Field Experiments

1 Introduction[1]

Although field experiments have been conducted by earlier researchers, their popularity has increased dramatically since the late 1990s. There is currently (circa mid-2000's) a heated debate over whether field or laboratory experiments are the "best" methodology.[2] I will make the same statement I made earlier when discussing theory and experiments, that we should use the results from all types of research to inform our policies and opinions, provided that the research is conducted in a careful and rigorous manner. What these notes will do is use laboratory experiments as a basis of comparison for field experiments, and point out the differences between field and lab experiments.

HL point out that the lab is a sterile environment, but that this is not a negative if one realizes the role of lab experiments in the discovery process. The sterility of the lab allows the researcher to make sharp inferences about the effects of exogenous treatments. But it may be that experiments conducted in the lab suffer from the fact that the conditions in the lab are not akin to conditions in the field or the "real-world". Thus, if there is a difference between the lab and the field that alters behavior, then drawing inferences from the lab to the field may be misleading. HL make 3 points about field experiments.

1. Dissecting the characteristics of field experiments helps define the "ideal experiment".

   Their definition of the "ideal experiment" is to "observe the subject in a controlled setting but where the subject does not perceive any of the controls as unnatural and there is no deception involved". I suppose the ideal experiment is similar to the movie *The Truman Show*, in which Jim Carrey plays a man who's been in a reality TV show his entire life but does not know it. Thus, one could introduce all sorts of treatment variables to see how the subject would respond. For the experiment to be "ideal", I would say that we would want to know one more thing – the individual's values and costs for objects. If you combine the premise of *The Truman Show* with some sort of device which could read an individual's mind without the individual knowing it,[3] then perhaps one could obtain the ideal experiment.

2. Many of the characteristics of field experiments are also found in varying degrees in lab experiments.

   Basically, field and lab experiments cannot be differentiated on some points since both contain these points.

3. There is much to learn from field experiments when returning to the lab.

   This is essentially an interaction similar to the one between experiments and theory. With lab and field experiments, one may be able to learn from the field to help design better lab experiments.

2 What is a field experiment?

We begin by listing the manner in which lab and field experiments may differ, then by defining the taxonomy of field experiments, as in Harrison and List (HL).

[1] Based in large part on Harrison and List (2004), Field Experiments, *Journal of Economic Literature*, Vol. 42 pgs. 1009-1055.

[2] If you know who the identities of the participants in the debate, fine, if not, I'm not going to tell you. I do not have tenure yet.

[3] Think of another Jim Carrey character, the Riddler, in *Batman Forever*.

2.1 Criteria of field experiments

HL set out 6 criteria for field experiments. These are ways in which field experiments may differ from lab experiments.

1. The nature of the subject pool.

   The typical subject pool in a lab experiment is undergraduate students, which may not accurately represent individuals in the field.

2. The nature of the information subjects bring to the experiment

   A wider pool of subjects may bring a wider set of information to the experiment.

3. The nature of the commodity

   If the commodity itself influences behavior, should incorporate the commodity into the theory.

4. The nature of the task or trading rules applied

   Typical lab subjects may not have had the time necessary to develop heuristics that those in the field may have developed.

5. The nature of the stakes

   It is possible that the stakes in the laboratory are not as high as those in the field, and that the nature of the stakes might affect behavior.

6. The nature of the environment in which the subject operates

   It may be that when subjects perform a task in a particular environment the environment itself causes subjects to rely on certain heuristics, whereas if these features of the environment are missing in the lab then the subjects may not rely on these heuristics.

2.2 Classification of field experiments

Following is the HL classification list of field experiments.

1. Conventional lab experiment – uses a standard subject pool, abstract framing, and imposed rules
2. Artefactual field experiment – same as a conventional lab experiment only that a nonstandard subject pool is used
3. Framed field experiment – same as an artefactual field experiment but with field context in either the commodity, task, or information set that the subjects can use
4. Natural field experiment – same as a framed field experiment but where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know they are in an experiment (e.g. the LR Magic card experiments)

Note that there are some changes that do not have a classification. For instance, if one used a context-loaded script as opposed to abstract-framing, that type of experiment does not fit into any of the classifications listed if undergraduate students are the subject pool. But this is a starting point.

