The Role of Theory in Field Experiments[1]

David Card, Stefano DellaVigna, Ulrike Malmendier
UC Berkeley and NBER

February 8, 2011
Preliminary Draft for JEP special issue on Field Experiments

I. Introduction

When it comes to role of theory, empirical microeconomists are torn. On the one hand, we devote a large fraction of our graduate courses to models of consumer and firm decision-making, and to the interactions that determine market equilibrium. On the other hand, it is not always obvious how these theories are relevant to empirical research. Outside the academy, policy-makers and business leaders often demand "basic facts" and simplified policy guidance with little or no concern for theoretical nuances.

How then do empirical economists negotiate between the demands for theory and "facts"? In this paper, we focus on the role of theory in the rapidly-growing area of field experiments. We take an empirical approach and quantify the role of theoretical modeling in all published field experiments in the top five economics journals from 1975 to 2010. Specifically, we propose a new classification of experimental studies that captures the extent to which the experimental design is linked to economic theory. We distinguish between four types of studies: studies that do not contain an explicit model (*Descriptive*); studies that test a single model-based hypothesis (*Single Model*); studies that test competing model-based hypotheses (*Competing Models*); and studies that estimate structural parameters in a completely specified model (*Parameter Estimation*).

Applying the same classification to laboratory experiments, we show that theory has played a more important role in the laboratory than in field experiments. We discuss in detail three sets of field experiments that illustrate the potential promise and the pitfalls of a tight link between experimental design and theoretical underpinnings.

II. Quantifying the Role of Theory in Field Experiments

[1] We thank for helpful comments the editors (David Autor and John List), as well as Iwan Barankay, Glenn Harrison, Matthew Rabin, David Reiley, and participants in Berkeley, at the Wharton Conference on Field Experiment, and at the 2011 ASSA conference in Denver. We thank Ivan Balbuzanov and Xiaoyu Xia for excellent research assistance.

a. Defining Field Experiments

The use of "experimental" techniques -- random assignment of a manipulation of interest -- came relatively late to economics.[2] In the last 15 years, however, randomized experiments conducted in field settings have become more common, and in 2010 field experiments represent about 3 percent of the articles published in the top journals. As in other areas of applied economics, the role of theory ranges from "almost none" to fully model-based investigations. Nevertheless, there is a widespread perception that many experimental studies – particularly field-based random-assignment studies – are "black box" evaluations that provide only limited evidence on theoretically-relevant mechanisms (see, e.g., Deaton, 2010).

To assess the actual importance of theoretical modeling in field experiments, and the relative role in field versus laboratory experiments, we collected data on all experimental studies published in the top five economics journals – the *American Economic Review*, *Econometrica*, the *Journal of Political Economy*, the *Quarterly Journal of Economics*, and the *Review of Economic Studies* – over the 35-year period from 1975 to 2010. We screened all articles excluding comments, notes, and articles in the annual *Papers and Proceedings Issue* of the *American Economic Review*, and identified two sets of studies -- laboratory and field experiments.

A first issue that arises in such an exercise is the delineation of what qualifies as an "experiment." We restrict attention to studies based on the random assignment of a purposeful "treatment" or manipulation.[3] This definition includes social experiments funded by governmental agencies, such as the Moving To Opportunities study (Kling, Liebman, and Katz, 2007), randomizations induced by a researcher, such as studies of sports card markets (List, 2003), and randomizations induced by a firm for its own research or marketing purposes (Nagin et al., 2002). The definition excludes, however, many interesting studies that are conventionally viewed as experimental. For example, several self-described "field experiments", including Bandiera, Barankay, and Rasul (2007; 2009) and Chetty, Looney and Kroft (2009) evaluate a manipulation but do not

[2] According to Forsetlund, Chalmers, and Bjorndal (2007), the earliest documented use of randomization in the social sciences in 1928. In economics, we are unaware of any study using random assignment prior to the negative income tax experiments in the 1960s (see Greenberg and Shroder, 2004).

[3] We include as "random assignment" studies where treatment is deterministically assigned in a way that can be viewed as equivalent to random, such as assigning every second name in a list.

use random assignment. Other important studies (e.g., Angrist, 1990; Sacerdote, 2001) exploit random variation that was created for other purposes. By imposing the requirement of random assignment of a purposeful manipulation we do not mean to criticize studies that use non-random treatments, or that rely on incidental randomization. Rather, we use these criteria to narrow our focus to studies that are closest in spirit to the *randomized clinical trials* (RCT's) used in medicine and other sciences. Advocates of randomized experimental studies often point to RCT's as a gold standard for scientific evidence, while critics tend to emphasize the limitations of RCT's (e.g., Heckman and Smith, 1995).

We include papers that re-analyze data from previous experiments, provided that the study uses the original micro data (e.g., Lalonde, 1986). In the terminology of Harrison and List (2004) we include both "natural field experiments" (in which the participants have no knowledge of being involved in an experiment) and "framed field experiments" (in which the participants are aware that they participate in an experiment).

b. Classification of Studies

Within this universe of studies, we classify the role of economic theory using a simple four-way scheme that we believe captures the centrality of economic theory in a particular study[4]:

- *Descriptive (D) – studies that do not formally specify a model in the paper.*
- *Single Model (S) – studies that lay out a formal model and test one (or more) qualitative implications of the model against a null.*
- *Competing Models (C) - studies that spell out two or more alternative models with at least one contrasting qualitative implication and test between them on the basis of this implication.*
- *Parameter Estimation (P) – studies that specify a complete data generating process for (at least some subset of) the observed data and obtain estimates of structural parameters of the model.*

[4] The classification in Harrison and List (2004) focuses, instead, on experimental control and artificiality of the setting. We view our classification as complementary.

To illustrate our classification system, we selected four examples from the recent literature that are broadly representative of the four classes (see Table 1). An example of a *descriptive* field experiment is the study by Miguel and Kremer (2004). The authors evaluate a deworming treatment program operated by a non-profit organization in Kenya. The program contains several elements, including drug treatment and education, and was designed to reduce infection rates and improve other outcomes, including school attendance and educational achievements. The paper provides no formal model for the experimental program impacts, though it does discuss the expected effects on health and education outcomes as well as possible channels for these effects, including social spillover effects.

The second type of field experiment is studies that include a formal model of the experimental impact, and then evaluate the predictions of this model (*Single Model*). An illustrative example is the paper by Nagin et al. (2002). It includes a simple but formally specified model that isolates the response of a key endogenous variable (the number of "questionable" calls claimed by a sales associate) to a manipulation of interest (the monitoring rate of questionable calls). The qualitative prediction of the model is then tested contrasting various treatment groups.

