NBER WORKING PAPER SERIES

DO NATURAL FIELD EXPERIMENTS AFFORD RESEARCHERS MORE OR LESS
CONTROL THAN LABORATORY EXPERIMENTS? A SIMPLE MODEL

Omar Al-Ubaydli
John A. List

Do Natural Field Experiments Afford Researchers More or Less Control than Laboratory
Experiments? A Simple Model
Omar Al-Ubaydli and John A. List
NBER Working Paper No. 20877
January 2015
JEL No. C9,C91,C92,C93

ABSTRACT

A commonly held view is that laboratory experiments provide researchers with more "control" than
natural field experiments, and that this advantage is to be balanced against the disadvantage that laboratory
experiments are less generalizable. This paper presents a simple model that explores circumstances
under which natural field experiments provide researchers with more control than laboratory experiments
afford. This stems from the covertness of natural field experiments: laboratory experiments provide
researchers with a high degree of control in the environment which participants agree to be experimental
subjects. When participants systematically opt out of laboratory experiments, the researcher's ability
to manipulate certain variables is limited. In contrast, natural field experiments bypass the participation
decision altogether and allow for a potentially more diverse participant pool within the market of interest.
We show one particular case where such selection is invaluable: when treatment effects interact with
participant characteristics.

Omar Al-Ubaydli
Department of Economics and Mercatus Center
George Mason University
omar@omar.ec

John A. List
Department of Economics
University of Chicago
1126 East 59th
Chicago, IL 60637
and NBER
jlist@uchicago.edu

Do Natural Field Experiments Afford Researchers More or Less Control than Laboratory Experiments? A Simple Model

Omar Al-Ubaydli and John A. List[1]

December 2014

Abstract

A commonly held view is that laboratory experiments provide researchers with more "control" than natural field experiments, and that this advantage is to be balanced against the disadvantage that laboratory experiments are less generalizable. This paper presents a simple model that explores circumstances under which natural field experiments provide researchers with more control than laboratory experiments afford. This stems from the covertness of natural field experiments: laboratory experiments provide researchers with a high degree of control *in the environment which participants agree to be experimental subjects*. When participants systematically opt out of laboratory experiments, the researcher's ability to manipulate certain variables is limited. In contrast, natural field experiments bypass the participation decision altogether and allow for a potentially more diverse participant pool *within the market of interest*. We show one particular case where such selection is invaluable: when treatment effects interact with participant characteristics.

1. Introduction

There exists a vociferous and voluminous discussion in the literature about the relative merits of conventional laboratory experiments and natural field experiments (Frechette & Schotter, 2015). One view that has gained widespread currency is that laboratory experiments afford the inquiring researcher superior control over the environment than do natural field experiments. For example, according to Falk and Heckman (2009): "The laboratory allows tight control of decision environments," while Camerer (2015) claims that: "There is little doubt that the quality of control is potentially very high in lab experiments," going on to describe the level of control in the lab as "extreme." In contrast, the prevailing view in the literature is that natural field experiments occur in environments over which the experimenter has modest control, constituting a drawback to their deployment as part of a scientific inquiry.

The main area of contention among contributors to the laboratory and natural field experiment debate focuses on the *generalizability* of the conclusions drawn from an individual experiment, which refers to the extent to which the results apply outside the immediate confines of the initial experiment. Loosely speaking, one group of scholars feels that the results from laboratory experiments are no less generalizable than their natural field experimental counterparts, and that therefore the superior control typically associated with the laboratory means that they should account for a substantial share of our empirical knowledge (Falk & Heckman, 2009). Another group feels that natural field experiments are more generalizable, and that in many settings, this benefit outweighs the drawback of having limited control, meaning that we should focus our scholarly energy on natural field experiments (Levitt & List, 2007; and the model in Al-Ubaydli and List, 2015).

In this paper, we ignore the generalizability debate to more closely scrutinize the pseudo-accepted claim that laboratory experiments provide researchers with greater control over the environment. Our argument focuses on the experimental subject pool: in some settings, prospective subjects may refuse to participate in a laboratory experiment (the reasons could vary from the potential subject regards the decisions over which they must surrender control as being too important, to simply they have a high value of time). In these cases, a natural field experiment—by virtue of its covertness—becomes the only way to run an experiment as informed consent is bypassed. Thus, while the laboratory experiment does indeed provide the researcher with greater control over what happens to subjects *once they agree to participate*, it provides the researcher with <u>less control</u> than a natural field experiment over who actually participates.

As an illustration, imagine a researcher who wants to investigate the nature and extent of discrimination against the disabled in product or labor markets. Assuming a normal recruitment budget, one could ask prospective employers or sellers of a good or service to participate in a laboratory experiment. This will likely result in high rates of rejection as the prospective subjects might regard their time as too valuable. Alternatively, the researcher can design a natural field experiment to reach out to the prospective employers or sellers in their natural environments (see for example Gneezy et al., 2012). The laboratory experimenter suffers from a lack of control over recruitment compared to the natural field experiment (and this leaves on the sideline important behavioral effects, such as whether a person might actually engage in discrimination if he knows he is being observed as an experimental subject).

Naturally, in many important settings, recruitment constraints are of little consequence to the value of the data. For example a scholar studying the effect of ambient heat on the performance of college students on IQ tests would be well-served by conducting a standard laboratory experiment where college students are recruited, and excluding employers, CEOs, and airline pilots does not undermine the value of the results. However, if in the aforementioned discrimination experiment, the scholar was attempting to gauge sellers' propensity to engage in

discrimination, then the experiment's scientific value is clearly enhanced by ensuring that the participants are actual sellers.

In this paper, we explore the participation decision using a simple game-theoretic model. We also formally articulate the econometric advantages of natural field experiments in this regard. Our main contribution is in demonstrating the error in the blanket claim that laboratory experiments provide researchers a greater degree of control than natural field experiments, and in exposing the econometric consequences of failing to account for the issue.

