to labs at universities in locations where the exchange rate can be exploited, but none of them establish a relationship between behavior and stakes. Andersen et al. (2011) move the experiment out of the lab into villages in rural India, where wages are low enough to allow the authors to vary the pool size from roughly an hour's wage to a year's wage, and find that the frequency of rejection is decreasing in the stakes. It seems that more research is needed on the issue of the size of the stakes in this game.

4.2. Public-goods game

An example of a very close correspondence between lab and extra-lab results comes from the linear public goods game. List (2004b) conducts an extra-lab public-goods game with card traders at a sports-card show, using cards as the currency of the game. The experimenters set up a stall advertising the sale of a particular card and then recruited visitors to the stall for participation in either a one-shot or multiple round version of the game. This represents a different selection process on a different population than in a university laboratory. One might expect experienced hagglers to be more strategic than students. Despite this, the one-shot results of significant contributions match those of Marwell and Ames (1981) and the multiple round results of gradually decreasing contributions are in line with those from Andreoni (1988), Isaac et al. (1994) and Fehr and Gächter (2000).

4.3. Trust games

Fiedler et al. (2011) manipulate the social distance of partners in trust games. First-movers can select their partner (to whom they pass part of their endowment), but the multiplier on passed money is lower if the first-mover selects a partner in the same lab area as him- or herself. In other words, to minimize social distance, they must pay a premium in money potentially sent back. They take identical procedures with students in a standard lab and with users of the online game Second Life. It turns out that subjects are more willing to pay a social-distance premium online than in the lab and that coordination via chat is a bigger factor in determining behavior. Second movers who were selected despite having a lower multiplier (the closer of the two potential partners) responded to their selection by passing higher amounts back. This indicates that even within the framework of the online game community, social preferences come into play.

Charness et al. (2007) administer the lost-wallet game (a near cousin of the trust game) in three environments: laboratory, classroom and online. They also manipulate the social distance of partners. They find that even in online experiments, many people return considerable amounts. While the levels in the three environments depend on social across, between environments there are surprisingly few qualitative differences in results.

4.4. Risk and time preference elicitations

Andersen et al. (2008) draw a nationally representative sample of Danish adults, and elicit their time and risk preferences in what we classify as an extra-lab environment. They find estimates of annual discounting preferences that are above mortgage and auto rates, but below credit card rates; in sync with the rates in the markets adults are commonly involved in. However, using a similar elicitation instrument in one of their specifications, Andreoni et al. (2013) find discounting preferences to actually be about double the standard rate on credit cards in lab experiments with California university

students. This might help to explain the prevalence of credit-card debt for young adults, despite similar rigorous controls across the studies for transaction costs and any uncertainty of actually receiving payment in the future. Circumventing the selection process clearly allows for an elicitation of preferences that policy makers might find more useful.

In the realm of risk preferences, countless lab studies have utilized the elicitation methodology developed by Holt and Laury (2002). A fairly standard result is that subjects exhibit quite high levels of risk-aversion in the laboratory. Eckel et al. (2009) utilize a somewhat different lottery choice task to elicit the risk preferences of Hurricane Katrina evacuees in Houston shelters shortly after the event. The authors find a remarkable degree of risk-seeking behavior in this population. Follow up surveys with a similar population roughly a year later indicated that this behavior diminished over time. The extra-lab experiment in this case allowed for the identification of an emotional-state specific effect on preferences.

In sum, whether results differ across experimental methodologies is an empirical question worth investigating. One concern, for example, is whether levels of a certain type of behavior vary across methodologies, but even if this were true, it may well be the case that treatment effects are robust with respect to experimental methodology. Expanding the breadth of subject pools and locations is a key step in the process of generalizing a lab result to make it more policy relevant and predictive.

5. Conclusion

We suggest a new classification of experiment methodologies that includes extra-laboratory experiments, which are generally somewhere between classic laboratory experiments and field experiments. We present organizing principles for this classification, highlighting a number of the distinct advantages (and thus motivations) for performing extra-lab experiments. These include changing the selection process, changing the subject pool, leveraging stakes, decreasing cost, or finding variation that simply would not exist in the laboratory. We briefly discuss the methodology of such papers, and examine results in canonical experiments across different venues and subject pools.