3 Differences between the lab and the field

3.1 Differences in the subject pool

An artefactual field experiment is one where the only change from a conventional lab experiment is that a non-standard subject pool (i.e. non-student) is used. The primary reason for this is that the student population may not adequately reflect the characteristics of the population that is to be represented by the study. A harsher way of stating this is that "Students are not real people." Common responses to this claim are "If

you think 'real' people will generate different results go ahead and run the experiments with 'real' people" or "Prove that students are not real people" or "Theory does not make much of a statement about the realness of people, only that a rational agent have complete and transitive preferences". However, intuitively we might feel that a college student and a 70-year-old grandmother have different primitive characteristics given their likely differences in experiences. Thus, if a question really hinges on socio-demographic characteristics of an individual, such as estimating their degree of risk aversion or trust or some other "innate" parameter, then perhaps an artefactual field experiment need be conducted so that one does not assume that the risk preferences of the 20-year-old and the 70-year-old are identical.[4]

Now, how might the differences between students and non-students in socio-demographic characteristics and other experiences affect outcomes in an experiment? Consider an experiment on the hold-out problem in land acquisition. If both students and homeowners (assuming the two sets do not completely overlap) are placed in an experiment where they have a private value for a home and are asked to submit offers to sell that home to a developer, it is quite possible that the homeowners, who have the experience of living in a home, will submit higher offers to sell simply because they are internalizing all the aspects of their actual homes and communities when making their offers while the students, who do not have the experience of owning a home, submit lower offers to sell. Of course it could be the other way around if homeowners are more risk-averse than students or if homeowners have a bad experience owning their home.

3.2 Differences in context

3.2.1 Field context

In a framed-field experiment there is a non-standard subject pool as well as some field context in either the task, commodity, or information set that the subjects might have.[5] Thus, if there is something specific about the context of the task or commodity that causes people to implement a heuristic that they have found useful in this situation in the past then NOT running the experiment using that context will provide misleading results about how people behave when the context is applied. For example, consider a 1^{st}-price sealed bid common value auction. Note that this is different from the SIPV auction environment we discussed earlier. In a common value auction there is an item up for sale which has the same value to all participants, only the participants do not know the actual value of the item until they receive it. They can, however, form an estimate of the item's value. Many people will use oil deposits as an example for common values, but if I owned an oil deposit the only thing I could do is sell it since I know zero about drilling for oil (well, I do have a friend who's father used to work in drilling, so maybe my value is not quite zero). At any rate, think about a jar of coins – everyone can see the jar and form an estimate of how much money is in the jar, but no one is allowed to open the jar unless that person wins the auction. It is clear that all individuals have the same value for the item (unless of course one has a utility or disutility spike from receiving the coins rather than cash – I suppose we could say that the winner does not necessarily receive the jar but an amount in bills equal to the amount of money in the jar).

In the laboratory common value environments many subjects fall prey to the "winner's curse", where the subject who wins the item is the one who has formed the highest estimate of its value, and the estimate is most likely above the actual value of the item, and so is the bid. Theoretically this should not happen, and one can imagine that if individuals in the real world were constantly being drained of money by the winner's curse that they would eventually stop participating in these auctions. The only individuals in the laboratory who do not fall prey to the winner's curse are super-experienced subjects, chosen specifically because they did not fall prey to the winner's curse when they were inexperienced subjects.

In a framed-field experiment using sportscard dealers, the winner's curse is avoided. This does not mean that all field contexts have experienced subjects that avoid the winner's curse, but that the context might be important so that the individuals can recall heuristics they have used to avoid losses in the past. HL make a statement that, "It is not the case that abstract, context-free experiments provide more general findings if the context itself is relevant to the performance of subjects."

[4] On the other hand, if one can design a market or a mechanism in which behavior does not depend on those factors then knowledge of them is unnecessary. It is quite easy to believe that both the 20 and 70-year-old would be able to understand that they should remain in an English auction until their value for an item is reached.

[5] I suppose the holdout problem experiment with homeowners and calling the good a "home" might make it a framed-field experiment, although I doubt any individuals would be willing to play the game with their actual homes.