2. THE BENEFITS AND COSTS OF RESEARCH

Scientists conducting experimental research using human subjects require the ability to assign human subjects to different treatments. Most frequently, this control is exercised overtly, as in the case of a patient knowingly participating in a medical trial, or a college student participating in a psychology experiment. In the language of Harrison and List (2004), conventional laboratory, artefactual, and framed field experiments all fall into this category of study as the human subjects are aware of their participation in an experiment and the participation requires the subjects' informed consent.

Alternatively, in some studies, the scientist assigns the human subjects to different treatments covertly, and without the subjects' explicit consent. These constitute defining characteristics of natural field experiments (Harrison and List, 2004), and examples include US voters receiving a free subscription to a randomly chosen newspaper (Gerber et al., 2009) or Peruvian taxi drivers being assigned a passenger with a randomly assigned bargaining stance (Castillo et al., 2013).

The steps involved in running overt versus covert experiments carry important implications for the type of subjects that the scientist can secure as participants in the experiment, and how representative those subjects are of the broader population that the scientist is seeking to draw conclusions about. In particular, in overt experiments, the scientist needs to convince the human subject to willingly cede control over treatment assignment to the scientist, and this is most commonly achieved via the offer of financial compensation.

2.1. THE COST OF ENSURING A REPRESENTATIVE SAMPLE

Consider an environment where agents are organically assigned to a treatment in a wide range of decisions. For any given decision, if agents are passive, they accept whatever they are assigned and bear no decision-making cost, which includes cognitive and adjustment costs. The drawback of being passive is the risk of the assignment being ex post suboptimal from the agent's perspective. On the other hand, if agents are active, they investigate and possibly modify their treatment assignment to ensure that it is ex post optimal, while incurring decision-making costs.

3

This situation presents a classic tradeoff, and the optimal decision on whether to be active versus passive depends on a comparison of the costs of an inferior treatment assignment vis-à-vis the costs of making a considered decision, be they cognitive exertion costs, adjustment costs, or other costs (Smith & Walker, 1993).

A simple example would be seats on an airplane. At the booking stage, you can accept the airline's seat assignment and risk being in an undesirable seat. This constitutes passive decision-making. Alternatively, you can take steps to guarantee a desirable seat as part of an active decision on your part. These steps may involve financial cost in addition to the cognitive cost of making the choice.

Now imagine that an agent is a prospective subject in an experiment. Overt experiments require the agent to be a passive decision-maker, possibly (though not necessarily) as the result of an informed participation decision resulting from an offer of financial compensation. Covert experiments can in principle run whether the agent is actively or passively decision-making, but they are easier (and likely cheaper) to organize when the agent is passive since passivity affords the researcher greater latitude in manipulating the environment. Thus in both overt and covert experiments, passive decision-making is associated with a greater likelihood of participation.

A key factor in determining the likelihood of passive decision-making—and hence the size of the potential participant pool—is the importance of the different treatments to prospective human subjects: if the choice between treatments is considered very important to potential participants, then they will be unwilling to surrender control to nature or to the scientist even if offered considerable sums of money, creating recruitment problems. For example if a scholar wished to study the consequences of being late to work to a worker's promotion prospects, then she will likely struggle to convince executives to participate in the experiment since they value being on time to work so highly. The available budget will probably only be enough to secure the participation of very ill-disciplined workers, limiting the samples' representativeness to the greater population.

In contrast in some settings, prospective human subjects are virtually indifferent between treatments and are therefore happy to passively accept external treatment assignments, implying that obtaining a broad sample is straightforward. An example would be a scientist studying people's preferences between different brands of orange juice; a scientist who sets up a stand in a shopping mall and asks people to taste blindly would likely obtain a diverse sample with minimal effort.

For the remainder of this paper, when comparing overt and covert experiments, we restrict our focus to the ease with which each type of experiment permits the scientist to obtain a representative sample of participants. In the interests of parsimony, we will abstract from the broader generalizability debate that informs a scientist's decision between overt and covert experiments, noting that such issues have been discussed extensively (Frechette & Schotter,

2015). The arguments that we consider below are complementary to those in the broader generalizability debate.

By way of simple operationalization, a scientist interested in estimating a causal effect by randomly assigning human subjects to different treatments has three options: run an overt experiment at cost $c_L > 0$, run a covert experiment at cost $c_{F,A} > 0$ when the participant is an active decision-maker and $c_{F,P}$ when she is passive (where $c_{F,A} > c_{F,P}$), or decline to run an experiment thereby avoiding any cost. For notational clarity, we use the subscript L to denote overt experiments because we are using the term "laboratory experiment" to refer to the union of conventional, artefactual, and framed field experiments; and we use the subscript F to denote covert experiments because they constitute natural field experiments.

2.2. THE VALUE OF THE RESEARCH QUESTION

Whether or not an experiment is worth running at all depends on the value of the research question to the scientist. Some causal effects are of manifestly no interest to any scientist regardless of any estimation cost. For example, it is unlikely that many economists value knowing the causal effect of using blue versus black ink on driving license application forms. Conversely, the scientific community places great value on studies that can shed light on the most effective teaching methods for young children, or that help central bankers understand the relationship between interest rates and inflation.

One configuration of factors that typically yields a valuable research question is as follows.

1. The prevailing perception is that a certain causal effect is zero.
2. A scientist has a strong suspicion (due to her specialized knowledge) that the causal effect is in fact non-zero.
3. If the causal effect is demonstrated to be non-zero, people (including possibly policymakers) will update their priors and potentially change their behavior.

An illustration would be Scottish physician James Lind's 18th century experiment on a cure for scurvy. Prior to his experiment, the average sailor believed that consuming citrus fruits had a zero causal effect on the likelihood of suffering from scurvy. Lind's experiment—regarded as the first ever clinical trial—demonstrated otherwise, and paved the way for the Royal Navy's adoption of citrus fruits to guard against scurvy.