We wish to emphasize the importance of extra-lab experiments as links between experimental economics and policy making. Extra-lab experiments taken directly to a potential target population of a policy intervention can be a powerful tool for practitioners, especially when the extra-lab result is benchmarked against its lab precursor.

References

Acland, D., Levy, M., 2013. Habit formation, naiveté, and projection bias in gym attendance, GSPP Working Paper No. GSPP 13-002.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., List, J.A., 2011. Stakes Matter in Ultimatum Games. American Economic Review 101 (7), 3427–3439.

Andersen, S., Harrison, G.W., Lau, M.I., Elisabet Rutström, E., 2008. Eliciting risk and time preferences. Econometrica 76 (3), 583–618.

Andreoni, J., 1988. Why free ride? Strategies and learning in public goods experiments. Journal of Public Economics 37 (3), 291–304.

Andreoni, J., Kuhn, M.A., Sprenger, C., 2013. On measuring time preferences, UC San Diego. Working paper.

Angrist, J., Lavy, V., 2009. The effects of high stakes high school achievement awards: evidence from a randomized trial. American Economic Review 99 (4), 1384–1414.

Ariely, D., Gneezy, U., Loewenstein, G., Mazar, N., 2009. Large stakes and big mistakes. The Review of Economic Studies 76 (2), 451–469.

Babcock, P., Bedard, K., Charness, G., Hartman, J., Royer, H., 2011. Letting down the team? Evidence of social effects of team incentives. NBER working paper No 16687.

Babcock, P., Hartman, J., 2010. Exercising in herds: treatment size and status specific peer effects in a randomized exercise intervention, Mimeo.

Bargh, J.A., Chen, M., Burrows, L., 1996. Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. Journal of Personality and Social Psychology 71 (2), 230–244.

Bosch-Domènech, A., Montalvo, J., Nagel, R., Satorra, A., 2002. One, two, (three), infinity,. . .: newspaper and lab beauty-contest experiments. American Economic Review 92 (5), 1687–1701.

Bosch-Domènech, A., Nagel, R., 1997. Guess the number: comparing the financial times and expansion's results. Financial Times, June 30, Sec. 8, p. 14.

Burger, N., Charness, G., Lynham, J., 2011. Field and online experiments on self-control. Journal of Economic Behavior and Organization 77 (3), 393–404.

Cameron, L.A., 1999. Raising the stakes in the ultimatum game: experimental evidence from Indonesia. Economic Inquiry 37 (1), 47–59.

Charness, G., Gneezy, U., 2009. Incentives to exercise. Econometrica 77 (3), 909–931.

Charness, G., Haruvy, E., Sonsino, D., 2007. Social distance and reciprocity: an Internet experiment. Journal of Economic Behavior and Organization 63 (1), 88–103.

Charness, G., Villeval, M.-C., 2009. Cooperation and competition in intergenerational experiments in the field and the laboratory. The American Economic Review 99 (3), 956–978.

Cleave, B.L., Nikiforakis, N., Slonim, R., 2011. Is there selection bias in laboratory experiments? The case of social and risk preferences. IZA Disucssion Paper No. 5488.

Eckel, C.C., Grossman, P.J., 2000. Volunteers and pseudo-volunteers: the effect of recruitment method on subjects' behavior in experiments. Experimental Economics 3 (2), 107–120.

Eckel, C.C., El-Gamal, M.A., Wilson, R.K., 2009. Risk loving after the storm: a Bayesian-network study of hurricane Katrina evacuees. Journal of Economic Behavior and Organization 69 (2), 110–124.

Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. The American Economic Review 90 (4), 980–994.

Fiedler, M., Haruvy, E., Lin, S.X., 2011. Social distance in a virtual world experiment. Games and Economic Behavior 72 (2), 400–426.

Garratt, R.J., Walker, M., Wooders, J., 2012. Behavior in second-price auctsions by highly experienced eBay buyers and sellers. Experimental Economics 15 (1), 44–57.

Giordano, L.A., Bickel, W.K., Loewenstein, G., Jacobs, E.A., Marsch, L., Badger, G.J., 2002. Mild opioid deprivation increases the degree that opioid-dependent outpatients discount delayed heroin and money. Psychopharmacology 163 (2), 174–182.

Gneezy, U., Imas, A., Madarasz, K., 2013. Conscience accounting. Working paper.

Gneezy, U., Leonard, K.L., List, J.A., 2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. Econometrica 77 (5), 1637–1664.