In many cases the distinction between *descriptive* and *single-model* studies is rather arbitrary. Consider, for example, a field experiment designed to test some implications of a well known model. If the paper contains at least one mathematical equation we assign the paper to the *single model* category.[5] If not, we classify the study as *descriptive*. This distinction may seem especially troubling to researchers who – like us – have had to remove the formal statement of a model from a paper to satisfy an editor or referee.

Nevertheless, we believe that the presence of a mathematical statement of the model is a useful (if crude) indicator of the importance of economic theory in the paper. While descriptive papers often refer to a model in general terms (for example, references to a model of peer effects in Miguel and Kremer), the formal statement of the model clar-

[5] We did not assess the logical completeness of the formal model specification: we simply checked for the presence of at least one line of offset mathematical text. We also did not count statistical models or statements of payoffs in laboratory experiments.

ifies the underlying assumptions and details the exact form of the model that is purported to apply to the empirical setting.[6]

A criticism of studies that test a single model is that the results provide little guidance in the event that the model is rejected: Which of the assumptions does the data reject? Would alternative models have fared differently? A parallel issue arises when the model is *not* rejected: Competing models may make the same prediction, so simple "one sided" tests do not distinguish between theories (Rabin, 2010). A text-book example for the latter case is provided by Becker (1962): finding that a demand curve is downward sloping does not provide evidence of utility maximization; even when agents choose randomly demand curves are downward-sloping as long the budget constraint is sometimes binding.

These concerns are partially addressed by *competing model* studies that lay out two or more competing models, with differing predictions for the response to a manipulation. The study by Fehr and Goette (2007), for example, compares a standard intertemporal labor supply model against an alternative model with reference-dependent preferences. The two models have similar predictions for the response of earnings to a short term change in the effective wage rate, but differing predictions for effort per hour: effort increase under the standard model, decrease under reference dependence. The latter predictions provide the basis for a test between the models.

[6] A useful case to consider is the influential set of findings on the disposition effect (Odean, 1999), i.e., on the propensity to sell stocks that are "winners" rather than "losers" compared to the purchase price. Odean uses a graph and an intuitive explanation to suggest that the phenomenon is explained by prospect theory. However, Barberis and Xiong (2009) show that once one actually writes down an explicit model of prospect theory in asset prices, the disposition effect is not generally predicted by the model. In this case, the intuitive explanation had focused on the concavity and convexity of the value function, but had neglected the impact of the kink at the reference point.

Table 1: Classification Examples

Study	Description	Classification
1. E. Miguel and M. Kremer "Worms: Identifying Impacts on Education and Health in the Presence of Treatment" *Econometrica* , 2004	Evaluation of deworming treatment program in Kenya. School-level assignment. Treatment delayed at control groups.	**Descriptive**
2. D. Nagin et al. "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment" *American Econ. Rev,* 2002	Random assignment of monitoring rate of call-center employees. Center-level assignment. Model of optimal cheating predicts greated cheating when monitoring is reduced.	**Single Model**
3. E. Fehr and L. Goette "Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment" *American Econ. Rev,* 2007	Random assignment of temporary increase in piece rate for bicycle messengers. Neoclassical model of intertemporal labor supply contrasted with reference-dependent preferences.	**Competing Models**
4. P. Todd and K. Wolpin "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment toValidate a Dynamic Behavioral Model of Child Schooling and Fertility" *American Econ. Rev,* 2006	Random assignment of schooling subsidies. Village-level assignment. Dynamic structural model of fertility and schooling fit to control group and used to forecast experimental impacts.	**Parameter Estimation**

Notes: Studies selected and summarized by authors. See text for description of relevant universe of studies, and classification system.

The fourth category is field experiments with fully specified models with *parameter estimation*. The estimation of the underlying parameters of the model allows for welfare and policy evaluations that are not possible otherwise. An interesting example of this approach is Todd and Wolpin (2006), who specify a dynamic choice model for schooling and fertility choices of families in rural Mexico. They estimate the model parameters using data from the control group of the Progresa experiment, and then compare the predicted responses of these families to the subsidies offered to the experimental treatment group. Since the predicted and actual responses are relatively similar, they conclude that the model provides an adequate description of the behavior of families in the experimental population.

III. The Role of Theory in Experiments: The Last 35 Years

In this section we turn to a quantitative analysis of the role that theory has played in field experiments in the past 35 years. To provide a useful contrast, we also classified all laboratory experiments published in top-five journals, including laboratory-like experi-

ments conducted in a field environment (and labeled "artifactual field experiments" in Harrison and List (2004)).

Figure 1 displays a count of all published field and laboratory experiments under this definition. Between 1975 and 1984, seven field experiment were published in top-five journals, six of which are analyses of the NIT experiments discussed in Section VI. Between 1985 and 1994, four more field experiments were published, including the Blank (1991) study of the impact of double anonymity in the refereeing process, and the LaLonde (1986) re-evaluation of training programs. Since 1995, the number of field experiments published has increased steadily and the variety broadens, ranging from behavioral topics to development economics. After 2005, the number of articles published is 8-10 per year. Over our entire 35 year sample period, a total of 86 field experiments have been published in the top-five journals.