More generally, a common feature of much high-impact scientific research is finding a non-zero causal effect when people think it is zero (Ioannidis, 2005), or finding a zero causal effect when the community is convinced that it is non-zero (Young et al., 2008). In fact the incentive to disprove conventional views combined with the ability to selectively report data is a well-known source of publication bias (Young et al., 2008).

Let $\pi \geq 0$ denote the value to the scientist of conducting the research.

3. INDEPENDENT BENEFITS AND COSTS OF RESEARCH

We will consider a game that is highly stylized and at the same time seemingly excessively complicated. The latter is motivated by the desire to smooth the transition to the game in the next section, which is the more enlightening of the two.

3.1. GAME FEATURES

In the extensive-form game in Figure 1, Nature (N) plays first, determining whether or not the experiment is *Important* (I) or *Unimportant* (U) from a research perspective. Let $p_I = 1 - p_U$ denote the probability that nature plays *Important*. The result of nature's decision is the experimenter's (E) private information. The experimenter's information advantage is the result of both her idiosyncratic preferences and her investment in acquiring knowledge of the cutting-edge literature; however, this information asymmetry is irrelevant since it does not affect the prospective subject's payoffs.

Figure 1: The Experiment Game
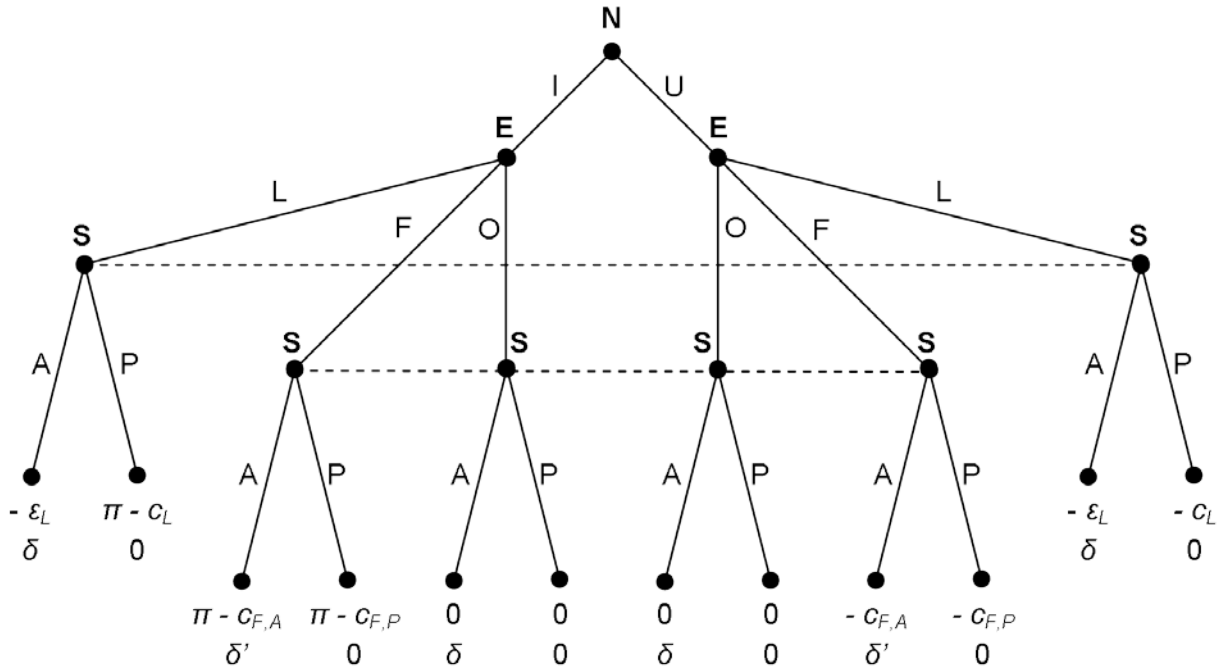
Figures below the terminal nodes are payoffs, with the higher figure denoting the experimenter's payoff and the lower figure denoting the subject's payoff

After observing nature's play, the experimenter has three options: a *Laboratory* (L) experiment, a natural *Field* (F) experiment, or *No Experiment* (O).

In the final stage, the subject (S) chooses between *Active* (A) and *Passive* (P) decision-making. In the case of a *Laboratory* experiment, playing *Active* corresponds to non-participation in the experiment.

In addition to being unaware of nature's play (the situation's research importance), the subject is also unable to distinguish between an experimenter who has opted to run a natural *Field* experiment and one who has chosen *No Experiment* since the former is run covertly. However he can determine when the experimenter has chosen to run a *Laboratory* experiment since that requires informed consent.

The experimenter's payoff is composed of the earnings from collecting data minus the costs of running an experiment. Regardless of the situation's research importance, if the experimenter chooses *No Experiment*, since no data are collected and no costs are incurred, her payoff is zero. In the event that an experimenter chooses *Laboratory* experiment and the subject responds by choosing *Active*, implying non-participation, the experimenter receives zero data-based earnings.

When an experimenter successfully collects data (*Laboratory* followed by *Passive*; or *Field*), her data-based earnings depend upon the research question's importance. When it is *Important*, she earns π, and when it is *Unimportant*, she earns zero.

As described above, *Laboratory* experiments cost $c_L > 0$; if a subject declines participation (plays *Active*), then the experimenter still incurs a cost $\varepsilon_L > 0$ reflecting sunk costs such as experimental design formulation and equipment purchases. Natural *Field* experiments cost $c_{F,A}$ when the subject plays *Active* and $c_{F,P}$ when the subject plays *Passive*, with the property $c_{F,A} > c_{F,P} > 0$ reflecting the greater ease with which natural field experiments are run when subjects are passive toward their treatment.