Gneezy, U., Haruvy, E., Yafe, H., 2004. The inefficiency of splitting the bill. Economic Journal 114 (495), 265–280.

Güth, W., Schmittberger, R., Schwarze, B., 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior and Organization 3 (4), 367–388.

Haigh, M.S., List, J.A., 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. The Journal of Finance 60 (1), 523–534.
Harbaugh, W.T., Krause, K., 2000. Children's altruism in public good and dictator experiments. Economic Inquiry 38 (1), 95–109.
Harrison, G.W., Lau, M.I., Williams, M.B., 2002. Estimating individual discount rates in Denmark: a field experiment. American Economic Review 92 (5), 1606–1617.
Harrison, G.W., List, J.A., 2004. Field experiments. Journal of Economic Literature 42, 1009–1055.
Henrich, J., Boyd, R., Bowles, S., Gintis, H., Camerer, C., McElreath, R., 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. American Economic Review 91 (2), 73–78.
Hoffman, E., McCabe, K.A., Smith, V.L., 1996. On expectations and the monetary stakes in ultimatum games. International Journal of Game Theory 25 (3), 289–301.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. American Economic Review 92 (5), 1644–1655.
Horton, J.J., Chilton, L.B., 2010. The labor economics of paid crowdsouring. In: Proceedings of the 11th ACM Conference on Electronic Commerce.
Isaac, R.M., Walker, J.M., Williams, A.W., 1994. Group size and the voluntary provision of public goods: experimental evidence utilizing large groups. Journal of Public Economics 54 (1), 1–36.
Kapelner, A., Chandler, D., 2010. Preventing satisficing in online surveys: a 'Kapcha' to ensure higher quality data. In: Proceedings of 2010 CrowdConf.
Krosnick, J.A., Narayan, S., Smith, W.R., 1996. Satisficing in surveys: initial evidence. New Directions for Evaluation 1996 (70), 29–44.
Levitt, S.D., List, J.A., Reiley, D.H., 2010. What happens in the field stays in the field: exploring whether professionals play minimax in laboratory experiments. Econometrica 78 (4), 1413–1434.
List, J.A., 2001. Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. American Economic Review 91 (5), 1498–1507.
List, J.A., 2004a. The nature and extent of discrimination in the marketplace: evidence from the field. Quarterly Journal of Economics 119 (1), 49–89.
List, J.A., 2004b. Young, selfish and male: field evidence of social preferences. The Economic Journal 114 (492), 121–149.
List, J.A., 2006. Field experiments: a bridge between lab and naturally occurring data. The B.E. Journal of Economic Analysis and Policy 6 (2), Article 8.
Loewenstein, G., Volpp, K., Asch, D.A., 2012. Incentives in health: different prescriptions for physicians and patients. Journal of the American Medical Association 307 (13), 1375–1376.
Lucking-Reiley, D., 1999. Using field experiments to test equivalence between auction formats: magic on the internet. American Economic Review 89 (5), 1063–1080.
Marwell, G., Ames, R.E., 1981. Economists free ride, does anyone else?: Experiments on the provision of public goods, IV. Journal of Public Economics 15 (3), 295–310.
Nagel, R., 1995. Unraveling in guessing games: an experimental study. American Economic Review 85 (5), 1313–1326.
Paolacci, G., Chandler, J., Ipeirotis, P.G., 2010. Running experiments on Amazon Mechanical Turk. Judgment and Decision Making 5 (5), 411–419.
Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh and Tokyo: an experimental study. The American Economic Review 81 (5), 1068–1095.
Selten, R., Nagel, R., 1997. 1000 DM zu Gewinnen, Spektrum der Wissen-schaft, November. p. 10.
Slonim, R., Roth, A.E., 1998. Learning in high stakes ultimatum games: an experiment in the Slovak Republic. Econometrica 66 (3), 569–596.
Straub, P.G., Murnighan, K., 1995. An experimental investigation of ultimatum games: common knowledge, fairness, expectations and lowest acceptable offers. Journal of Economic Behavior and Organization 27 (3), 345–364.
Sutter, M., Rützler, M., 2010. Gender differences in competition emerge early in life. IZA Discussion Paper No. 5015.
Thaler, R., 1997. Giving markets a human dimension. Financial Times, June 16, Sec.1, p. 29.