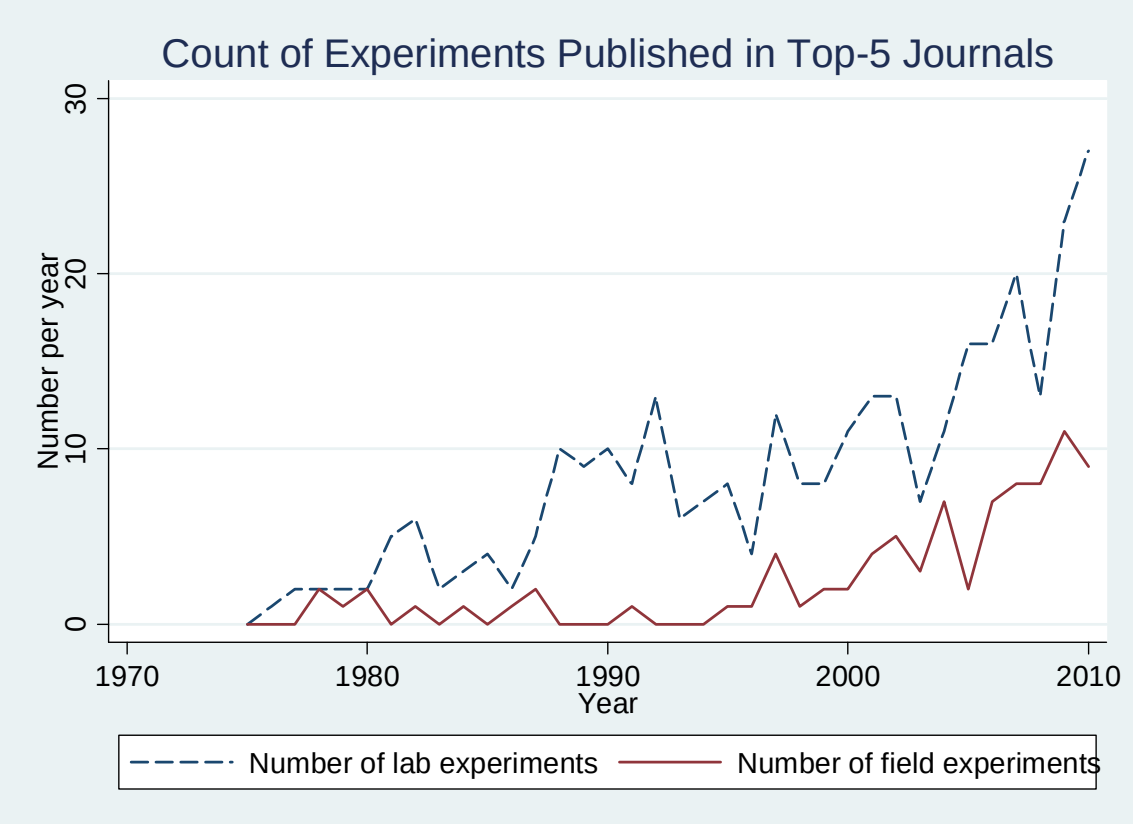

Figure 1. Number of laboratory and field experiments published in top-five journals from 1975 to 2010

Laboratory studies are significantly more frequent. In every single year since 1981, more laboratory experiments than field experiments were published in the top-five journals. Between 1985 and 1995, the number per year ranges from 5 to 10, yielding a total of 82 laboratory experiments, compared to only five field experiments in the same period. The annual number of laboratory experiments increased to 15-25 articles per year from 2005 to 2010. Indeed, in the most recent year (2010), laboratory experiments account for 10 percent of all top-five journal articles, compared to 3 percent for field experiments. The total number of laboratory experiments in our sample is 309, three and a half times the number of field experiments.

The *American Economic Review* accounts for 54% of all laboratory studies in our data, followed by *Econometrica* (19%) and the *Quarterly Journal of Economics* (13%). Field experiments are more evenly distributed across journals, with the *American Economic Review* (35%) and the *Quarterly Journal of Economics* (27%) publishing the most, followed by *Econometrica* (17%). Within each of these journals, the trends over time are similar to the ones documented in Figure 1. (Corresponding figures for each journal are in the Online Appendix).

How many of these experiments fall into each of our four categories? Figure 2 shows the numbers in each category for the first 10 years (1975-84), and the subsequent five-year periods (except for 2005-2010, which is a six-year period). Interestingly, the field experiments published in 1975-84 are all model-based. These papers are almost all analyses of the NIT experiments on the basis of a labor supply model (see Section VI). The few field experiments published thereafter in 1985-89 and 1990-94 are all descriptive; so too are the nine field experiments published from 1995 to 1999. Among the 21 field experiments published in the 2000-04 period, 18 are descriptive while 3 have a higher theoretical content (as judged by our criteria): 1 experiment with Single Model, 1 experiment with Competing Models, and two studies with Parameter Estimation. The first published field experiment with a more explicit theoretical framework excluding the NIT experiments is the Nagin et al. (2002) paper described above (in the *American Economic Review).*

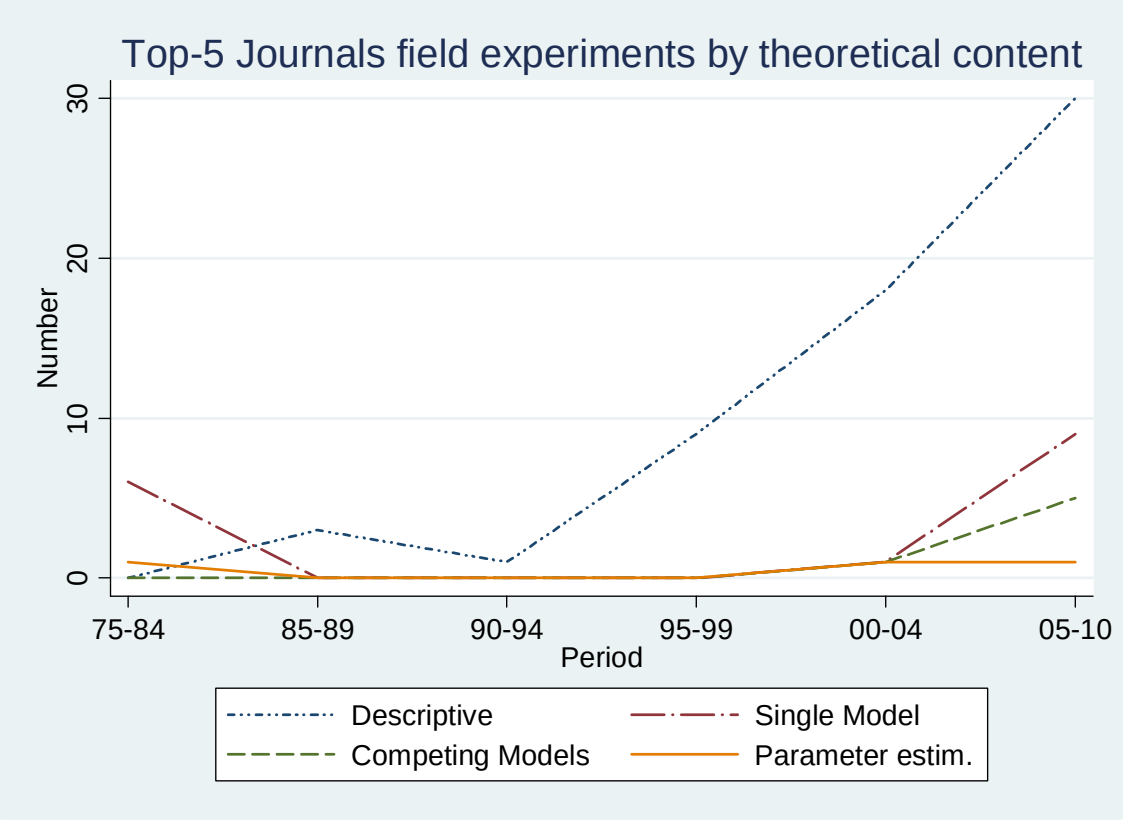

Figure 2. Field experiments by theoretical content

In the most recent period, 2005-2010, theory has played a more important role in field experiments, with nine experiments with Single Model, five with Competing Models, and one study with Parameter Estimation. Still, the dominant category remains *Descriptive*, with 30 articles.