Subject payoffs are as follows. Playing *Active* earns the subject $\delta \neq 0$ compared to zero for playing *Passive*. Whether or not $\delta > 0$ depends upon the aforementioned tradeoff regarding the importance of treatment assignment: when the gains to being an active decision-maker are large compared to the decision-making choice, $\delta > 0$; otherwise, $\delta < 0$. Finally, in the event that the experimenter plays *Field* and the subject plays *Active*, the subject earns $\delta' \leq \delta$ as the experimenter's covert interventions blunt the subject's ability to optimize treatment assignment. We assume that $\delta > 0 \Leftrightarrow \delta' > 0$.

3.2. EQUILIBRIUM CHOICE OF EXPERIMENT

We now consider some of the game's potential equilibria. Our goal is not to provide an exhaustive list of the equilibria, nor will we delve into the conditions that guarantee a unique equilibrium. Rather, our goal is to consider a subset of the equilibria to understand some of the mechanisms that may arise as a way of illustrating and exploring the participation decision.

Lemma 3.2.1: The experimenter runs experiments only if the research question is *Important*.

Proof: If the research question is *Unimportant*, running *No Experiment* is a strictly dominant strategy. ∎

Lemma 3.2.2: The subject plays *Active* if and only if $\delta > 0$.

Proof: When $\delta > 0$, it is strictly dominant for the subject to play *Active*, and when $\delta < 0$, it is strictly dominant for the subject to play *Passive*. ∎

The first proposition examines one set of circumstances under which a natural *Field* experiment is preferable to a *Laboratory* experiment.

Proposition 3.2.1: When $\delta > 0$:

- *Laboratory* experiments are impossible
- The experimenter runs a natural *Field* experiment if and only if the research question is *Important* and the cost of a natural *Field* experiment on *Active* subjects is sufficiently low ($\pi > c_{F,A}$).

Proof: By Lemma 3.2.2, when $\delta > 0$, the subject will always play *Active*, rendering *Laboratory* experiments impossible. By Lemma 3.2.1, the experimenter will only consider running an experiment under *Important*. Playing *Field* is preferred to *No Experiment* if and only if $\pi > c_{F,A}$. ∎

Proposition 3.2.1 conveys the principle that when people care sufficiently about a decision such that they do not wish to surrender that decision to an experimenter, the only way to get them to participate in an experiment is to do so covertly. The experimenter will pursue this option if the research question is sufficiently interesting and the natural field experiment is sufficiently cheap.

A perfect illustration is Bertrand and Mullainathan (2004), where the authors wanted to examine the incidence of racial discrimination among companies at the hiring stage. Given a plausible budget, it would be difficult to obtain a representative sample (and in the extreme virtually no companies would wish to participate in a controlled experiment since their human resource departments are likely too busy with their day-to-day responsibilities). The authors opted for a natural field experiment where they sent fake CVs to employers and examined how call-back rates varied with respect to the applicant's name. Though we do not have hard evidence, the experimental design suggests that the experiment was modestly priced. That the paper was published in one of the leading journals in economics confirms that the research question was highly interesting.

> **Proposition 3.2.2**: When $\delta < 0$, the experimenter will run a *Laboratory* experiment if and only if:
>
> - The research question is *Important*
> - A *Laboratory* experiment is cheaper than a natural *Field* experiment run on *Passive* decision-making subjects ($c_L < c_{F,P}$)
> - A *Laboratory* experiment is sufficiently cheap ($\pi > c_L$).
>
> The experimenter will run a natural *Field* experiment if and only if:
>
> - The research question is *Important*
> - A natural *Field* experiment is cheaper ($c_L > c_{F,P}$)
> - A natural *Field* experiment is sufficiently cheap.
>
> If $\pi < \max(c_L, c_{F,P})$, the experimenter always runs *No Experiment*.

Proof: By Lemma 3.2.2, when $\delta < 0$, the subject will always play *Passive*, opening the door for both types of experiment. By Lemma 3.2.1, the experimenter will only consider running an experiment under *Important*. The choice between the two experiment types and not running an experiment depends upon a direct comparison of the costs. ∎

Proposition 3.2.2 conveys the principle that when prospective subjects are willing to participate in a *Laboratory* experiment, a straightforward horse race in cost terms decides the "winning" type of experiment. In many real-life situations, *Laboratory* experiments offer significant cost advantages, and so they will be preferred. However there are some cases where a natural *Field* experiment may turn out cheaper. For example in Falk (2007), the author sent out 10,000 charitable solicitation letters by mail, and he included gifts in some of them as an attempt to investigate the causal effect of a gift on contributions. This was almost certainly cheaper than arranging for 10,000 solicitees to come to a laboratory, receive a show-up fee and then participate in wave-after-wave-after-wave of sessions. Admittedly, the extra control *in the lab* may well have reduced data demands because Falk could have ensured that all participants actually open and read the letter, unlike many of the solicitees in the natural field experiment who presumably discarded the envelope without even opening. However if Falk's goal was to simulate what a charity would actually do, then the natural field experiment was certainly cheaper.

Note that depending on the precise component of the subject pool, a laboratory or natural field experiment may be favored. Thus, while laboratory experiments are often a cost-effective way to gather data where the subjects are college students, an experimenter looking to collect data on the same causal effect but with professionals as the subject may need to resort to a natural field experiment. For example, in Camerer (1998), the author studied the effect of manipulating betting odds at a horse-racing track by placing large bets. Insofar as the author was seeking

professional gamblers and a professional betting shop as participants, the natural field experiment was likely the cheapest approach (provided Camerer could get his bets retracted on time!).

4. STATISTICAL DEPENDENCE BETWEEN THE BENEFITS AND COSTS OF RESEARCH

In the previous section, while the importance of the research question was the experimenter's private information, it had no payoff implications for the subject. This was to capture the possibility that the importance of a research question does not vary systematically with the treatment assignment's importance to the subject. As a result, the subject's choice was decision-theoretic rather than strategic.