They also state that context-free experiments are unusually tough tests of the theory (although many theories have little or no context, and theorists are generally praised for making theories as general as possible so that they apply to a multitude of contexts). The primary point is that if "individuals want to draw conclusions about the validity of the theory in the field, they must pay attention to the myriad of ways in which field context can affect behavior". I don't necessarily disagree with this point of view, as I think that providing some context may impose some additional control in the experiment. For instance, earlier in the semester we saw a "trust game" or an "investment game". The fact that economists call this game two different things suggests that no one really knows what subjects are thinking, so using the game to measure trust or investment might lead to false inferences about trust or investment behavior if the subject's are not envisioning that game.

A final point they make is about the exogenous assignment of roles in the laboratory. It may be the case that some subjects are assigned roles (such as sellers) with which they have little or no actual experience. This may be in part because they have endogenously determined to NOT play that role in the "real-world".[6] Or consider the role of a financial trader in a laboratory experiment – again, students may have little to no experience in that role. Thus it is possible that individuals who do actively participate in that role in the "real-world" use different heuristics when making transactions.

3.2.2 Commodity context

Many field experiments involve real, physical goods and the values subjects place on them, as well as real tasks that the subjects may be used to performing. If the experiment is designed to restrict the actions that the subject may take when solving the task at hand, it may be that the subject would like to choose an action that is not available to find the solution. In the Tower of Hanoi game, there are 3 pegs. On the first peg there are 5 discs of different sizes, and the discs are arranged in increasing size, with the smallest disc on the top of the stack and the largest disc on the bottom. The goal is to move the discs so that the discs are in the same order but on the third peg. There are two restrictions. First, only 1 disc may be moved at a time. Second, at no point in time can a larger disc be placed above a smaller disc. If one were to visualize the goal state (the discs on the third peg), one can see that to achieve the goal state the smallest disc must be alone on either the first or second peg immediately prior to reaching the goal state. Taking it one more step backward, one must realize that the second-smallest disc must be on the peg that the first disc is not on, and so forth. Thus one can use backward induction to solve the game. However, when children are asked to solve the game they tend to violate the rules of the game and move all discs to the goal state and *then* work backwards.[7] Thus, if the subjects were restricted from violating the constraint it may inhibit learning in this environment. Similar problems may occur in other environments.

When running field experiments using homegrown values,[8] some institutions may lack control, so caution needs to be used to ensure that the field environment is what the researcher believes it is. For instance, both a 2^{nd}-price sealed-bid auction and an English auction theoretically provide truthful revelation of an individual's private value for an object in a SIPV environment. Thus, either should be able to be used to elicit truthful revelation of private values in the field. However, it may be that values are not private values for some field objects, but affiliated. An easy way of describing an item with affiliated values is to say that the item has both private and common value features. Consider a piece of artwork. Individuals have their own private values for a piece of artwork, but prices are also influenced by what others are willing to pay for the item. If an English auction is used in an affiliated values context, then it is unlikely that the researcher will get a truthful estimate of the participant's value before the auction. What the researcher would observe is the participant's value of the item conditional on the information the participant receives throughout the auction. In this context, a 2^{nd}-price sealed bid auction would be a better method of eliciting truthful revelation of value without the participant observing the additional information.

[6] The endogenous assignment of roles does not really carry over to the role of the buyer or consumer, as nearly everyone has practical experience in this role.

[7] Yes, this is the exact same problem. I realize they could just as easily move to the other side of the table or spin the pegboard around if it is movable. But what they do is actually physically move the discs to the goal state, and who am I to argue with children?

[8] Homegrown simply means the value the individual places on the item.

3.3 Differences in the task and environment

3.3.1 Nature of the task

I love this line – "Who cares if hamburger flippers violate expected utility theory?" The hamburger flipper's job description, satisfaction, and evaluation may not hinge upon whether or not he violates EUT. However, there is no "hamburger flipper theory" to my knowledge, and we know that hamburger flipper's are people too.[9] So even hamburger flippers should follow EUT. The point is as above – if subjects are placed in an unfamiliar task they may perform poorly relative to the theory. If one takes a hamburger flipper and places him in an abstract environment then it may be that the individual is unable to recall heuristics developed in an environment with context that he can use for the task. Thus, if these heuristics exist in the field but not in the lab then it may be the case that the results from the lab are misleading.[10]

3.3.2 Nature of the stakes

Another common criticism of lab experiments concerns the stakes. One criticism is that the subjects are playing with house money and that their behavior may be different than when playing with their own money. A second criticism concerns the size of the stakes in lab experiments. A typical lab experiment might have making decisions for 1-2 hours and receiving \$15-\$30.