Overall, 71% of the 86 field experiments published in top-five journals are descriptive, 19% contain a single model, 7% contain competing models, and only 3% of field experiments contain a model with parameter estimation.

The patterns are quite similar across journals, including *Econometrica* and the *Review of Economic Studies*. While empirical papers in these two journals are in general more likely to include models, in the case of field experiments the models are typically statistical, rather than economic models.

This pattern is quite different for laboratory experiments, as shown in Figure 3. While the descriptive type of experiments has typically been, and remains, the most

common type of laboratory experiment, model-based experiments (either with a single model or with competing models) were relatively common since the beginning, and have remained so. The main discernible trend in the last decade is an increase in the number of laboratory experiments with parameter estimation. This latter category includes, among others, the estimation of Quantal Response Equilibria models, of models of k-levels of thinking, and of time and risk preferences.

Overall, it is clear that the role of explicit theoretical models is very different in laboratory than in field experiments: 26% of the laboratory experiments contain a single model, 9% contain competing models, and 19% of papers contain a model with parameter estimation, while only about one-half (46%) are descriptive in nature.

These patterns differ somewhat by journal. In particular, *Econometrica* and the *Review of Economic Studies* have a higher incidence of model-based experiments than the other journals: in the last decade, the most common type of laboratory experiment is one with Parameter Estimation in these two journals.

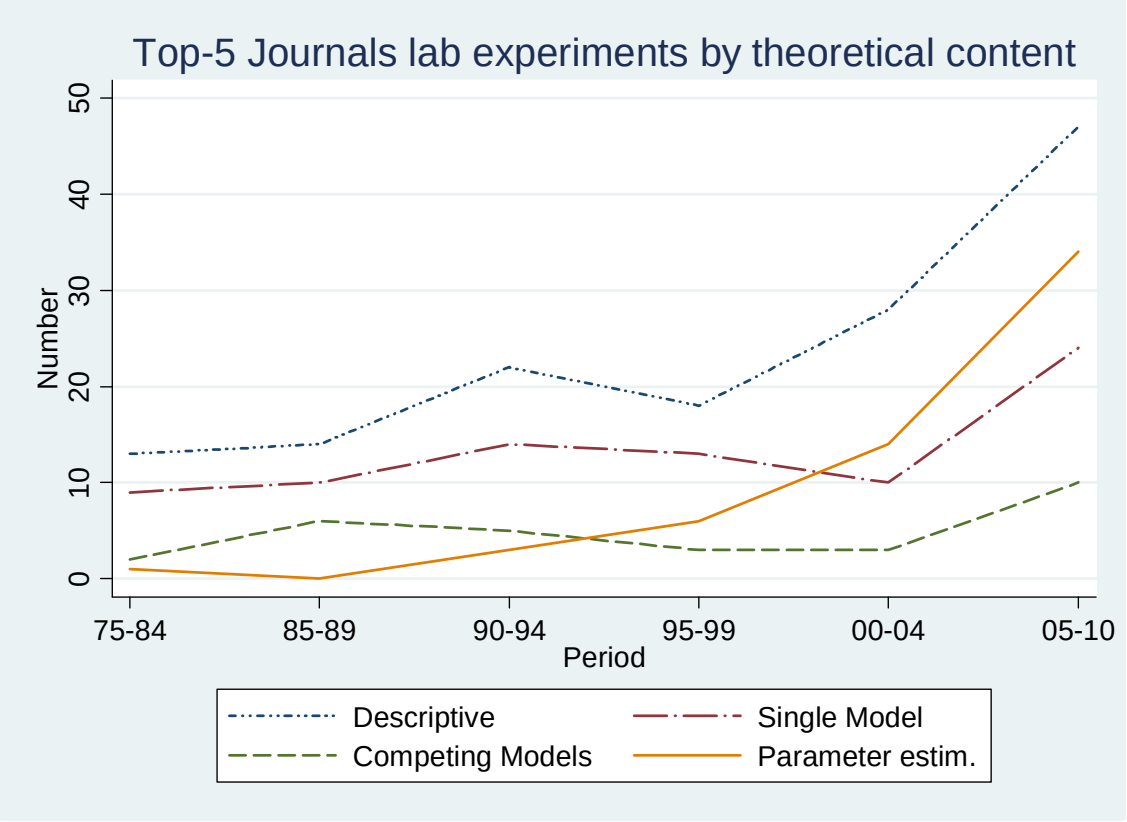

Figure 3. Laboratory experiments by theoretical content

. This brief historical review shows just how different the role of theory is in laboratory and field experiments. Models have always played a key role in laboratory experiments, with an increasing trend. Field experiments have been largely descriptive, with only a recent increase in the role for models. In the two journals typically most devoted to theory, *Econometrica* and the *Review of Economic Studies*, in the last decade the most common laboratory experiment is an experiment with Model and Parameter Estimation, while the most common field experiment is descriptive.

The question then is – what would be gained, and what would be lost, if field experiments were more like laboratory experiments, with respect to theory. We discuss this question using three exemplar types of field experiments – gift exchange experiments, charitable giving field experiments, and negative income tax studies.

IV. Gift Exchange Field Experiments

Akerlof (1982) argued that a gift exchange mechanism between employers and employees can play an important role in labor markets. If employees respond to a kind wage offer by work harder, employers may find it optimal to offer wages above the reservation utility. Gift exchange, hence, is a possible rationale for efficiency wages. This theory has proven hard to test empirically. For one thing, the repeated nature of employment contracts makes it difficult to separate genuine gift exchange from repeated game equilibria: the worker exerts extra effort in anticipation of future compensation and so on. In a genuine gift exchange, instead, the worker exerts extra effort because the "gift" by the employer induces social preferences towards the employer.

In a highly-cited laboratory study, Fehr, Kirchsteiger and Riedl (1993) test for gift exchange. In the experiment, some subjects are assigned the role of firms, others the role of workers. Firms move first and make a wage offer $w \in \{0, 5, 10, ...\}$. Workers then chose effort $e \in [0.1, 1]$. Workers and firms engage in one-shot interactions, so repeated-game effects are eliminated by design. Since effort is costly, the sub-game perfect equilibrium for self-interested workers is to exert the minimal effort $e^* = 0.1$, no matter what the wage offer. In anticipation of this, self-interested firms should offer workers their reservation utility, in this case $w^* = 30$.