Prima facie, decisions that bear research importance also tend to be the ones that would warrant active decision-making by agents, and vice versa. For example, browsing a recent issue of the journal *Experimental Economics* (Volume 17, Issue 1) reveals papers that study the causal effect of permitting third party punishment and reward, the causal effect of parental background on children's preferences, and the causal effect of intermediaries on corruption, all of which are scenarios where one imagines that agents would be keen to be active rather than passive decision-makers.

Naturally, the relationship is not perfect. For example, List's (2003) study of the endowment effect reflected a scholar's interest in studying the effect of being offered a Kansas City Royals ticket stub first and then being offered the opportunity to exchange it for a Nolan Ryan certificate versus the same offerings in reverse. It is fair to say that most people do not care about being actively involved in this particular treatment assignment.

As a general rule, however, if the situations where a scientist wants to exercise control over agents' treatment assignment for research purposes tend to be the situations where agents themselves care enough about the decision that they do not want to relinquish their influence over treatment assignment, then scientists may face a problem convincing people to participate in experiments.

4.1. GAME FEATURES

Figure 2 shows the game in extensive form. It is isomorphic to the game considered in Figure 1, but with one key difference: the subject's payoff for selecting *Active* is $-\gamma$ when the research question is *Unimportant* (and $-\gamma'$ in the event that the experimenter selects *Field*, where $-\gamma' < -\gamma$).

Figure 2: The Evolved Experiment Game
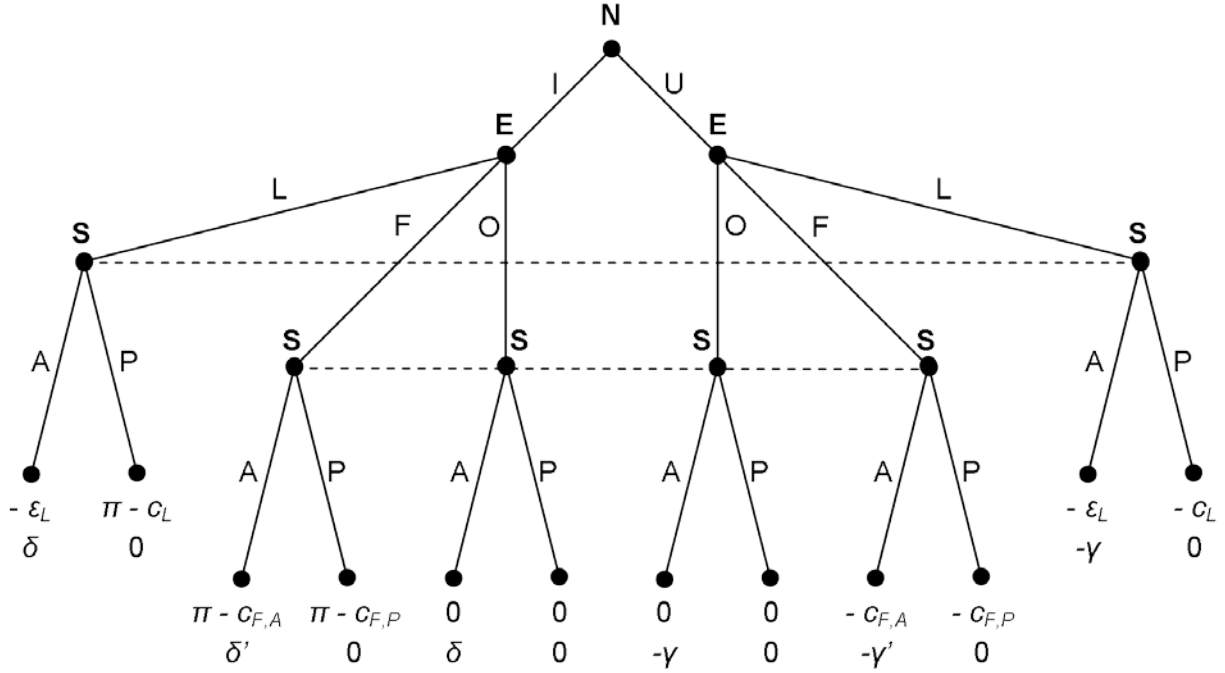
Figures below the terminal nodes are payoffs, with the higher figure denoting the experimenter's payoff and the lower figure denoting the subject's payoff

The experimenter's private information is now of interest to the subject since the subject's payoffs depend upon it. This version does not only capture the fact that there may be statistical dependence between the importance of the research question and the optimality of active decision-making. It represents a situation where the subject is imperfectly informed about the value of active decision-making, the experimenter is *better* informed, and the subject seeks to make inferences based on the experimenter's choice of experiment.

4.2. EQUILIBRIUM CHOICE OF EXPERIMENT

As before, we will consider a subset of the game's equilibria for illustrative and exploratory purposes. To construct a sequential equilibrium, we introduce the following notation for the subject's beliefs: $(\mu_{IL}, \mu_{IF}, \mu_{IO}, \mu_{UL}, \mu_{UF}, \mu_{UO})$ where in each element, the first letter in the subscript denotes nature's decision and the second letter denotes the experimenter's decision. The adding up constraints are $\mu_{IL} + \mu_{UL} = 1$ and $\mu_{IF} + \mu_{IO} + \mu_{UF} + \mu_{UO} = 1$.

Lemma 4.2.1: The experimenter runs experiments only if the research question is *Important*.

Proof: Playing *No Experiment* is the strict best response for the experimenter when the decision is *Unimportant* from a research perspective. ∎

Lemma 4.2.2: When $\delta > 0$, *Laboratory* experiments never occur.