To address both criticisms, consider the typical economic agent. The typical economic agent is assumed to have preferences and a utility function. The economic agent's goal is to maximize utility. There is typically no discussion about house money or how large utility is, simply that utility is to be maximized. Now, one thing that people often confuse is utility with profit, and they are two distinct concepts. Utility tends to capture ALL aspects of behavior, while profit tends to focus solely on monetary aspects. Thus, it may be that with house money a subject is more risk-loving in an experiment than he would be in reality because the subject is deriving utility from acting in a risk-loving manner. It may also be that the cognitive costs of the task are high relative to the benefit of the task, and thus the subject chooses to behave an indifferent manner in the experiment because he does not wish to derive disutility from thinking about the task at hand. This is where Smith's precept of dominance is of paramount importance – the goal is to minimize these extraneous factors.

As to the size of the stakes, there are many individuals who have found that increasing the size of the stakes leads to less variance in behavior, not different behavior. Thus, individuals focus more the higher the nature of the stakes (which is natural), but they do not alter behavior in many instances. Also, there is no guarantee that the stakes are "higher" in the field than in the laboratory. If an individual pays \$100 for an item in the field the exact amount of consumer surplus of utility the individual receives from the item is unknown – it may be that the individual pays \$100 for something the individual values at \$100.50. Finally, there is the paper by Hendel and Nevo (2006) that provides very telling evidence about the nature of the stakes in some "real-world" environments.[11] In this paper the authors provide summary statistics on laundry detergent purchases. The average household in their sample purchases laundry detergent every 43.7 days and spends \$4.38 per purchase. The average sales discount for laundry detergent is 67 cents, the 25^{th} percentile is 20 cents, and the 75^{th} percentile is 90 cents. These are decisions made by real people in the real-world, and they are apparently motivated by 67 cents every 43.7 days since approximately 40% of laundry detergent sold while on sale. That is less than \$6 per year. And lab experiments pay \$10, \$15, possibly even \$20+ per hour, with each individual decision possibly worth 67 cents. I believe what many people fail to realize is that most of the transactions individuals make are NOT for multi-billion or multi-million or multi-thousand or multi-hundred dollar deals. They are exactly the type where an individual saves 67 cents on a purchase of laundry detergent, and my conjecture is that most individuals would be in an unfamiliar context if making decisions involving hundreds of thousands of dollars.

[9] Recall the Kevin Federline Super Bowl commercial for Nationwide where he winds up working in a fastfood restaurant. He later apologized to the 12.8 million fastfood workers.
http://www.msnbc.msn.com/id/16949009/

[10] However, they may not be misleading about novice behavior in the field, as novices likely would not have had time to develop heuristics.

[11] The citation is Hendel and Nevo (2006), Measuring the Implications of Sales and Consumer Inventory Behavior, Econometrica, Vol. 74:6, pgs. 1637-1674.

3.3.3 Nature of the environment

It is also possible that the laboratory environment itself changes behavior. First, lab experiments are carried out in the lab, where certain stimuli that may occur naturally when undertaking a particular task may be missing. Second, subjects in a lab experiment know that they are in an experiment. People sometimes behave differently when they know they are being watched than when they may or may not be being watched with every decision recorded. Individuals may be especially prone to change their behavior when they know they are being watched and know the results the experimenter hopes to obtain. This is known as the Hawthorne effect. One might call it the "reality TV show effect" today, as most people are fairly certain that the individuals in the reality TV show do not act in the same manner when they are off-camera.[12] For all of these concerns, dominance is again important. If the reward to the task at hand dominates these other effects, then inferences about behavior in the lab should carry over to the "real-world".

[12] Of course, the individuals on reality TV are likely trying to set themselves apart from others in hope of securing future TV work.