Fehr, Kirchsteiger, and Riedl observe behavior that is starkly different from these predictions. Almost all subjects in the role of firms offer wages higher than 30, and subjects in the role of workers respond by exerting higher effort (Figure 4). This is precisely, in a laboratory setting, the gift exchange that Akerlof (1982) postulated. The reciprocal behavior by the workers makes it rational for firms to offer efficiency wages. A number of laboratory experiments have confirmed and extended the findings of this paper.

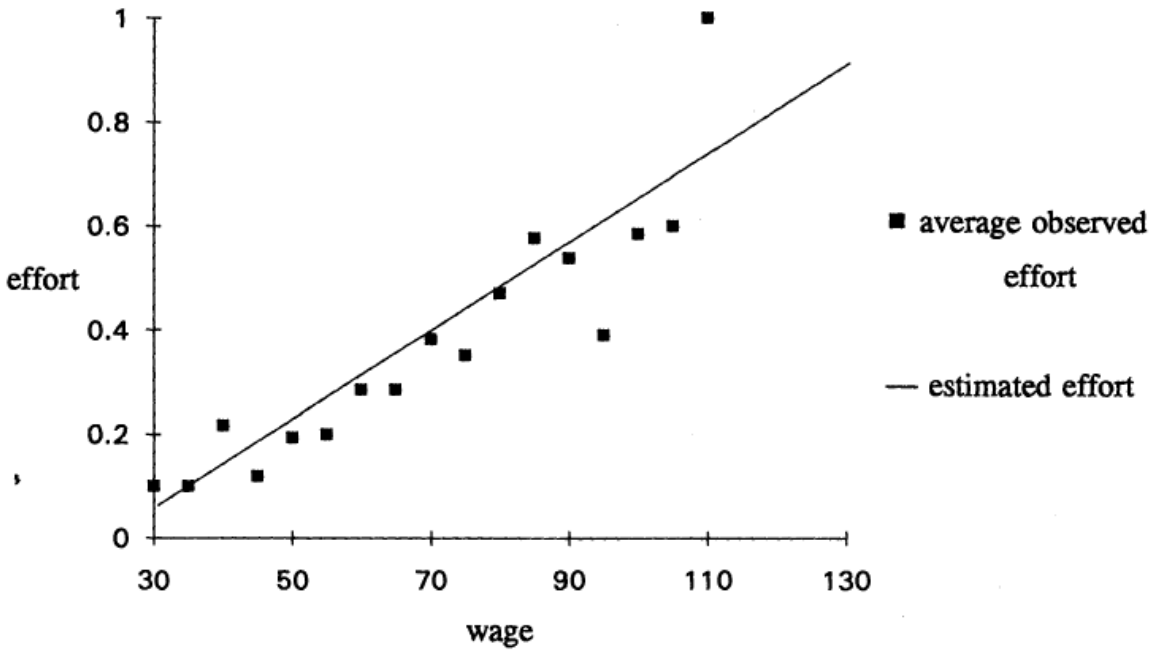

Figure 4. (reproduced from Fehr, Kirchsteiger, and Riedl (1993))

One may argue that, as interesting as this evidence is, behavior in an actual employment contract differs from behavior in the laboratory. Yet, employment relationships with their repeated nature make testing of gift exchange behavior very difficult.

Gneezy and List (2006) designed a field experiment that resolves this difficulty. They hire workers for two tasks, coding library books and running a fund-raising drive. They make it clear that the jobs are one-time tasks, hence removing repeated-interaction incentives. Once subjects show up for their task, a sub-set is randomly assigned a surprise pay of $20 per hour, while the control group is paid $12 per hour, as promised. Gneezy and List then examine whether effort responds to the higher pay, as predicted by the gift exchange hypothesis. (Notice that the higher pay is a flat payment, and as such does not

alter the incentives to exert effort) The main finding in the paper is that work effort is substantially higher in the first three hours of the job in the gift treatment relative to the control treatment, but it is indistinguishable after that. This suggests that gift exchange is present, but short-lived. This innovative design spawned a whole literature of field experiments using similar short-term, but real, employment contracts.

What neither Gneezy and List (2006) nor most of the follow-up papers do is to provide a model for the observed behavior; as such, they are descriptive field experiments. However, while gift exchange is indicative of non-standard preferences (else the worker would not reciprocate in a one-shot interaction), multiple models of social preferences can explain the evidence.

For the findings in Gneezy and List (2006), consider two prominent classes of explanations: inequity aversion and reciprocity. Under inequity aversion, put forward by Fehr and Schmidt (1999) and Bolton and Ockenfels (2000), individuals dislike inequity: while they want higher payoffs for themselves, they are willing to forgo some payoff to help another player who is behind them – though not someone who is ahead of them. This simple model of social preferences has been remarkably successful in accounting for behavior in a variety of contexts, including behavior in the dictator game, the ultimatum game, and, most relevant for us, gift exchange in the laboratory. In the Fehr et al. experiment, the "firm" falls behind by paying a (higher) wage. The worker can mitigate this inequity by exerting effort which benefits the firm at, at least initially, limited cost to the worker (since the cost function is convex). The model also predicts that the worker will not instead put this effort if the firm has not paid a generous wage. In this latter case, the firm is ahead in payoffs, and as such putting in effort would increase, not decrease, inequality.

Under reciprocity models instead (such as the intention-based models in Rabin, 1993, and Dufwenberg and Kirchsteiger, 2004, or type-dependent preferences in Levine, 1998, or action-based models as in Cox, Friedman and Sadiraj, 2008), individuals have positive social preferences towards others who they think are nice or behave nicely, but not (as much) towards individuals who are not nice. In the laboratory gift exchange, workers exert effort if the firm pays a higher wage because of the inference workers

make about how nice the firm is. And they do not exert effort under a low wage because they do not care for firms that prove to be selfish.

Can gift exchange experiments in the field then help separate the two explanations? It is simple to show that they do, even though this point has not been made in any of the papers cited above. The inequity aversion model predicts gift exchange in the laboratory because the generous wage payment by the firm causes the firm to fall behind in payoffs relative to the worker, triggering the inequity-diminishing effort by the worker. But in the field experiment, it is highly implausible that a higher wage payment by the firm for a 6-hour task causes the firm to fall behind in payoffs relative to the workers. But if the "gift" payment does not alter the inequity between the worker and the firm, it will not induce gift exchange behavior. Hence, any observed gift exchange in firms cannot be due inequity aversion but to other social preferences such as reciprocity. This point applies, beyond the labor market, to other economic settings where gifts are given to influence behavior, such as gifts to doctors in the pharmaceutical industry, or vote-buying in the case of politicians. [7] These gifts, when not explained by standard incentives, cannot be explained by inequity aversion, but only by some of the existing reciprocity-based theories (action-based reciprocity, Cox et al., 2008).