Proof: Given Lemma 4.2.1, the subject knows that if the experimenter plays *Laboratory*, it is because the decision is *Important* from a research perspective. The subject's strict best response is *Active*; the experimenter can guarantee herself a higher payoff by playing *No Experiment*, and so she will never play *Laboratory* in the first place. ∎

Proposition 4.2.1a: If $\delta > 0$, $\gamma < 0$ and $\pi > c_{F,A}$, there exists a pure strategy sequential equilibrium where:

- The experimenter plays *Field* when the decision is *Important* from a research perspective and plays *No Experiment* when it is *Unimportant*
- The subject always plays *Active*
- The subject's beliefs are unrestricted when the experimenter plays *Laboratory*, and when the experimenter does <u>not</u> play *Laboratory*, the subject believes that he is at the *Important-Field* node with probability p_I and that he is at the *Unimportant-No Experiment* node with probability $1 - p_I$, i.e., $\mu_{IF} = p_I = 1 - \mu_{UO}$ and $\mu_{IO} = \mu_{UF} = 0$.

Proof: This is largely analogous to the proof of Proposition 3.2.1 since the initial conditions guarantee that playing *Active* is a strictly dominant strategy for the subject, which implies that it is optimal for the experimenter to play *Field* when *Important* and *No Experiment* when *Unimportant*. The absence of *Laboratory* experiments in equilibrium renders the beliefs at that node unrestricted. The beliefs at the remaining node reflect the equilibrium play of the experimenter, and they have no effect on the subject's play because *Active* is strictly dominant. ∎

As with Proposition 3.2.1, Proposition 4.2.1a conveys the principle that when subjects refuse participation in laboratory experiments, the only option is the field, and the option will be exercised if the research question is sufficiently interesting and the natural field experiment is sufficiently cheap.

The next equilibrium is substantively different to the equilibria in the simpler game considered in Section 3.

Proposition 4.2.1b: If $\delta > 0$, $\delta' p_I < \gamma'(1 - p_I)$, and $\pi > c_{F,P}$, then there is a pure strategy sequential equilibrium where:

- The experimenter plays *Field* when the decision is *Important* from a research perspective and plays *No Experiment* when it is *Unimportant*
- The subject plays *Active* when the experimenter plays *Laboratory* and the subject plays *Passive* when the experimenter does <u>not</u> play *Laboratory*
- The subject believes that the decision is *Important* with certainty if the experimenter plays *Laboratory* ($\mu_{IL} = 1$), and when the experimenter does <u>not</u> play *Laboratory*, the subject believes that he is at the *Important-Field* node with probability p_I and that he is at the *Unimportant-No Experiment* node with probability $1 - p_I$, i.e., $\mu_{IF} = p_I = 1 - \mu_{UO}$ and $\mu_{IO} = \mu_{UF} = 0$.

Proof: If, when the experimenter chooses <u>not</u> to play *Laboratory*, the subject responds by playing *Passive*, then it is optimal for the experimenter to play *Field* when the decision is *Important* if and only if $\pi > c_{F,P}$, which is true by assumption. The rest of the experimenter's strategy follows from Lemma 4.2.1 and Lemma 4.2.2. This strategy profile is also congruent with the proposed equilibrium beliefs of the subject, with the exception of those at the *Laboratory* nodes since they are off equilibrium and therefore unrestricted.

Finally, if the experimenter is playing *Field* if and only if the decision is *Important* and *No Experiment* otherwise, and the subject's beliefs reflect this, then the payoff for playing *Active* is $\delta' p_I - \gamma'(1 - p_I)$ compared to a payoff of zero for playing *Passive*. ∎

This equilibrium illustrates how—contrary to conventional wisdom (Falk & Heckman, 2009; Camerer, 2015)—natural field experiments can offer experimenters *more* control than laboratory experiments, specifically in terms of the participation decision. The intuition is as follows. Subjects' passive acceptance of their treatment assignment is necessary for a laboratory experiment. In some cases, the factors that tend to make a decision important from a research perspective also tend to make it important from the subject's own viewpoint in terms of his own welfare, and so he is unlikely to want to surrender his ability to determine his treatment assignment. Since experiments are costly to the experimenter, her invitation to a subject to participate in an experiment signals to the subject that this is an important decision that may be worth his consideration, pushing him toward declining the invitation. Were the experimenter to opt for a covert natural field experiment, then the subject may just classify the decision as one of the hundreds of decisions he passively makes on a daily basis, thereby implicitly consenting to the experimenter's treatment assignment.

As an example, imagine that scientists and non-scientists alike are unaware that vitamin D can be acquired from sunshine. Every day, people for the most part passively decide how much sunshine to expose themselves to. Next, a scientist theorizes about the capacity of the sun to deliver vitamin D, and seeks to run an experiment to investigate it. If she runs a laboratory

experiment then potential subjects will learn of the hypothesis and they will want to take steps to avoid being a control (low sunshine exposure). Since ensuring treatment status (high sunshine exposure) is easy outside the confines of an experiment, the scientist will probably find it very difficult to recruit subjects, and she will have to worry about some sort of systematic unobservable difference between those who choose to participate and those who do not. In contrast, were she to run a natural field experiment surreptitiously (setting aside the ethics of such a decision), then the participation decision will be effectively bypassed and her results more representative. In this sense, the natural field experiment has afforded her *more* control over the environment (specifically over who does and does not participate in the experiment) than a laboratory experiment.

An analogous model can be constructed where potential subjects opt out of experiments because they are averse to being randomized, which is a source of randomization bias (Harrison et al., 2009).

Proposition 4.2.2: If $\delta < 0$ and $\gamma > 0$, the experimenter will run a *Laboratory* experiment if and only if

- The research question is *Important*
- A *Laboratory* experiment is cheaper than a natural *Field* experiment run on *Passive* decision-making subjects ($c_L < c_{F,P}$)
- A *Laboratory* experiment is sufficiently cheap ($\pi > c_L$)

The experimenter will run a natural *Field* experiment if and only if:

- The research question is *Important*
- A natural *Field* experiment is cheaper ($c_L > c_{F,P}$)
- A natural *Field* experiment is sufficiently cheap.