Adding a simple model of two (or more) competing social-preference models would thus add insights beyond the Descriptive contribution of the field experiments. As a further step, using a model of reciprocal preferences, one can ask how much reciprocity is implied by the observed gift exchange in the field. In Gneezy and List, the increase in pay raises (temporarily) productivity in book coding by 30%. But was that gain achieved easily, or did it require great effort? In the latter case, the observed gift exchange indicates substantial reciprocation, in the former case not. To estimate the extent of reciprocity would require knowing the shape of the cost function of effort. This can be done by varying in a randomized fashion the piece rate. As such, additional experimental treatments can be designed to estimate the nuisance parameters (in this case the curvature of the cost of effort) and thus shed light on the parameters of interest (the extent of reciprocity).

[7] On these topics, see Malmendier and Schmidt (2010) and Finan and Schechter (2010).

To summarize, the gift exchange experiments suggest that there is an important role played by both types of experimental evidence: The laboratory experiments in Fehr et al. (1993) were the first to suggest an experimental methodology to test for gift exchange, and find support for it in the laboratory. The Gneezy and List (2006) paper was a mile-stone in that it proposed a design for gift exchange in a real employment contract unconfounded by repeated game effects. While this field experiment falls in the Descriptive category, follow-up modeling can clarify its implications for the theoretical work on social preferences. Furthermore, studies that structurally estimate these parameters could build on the Gneezy and List design. Hence, scientific progress can be achieved by a sequence of papers, each adding to the previous work.

V. Charitable Giving Field Experiments

A series of field experiment have transformed the charitable giving field from an area mostly focused on modeling and stylized facts to an area constantly fueled by new experimental findings. A trail-blazing field experiment was List and Lucking-Reiley (2002). In a mailer requesting funds for a research center, the authors randomized both the seed money (the funding already available) and whether funds would be refunded in case the fund-raising targets were not met. This experiment was motivated by Andreoni's signalling model of charitable giving; however, since the List and Lucking-Reiley (2002) paper does not contain a model, we categorize it as Descriptive. Most recent field experiments in the area follow List and Lucking-Reiley: they are motivated by models on charitable giving, but they are ultimately Descriptive (see e.g., Falk, 2007).

In 2006, two of the authors of this paper, Stefano DellaVigna and Ulrike Malmendier, together with John List started discussing a field experiment along these lines. The idea was to attempt to discriminate between two groups of reasons for giving to a charity when asked for a donation. Giving may be associated with a utility increase, whether due to altruism, warm glow, or prestige. Alternatively, individuals may actually *dislike* giving money to a charity, but feel even worse saying no to the solicitor; in this latter case, giving is due to the social pressure that the individuals experience when being asked. We thought that it was important to distinguish the two motivations for giving

since they have very different welfare implications: giving is welfare-increasing in the first case, but welfare-diminishing for the donor in the second case.

Thus, we discussed a field experimental design and settled on a door-to-door campaign where we would randomize the extent to which people are informed about the upcoming campaign. In the treatment group, but not in the control group, we would post a flyer on the door-knob of the household, informing them of the upcoming fund-raiser (see Figure 5). Households could then vote with their feet—if giving is mostly due to altruism, households in the treatment group would sort into staying at home and give; if giving is mostly due to social pressure, they would sort out to avoid being asked.

Figure 5. Example of the flyers used by DellaVigna, List, and Malmendier (2010)

Our initial plan for the field experiment was in the Descriptive line of previous work: we intended to test a hypothesis which was intuitively suggested by theory, but without actually making the underlying model explicit. After some discussions, though, we decided to write the model to clarify what assumptions we were implicitly making. We assumed a cost function of shifting the probability of being at home (in response to

the flyer), and we allowed for competing models to explain sorting and giving behavior: altruism on the one hand and a social pressure cost from turning down an in-person giving request on the other hand.

In our case, the dividends from writing the model were substantial. In addition to clarifying the assumptions needed (for example, that there is no social pressure cost from avoiding the solicitor by not answering the door), the model suggested novel predictions. One such prediction relates to the size of donations. In our model, social pressure drives small, but not large donations. Hence, if social pressure is responsible for the observed donations, the flyer treatment should lower small donations, but not larger ones. The model also suggested new treatments. In particular, we added an "Opt-Out" treatment in which the flyer includes a box that can be checked if the household does not "want to be disturbed." This treatment makes sorting easier, i.e., lowers the cost of avoiding the solicitor relative to the regular flyer without opt-out box. Hence any (additional) decrease in giving allows us to identify social pressure more directly and to address confounding explanations such as information or self- and other-signaling models. In summary, making the model explicit before running the experiment, made for a tighter and more informative test of the initial hypothesis.

In addition, we realized that, were it not for one nuisance parameter, we would be able to estimate the key parameters of the model, including the social pressure cost of saying no to an in-person request, and the extent of altruism. The nuisance parameter is the elasticity of the cost of sorting in and out of the home, a key parameter to make inferences. Suppose for example that the flyer reduces the probability of home presence by 4 percentage points – is that much or is it little? Unfortunately, none of the experimental treatments allowed us to "monetize" the magnitude and estimate this elasticity parameter.

This led us to think of other ways to estimate this parameter. In the end, still in the design stage, we decided to run a parallel field experiment explicitly designed for the purpose. We posted flyers announcing that "Researchers will visit this address tomorrow (/) between … and … to conduct an X-minute survey. You will be paid $Y for your participation". Across treatments we varied the time duration X (10/5 minutes) and the payment Y ($0/$5/$10). The responsiveness in the presence at home with respect to the duration and the payment provided the identification to the elasticity parameters, hence allow-

ing us to back out all other parameters. Indeed, in the end these survey treatments made up the bulk of our field experiment, even though their only purpose was to estimate a nuisance parameter.

The reduced-form results in DellaVigna, List, and Malmendier (2010) point to the importance of social pressure for solicited donations, with the most important piece of evidence being the fact that the flyer with opt-out option lowers donations significantly, and especially the small donations. As discussed above, this is a key prediction of the social pressure framework which we had not honed in until we wrote the model. As such, writing the model provided us with a tighter reduced-form test.