If $\pi < \max(c_L, c_{F,P})$, the experimenter always runs *No Experiment*.

Proof: Analogous to Proposition 3.2.2. ∎

The conditions $\delta < 0$ and $\gamma > 0$ refer to a case where treatment assignment is not worth the subject's consideration regardless of the decision's research importance. In this case, the act of seeking the subject's informed consent as part of a laboratory experiment does not signal anything about the decision to the subject, and he is happy to passively accept assignment to control/treatment.

For example, if a company's marketing department wishes to conduct a randomized control experiment to compare the effectiveness of two of its television commercials, then the research team is unlikely to encounter the problems captured by Proposition 4.2.1 since the decision (control versus treatment commercial) is of manifestly of trivial importance to potential subjects.

As such, given the considerable cost advantage in assessing consequences that a laboratory experiment affords the research team, they will likely opt for one.

5. ARTICULATING THE CONCLUSIONS ECONOMETRICALLY

Proposition 4.2.1 has potentially important econometric implications: it suggests that in certain cases, relying on laboratory experiments can create gaps in our estimates of causal effects due to heterogeneity among subjects, and that these gaps may force scholars into precarious extrapolation. To demonstrate this more formally, we introduce some econometric terminology. The following draws heavily from Al-Ubaydli and List (2015), which is a version of more conventional models of causal inference (Heckman, 2005) that has been adapted to compare laboratory and field experiments.

Let Y be a random variable, denoted the *dependent variable*, whose realizations are in $S_Y \subseteq \mathbb{R}$; let X be a random variable, denoted the *explanatory variable of interest*, whose realizations are in $S_X \subseteq \mathbb{R}$; and let Z be a random vector, denoted the *additional explanatory variables*, whose realizations are in $S_Z \subseteq \mathbb{R}^k$. Further, Z contains all the explanatory variables (apart from X) that have an impact on Y, and they need not be observable.

In the all causes model (Heckman 2000), (X, Y, Z) are related according to the function $f: S_X \times S_Z \to S_Y$. Each $(x, x', z) \in S_X \times S_X \times S_Z$ is denoted a *causal triple*. The *causal effect* of changing X from x (control) to x' (treatment) on Y given $Z = z$ is described by the function $g: S_X \times S_X \times S_Z \to \mathbb{R}$, where:

$$g(x, x', z) = f(x', z) - f(x, z)$$

One central goal of empirical research is to estimate causal effects. Randomization is desirable because, in the absence of the ability to hold Z constant as X is varied, it allows the researcher to at least ensure that any variation in Z is statistically independent of variation in X.

Consider a family of causal triples $\{g(x, x', z)\}_{z \in U_Z \subseteq S_Z}$ that an investigator wants to estimate, where z is unidimensional. Z can be thought of as a potentially observable individual-level characteristic, such as preferences or IQ. As in the model above, individuals can influence their realization of X, i.e., their assignment to control/treatment by being *Active* rather than *Passive* decision-makers.

If, after deciding to conduct an experiment, the researcher chooses to conduct it covertly via a natural *Field* experiment, then inference will proceed as normal and the desired family of causal effects will be estimated.

On the other hand, should the investigator publicize her intention to conduct the experiment (a *Laboratory* experiment) due to (for example) ethical concerns, then she has to worry about

subjects not participating as a result of knowing about the experiment. Suppose some subset $U_Z' \subset U_Z$ decides to play *Active*, meaning that the investigator cannot estimate the causal triples for this subset. Inference for the remaining group, $U_Z \backslash U_Z'$, remains valid as they play *Passive*. One way to model this would be to assign the group U_Z' a value $\delta > 0$ and to assign the group $U_Z \backslash U_Z'$ a value $\delta < 0$.

As a consequence, the researcher will be forced to update her priors on causal triples associated with U_Z' by extrapolating/interpolating from $U_Z \backslash U_Z'$. In practice, this will be rendered even more precarious by the possibility that Z is unobservable, meaning that the researcher will be forced to assume that the causal triple is simply unaffected by the participation decision. In the case when $U_Z = \{z_1, z_2\}, U_Z' = \{z_2\}$, the extrapolation bias, which we term Treatment Specific Selection Bias, will be:

$$B = g(x, x', z_2) - g(x, x', z_1)$$

This corresponds to the bias that arises from what is commonly referred to as heterogeneous treatment effects (Heckman, 2005). An example would be Jorenby et al. (1999), where the authors wanted to compare the effectiveness of various methods of smoking cessation. They advertised for participants in local media, all of whom gave their informed consent. One year after the start of the experiment, subjects using sustained-release bupropion reported a smoking abstinence rate of 30% compared to 16% for a placebo group. What does this imply about the success rate of sustained-release bupropion for smokers who did not volunteer for the experiment—a figure that is surely of interest to policymakers? Given that non-participants include many people who do not actually wish to quit smoking, and people who might be too ill (from smoking) to participate in the experiment, extrapolating is clearly a precarious exercise.

6. DISCUSSION

There are many dimensions on which to compare laboratory and natural field experiments. For example much of the debate has centered on the impact of scrutiny on behavior, especially in studies of social preferences (Levitt & List, 2007). In the interests of parsimony, we have abstracted from these fundamentally important issues, and direct interested readers to the recent volume edited by Frechette and Schotter (2015). For the purposes of this paper, we are differentiating between laboratory and natural field experiments exclusively on their ability to secure participants.

Our paper's insights regarding the choice between a laboratory and a natural field experiment can be summarized as follows.

1. Laboratory experiments require informed consent, whereas natural field experiments do not.
2. Acquiring informed consent is often easier and cheaper than designing and implementing a covert (field) experiment, in which case a laboratory experiment will be favored. However

sometimes informed consent is relatively costly to acquire and natural field experiments will be favored on cost grounds alone.