What do the survey treatments and the ensuing parameter estimation add to these results? We estimate the effect of a fund-raising campaign on the welfare of the households contacted. In a model with no social pressure, the welfare effect of a campaign can only be positive since a donor can always costlessly say no. But in the presence of social pressure, this free-disposal condition does not hold: the benefits of a campaign for the willing donors have to be weighed against the cost non-donors pay for being asked and saying no, which we estimate to be about $4 for a local charity. In addition to this cost for non-donors, we estimate that as many as 50 percent of the donors would have preferred not to be asked, because social pressure induces them to give when they would not have given otherwise, or give too much.

Taking into account these forces, our benchmark specification indicates that our door-to-door campaign induces a welfare loss of about $1 on average per household contacted (including households that were not at home and hence did not suffer a welfare loss, and not counting the benefits associated with the public good provision). An interesting and counter-intuitive result is that the local and well-liked favorite charity is associated with more negative welfare impacts than an out-of-state and lesser-known charity because, yes, more people are willing to donate to the local charity, but at the same time the social pressure cost of saying no is significantly higher, and the second force dominates. These latter findings, which of course require some parametric assumptions, complement the descriptive findings.

VI. Negative Income Tax Experiments

The two previous cases discussed suggest that, while there is much that we learn from descriptive studies, there can be additional benefits in developing a fully specified behavioral model and obtaining estimates of the key parameters from that model. This process can follow from models and estimates that are obtained in follow-up papers, as may happen for the gift exchange experiments, or could be part of the design of the initial field experiment, as in the charity experiment described above.

But is it always advantageous to have a model with parameter estimation? In this Section we consider the case of the Negative Income Tax Experiments, one of the most famous large-scale social experiments conducted in the US (1968 to 1972). Funded by the Office of Economic Opportunity, this experiment was designed to test the effects of a negative income tax – a simplified two-parameter income support system proposed by Milton Friedman in the 1950s. The experiment attracted widespread professional interest at the time, and dozens of high-profile economists were involved in various aspects of its design and analysis.

Its ultimate design was closely tied to a specific parametric model: Rather than implement a simple "two-group" experimental design, the experiment included a total of eight different treatment arms, each with a specific value for the "guarantee level," i.e., the level of income support for a family with no earnings, and for the program tax rate. A complex optimal assignment model, developed by John Conlisk and Harold Watts, was designed to maximize the efficiency of the experiment, assuming a (parametric) model of the likely responses to the experimental incentives.

In principle, the design could have provided estimates of the incentive effects of various combinations of the guarantee level and tax rate. However, with the very small sample sizes (1,350 subjects, with 750 members of the treatment group, and 46-138 treatments per arm), even the pooled experimental impacts were quite imprecise. The only possible inferences that could be made from the data were under the assumption of the structural model.

Similarly complex designs were employed in the Rural Income Maintenance Experiment (operated in Iowa and North Carolina between 1969 and 1973), the Gary Income Maintenance Experiment (operated in Gary Indiana between 1971 and 1974), and the Seattle-Denver Income Maintenance Experiment (SIME-DIME), which ran between

1971 and 1982. As in the earlier NIT experiments, the SIME-DIME experiment was substantially under-powered – the SIME-DIME sample would have had to have been eight times larger to yield statistically significant treatment effect estimates for even the largest arm of the design.

From today's perspective, the obvious comfort that analysts at the time had with a model-based assigned mechanism is surprising. Equally remarkable, perhaps, was the nearly universal adoption of model-based analysis methods for the NIT experiments (e.g., the analysis in Johnson and Pencavel, 1982). As pointed out by Ashenfelter and Plant (1990), the final report of the SIME-DIME experiment did not include any "non-parametric" estimates of the impact of treatment.

As a result of the frustrations in dealing with the complex designs of the NIT experiments (and with the confusing message that emerged from such designs) many respected analysts adopted the view that social experiments should be designed as simply as possible. For example, Hausman and Wise (1985) argued: "… *we propose as a guiding principle the experiments should have as a first priority the precise estimation of a single or a small number of treatment effects.*" (page 188).

Subsequent social experiments – particularly those that focus on new programs – have tended to follow this advice. As noted by Greenberg, Shroder and Onstott (1999), 80% of the social experiments initiated after 1983 had only a single treatment-control contrast. This shift away from designs that explicitly attempt to model response variation to multiple treatments and toward a single manipulation has led to a new round of criticism (e.g., Heckman and Smith, 1995) that the social experiments are often "*black boxes*" that "… *contribute next to nothing to the cumulative body of social science knowledge…*" (page 108).

While there is no simple answer as to the optimal role of modeling in field experiment, the NIT example makes clear that reliance on a model is not always a plus, particularly in the evaluation of complex social programs that may affect a range of behaviors through multiple channels. Yet, the previous examples also suggest that much can be gained in many experimental settings (like the gift exchange experiment) from a careful consideration of the predictions of economic models.

VII. Conclusions

Over the last two decades, economics has witnessed a dramatic expansion of experimental research. Both laboratory and field experiments share the common advantage of studying a controlled setting in order to evaluate treatment effects. There is, however, as we documented, a noticeable difference in the evolution of these two types of experimental research: Laboratory experiments feature a much closer link to theory than field experiments. We discussed three cases to highlight the benefits, as well as the potential drawbacks, of the presence of theory in field experiments

References

Akerlof, George A. “Labor Contracts as Partial Gift Exchange.” *The Quarterly Journal of Economics*. Vol. 97, No. 4: 543-569

Joshua D. Angrist. “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records” *American Economic Review* Vol. 80, No. 3 (Jun., 1990), pp. 313-336.

Ashenfelter, Orley and Mark W. Plant. “Nonparametric Estimates of the Labor Supply Effects of Negative Income Tax Programs. *Journal of Labor Economics* 8 (January 1990, part 2), pp. S396-S415.

Bandiera, Oriana , Iwan Barankay,. and Imran Rasul. 2009. “Social Connections and Incentives in the Workplace: Evidence From Personnel Data.” *Econometrica*, 77: 1047–1094.

Bandiera, Oriana , Iwan Barankay,. and Imran Rasul. 2007, “Incentives for Managers and Inequality Among Workers: Evidence from a Firm Level Experiment”, *Quarterly Journal of Economics* 122: 729-74.

Barberis, Nicholas, and Wei Xiong. 2009. “What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation.” *Journal of Finance*, Vol. 64, Issue 2, pages 751–784.

Bolton, Gary E and Axel Ockenfels. Mar., 2000. “ERC: A Theory of Equity, Reciprocity, and Competition.” *The American Economic Review*, Vol. 90, 166–193.