3. Sometimes, acquiring informed consent is impossible because prospective subjects regard the treatment assignment as too important to be left to an experimenter, and so a covert (field) experiment will be the only option for data collection. In that case, assuming it is cheap enough, a natural field experiment will be selected.

4. Sometimes, the act of proposing an experiment to a prospective subject signals the importance of treatment assignment to the prospective subject, making him refuse to participate in a laboratory experiment, and leaving a natural field experiment as the only option.

5. When experiments are too expensive or the research question is insufficiently valuable, no experiment will be conducted.

Conventional wisdom in the experimental methodology literature is that laboratory experiments afford researcher greater control over the environment, and that this benefit has to be compared to the generalizability-related disadvantages of laboratory experiments, such as scrutiny of subjects or requiring subjects to assume unfamiliar roles. On the back of the arguments presented in this paper, we regard the expression "greater control" as being misleadingly broad. We believe that a more precise way of describing the benefits of laboratory experiments is to say that they afford experiments greater control *over the physical environment and the nature of permissible interactions*, but that they potentially offer researchers less control *over the nature of the participants*. This deficiency is accentuated by the possibility that prospective subjects deduce the benefits of non-participation by the act of a scientist initiating an investigation, since scientists tend to investigate important choices.

It is fair to say that in some situations, when choosing between laboratory and natural field experiments, the arguments presented in this paper are swamped by other factors, such as the role of stakes or subject pool composition (or if the scholar would like to use induced values to explore the hypotheses of interest (see, e.g., List, 2004)). However experiments are being deployed in an ever-widening range of environments, and their role in policy debates is expanding. As such, the mechanisms that we have described are of import, both at the experimental design stage and at the data interpretation stage.

Our study also highlights the important ethical questions faced by researchers at the experimental design stage. Is it acceptable to opt for a natural field experiment precisely because you know that disclosing the experiment will lead to non-participation? Questions such as these reinforce the importance of the oversight provided by human subject review boards (for more on this issue we direct the reader to List (2008; 2009).

Finally, one might imagine using laboratory experiments to explore how important, empirically, our conceptual arguments are in practice. This is the direction that an interesting new study due to Cleave et al (2013) takes. The authors measure risk and trust preferences for a group of

students who both select into the laboratory experiment using common recruitment procedures and those who did not. They report some behavioral similarities and some differences.

More generally, we hope that our study leads to a general practice whereby the experimenter uses different recruiting incentives and measures if there are subject differences—this can be as simple as liberal changes in the show up fee. This approach represents a viable means to resolve appropriately the selection issues discussed herein and provide empirical content to an important aspect of control.

REFERENCES

Al-Ubaydli, O., List, J. (2015). On the Generalizability of Experimental Results in Economics. In Frechette, G., Schotter, A. (Eds): The Methods of Modern Experiments. Oxford Univesity Press.

Bertrand, M., Mullainathan, S. (2004). Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. American Economic Review, 94, p991-1013.

Camerer, C. (1998). Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting. Journal of Political Economy, 106, p457-482.

Camerer, C. (2015). The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List. In Frechette, G., Schotter, A. (Eds): The Methods of Modern Experiments. Oxford Univesity Press.

Castillo, M., Petrie, R., Torero, M., Vesterlund, L. (2013). Gender Differences in Bargaining Outcomes: A Field Experiment on Discrimination. Journal of Public Economics, 99, p35-48.

Cleave, B., Nikiforakis, N., Slonim, R. (2013). Is there selection bias in laboratory experiments? The case of social and risk preferences. Experimental Economics, 16(3), 372-382.

Falk, A. (2007). Gift Exchange in the Field. Econometrica, 75, p1501-1511.

Falk, A., Heckman, J. (2009). Lab Experiments Are a Major Source of Knowledge in the Social Sciences. Science, 326, p535-538.

Frechette, G., Schotter, A. (2015). The Methods of Modern Experiments. Oxford Univesity Press.

Gerber, A., Karlan, D., Bergan, D. (2009). Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions. American Economic Journal: Applied Economics, 1, p35-52.

Gneezy, U., List, J., Price, M. (2012). Toward an Understanding of Why People Discriminate: Evidence from a Series of Natural Field Experiments. NBER Working Paper 17855.

Harrison, G., Lau, M., Rutstrom, E. (2009). Risk Attitudes, Randomization to Treatment, and Self-Selection into Experiments. Journal of Economic Behavior and Organization, 70, p498-507.

Harrison, G., List, J. (2004). Field Experiments. Journal of Economic Literature, 42, 1009-1055.

Heckman, J. (2005). The Scientific Model of Causality. Sociological Methodology, 35, p1-97.

Ioannidis, J. (2005). Why Most Published Research Findings Are False. Public Library of Science: Medicine, 2, e124.

Jorenby, D., et al. (1999). A Controlled Trial of Sustained-Release Bupropion, a Nicotine Patch, or Both for Smoking Cessation. New England Journal of Medicine, 340, p685-691.

Levitt, S., List, J. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? Journal of Economic Perspectives, 21, p153-174.

List, J. (2003). Does Market Experience Eliminate Market Anomalies? Quarterly Journal of Economics, 118, p41-71.

List, J. (2004). "Testing Neoclassical Competitive Theory in Multilateral Decentralized Markets," Journal of Political Economy, 112(5): 1131-1156.

List, J. (2008), "Informed Consent in Social Science," Science, (2008), 322(5902), p. 672.

List, J. (2009), "The IRB Is Key in Field Experiments," Science, (2009), 323(5915), pp. 713-714.

Smith, V., Walker, J. (1993). Monetary Rewards and Decision Cost in Experimental Economics. Economic Inquiry, 31, p245-261.

Young, N., Ioannidis, J., Al-Ubaydli, O. (2008). Why Current Publication Practices May Distort Science. Public Library of Science: Medicine, 5, e201.