Gary S. Becker. “Irrational Behavior and Economic Theory” *Journal of Political Economy*, Vol. 70, No. 1 (Feb., 1962), pp. 1-13

Rebecca M. Blank. “The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The American Economic Review” *American Economic Review*, Vol. 81, No. 5 (Dec., 1991), pp. 1041-1067

Card, David. and Dean. R. Hyslop. 2005. “Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers.” *Econometrica*, 73: 1723–1770

Chetty, Raj, Adam Looney and Kory Kroft. 2009. “Salience and Taxation: Theory and Evidence.” *The American Economic Review*. Vol.99(4) : 1145-1177

Cox, James C; Friedman, Daniel and Vjollca Sadiraj. 2008. “Revealed Altruism.” *Econometrica,* Vol. 76, 31–69.

Angus Deaton. “Understanding the Mechanisms of Economic Development” *Journal of Economic Perspectives*, vol. 24, iss. 3, pp. 3-16, August 2010.

DellaVigna, Stefano, Ulrike Malmendier, and John. A. List. 2010. “Testing for Altruism and Social Pressure in Charitable Giving” Working paper.

Falk, Armin. 2007. “Gift Exchange in the Field”. *Econometrica*, 75: 1501–1511.

Fehr, Ernst and Lorenz Goette. 2007. “Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment.” *The American Economic Review*. Vol. 97, No. 1: 298-317

Fehr, Ernst , Georg Kirchsteiger and Arno Riedl. 1993. “Does Fairness Prevent Market Clearing? An Experimental Investigation.” *The Quarterly Journal of Economics*. Vol. 108, No. 2 : 437-459

Fehr, Ernst, and Klaus M. Schmidt. 1999. “A Theory of Fairness, Competition, and Cooperation.” *Quarterly Journal of Economics*, 114(3): 817–68.

Finan, Fred and Laura Schechter. 2010. “Vote-buying and Reciprocity,” Working paper.

Forsetlund Louise, Iain Chalmers and Arild Bjørndal. 2007. “When Was Random Allocation First Used To Generate Comparison Groups In Experiments To Assess The Effects Of Social Interventions?” Economics of Innovation and New Technology. Vol. 16(5) : 371-384

Gneezy, Uri. and List, John. A. 2006. “Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments”. *Econometrica*, 74: 1365–1384.

Harrison, Glenn W., and John A. List. 2004. “Field Experiments.” *Journal of Economic Literature*. 42 : 1013–1059.

Greenberg, David, Mark Shroder and Matthew Onstott. “The Social Experiment Market.” *Journal of Economic Perspectives* 13 (Summer 1999), pp. 157-172.

Hausman, Jerry A. and David A. Wise. “Technical Problems in Social Experimentation: Cost versus Ease of Analysis.” In Jerry A Hausman and David A. Wise, editors. *Social Experimentation*. Chicago: University of Chicago Press, 1985.

Heckman, James J. and Jeffrey A. Smith. “Assessing the Case for Social Experiments.” *Journal of Economic Perspectives* 9 (Spring 1995), pp. 85-110.

Johnson, Terry R. and Pencavel, John H. 1982. “Forecasting the Effects of a Negative Income Tax Program”. *Industrial and Labor Relations Review*. Vol. 35, No. 2 : 221-234

Kling, Jeffrey R., Jeffrey B Liebman and Lawrence F Katz. 2007. “Experimental Analysis of Neighborhood Effects.” *Econometrica*, 75: 83–119.

LaLonde, Robert J. 1986. “Evaluating the Econometric Evaluations of Training Programs with Experimental Data.” *The American Economic Review*. Vol. 76, No. 4 : 604-620

Levine, David K. 1998. “Modeling Altruism and Spitefulness in Experiments.” *Review of Economic Dynamics*, Vol. 1, 593–622.

List, John A. “Does Market Experience Eliminate Market Anomalies?" *Quarterly Journal of Economics*, 2003, Vol. 118, No. 1 : 41-71

List, John A. and David Lucking-Reiley. 2002. “The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign”. *The Journal of Political Economy* Vol. 110, No. 1 : 215-233

Malmendier, Ulrike and Klaus Schmidt. 2010. “You Owe Me.” Unpublished Manuscript.

Miguel, Edward and Michael Kremer. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica*, 72: 159–217.

Nagin, Daniel S., James B. Rebitzer, Seth Sanders and Lowell J. Taylor. 2002. “Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment.” *The American Economic Review*. Vol. 92, No. 4 : 850-873

Odean, Terrance. 1998. “Are Investors Reluctant to Realize Their Losses?” *Journal of Finance*, 53(5): 1775–98.

Rabin, 2010. “Improving Theory with Experiments, Improving Experiments with Theory, all so as to Improve Traditional Economics.” Unpublished Manuscript.

Sacerdote, Bruce. 2001. “Peer Effects with Random Assignment: Results for Dartmouth Roommates.” *Quarterly Journal of Economics.* Vol.116:2, 681-704

Todd, Petra E. and ; Kenneth I. Wolpin. 2006. “Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility.” *The American Economic Review*. Vol.96 (5): 1384-1417

Appendix Table 1, Panel A. List of all Field Experiments Published in Top-5 Journals from 1975 to 2004

Year	Month	Journa	Pages	Authors	Title	Classification
1978	12	AER	pp. 873-887	Keeley, Michael C. and Robins, Philip K. and Spieg	The Estimation of Labor Supply Models Using Experimental Data	Single Model
1978	12	JPE	1103-1130	Burtless, Gary and Hausman, Jerry A.	The Effect of Taxation on Labor Supply: Evaluating the Gary Negative Income Tax Experiment	Single Model
1979	3	EMA	pp. 455-473	Hausman, Jerry A. and Wise, David A.	Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment	Single Model
1980	5	EMA	pp. 1031-105	Nancy Brandon Tuma and Robins, Philip K.	A Dynamic Model of Employment Behavior: An Application to the Seattle and Denver Income Mair	Single Model
1980	1	RES	75-96	Hausman, Jerry A. and Wise, David A.	Discontinuous Budget Constraints and Estimation: The Demand for Housing	Parameter Estim.
1982	6	AER	pp. 488-497	Burtless, Gary and Greenberg, David	Inferences Concerning Labor Supply Behavior Based on Limited-Duration Experiments	Single Model
1984	9	AER	pp. 673-684	Plant, Mark W.	An Empirical Analysis of Welfare Dependence	Single Model
1986	9	AER	604-620	LaLonde, Robert J.	Evaluating the Econometric Evaluations of Training Programs with Experimental Data	Descriptive
1987	9	AER	513--530	Woodbury, Stephen A. and Spiegelman, Robert G.	Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois	Descriptive
1987	6	AER	251--277	Manning, Willard G. and Newhouse, Joseph P. and	Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment	Descriptive
1991	12	AER	1041--1067	Blank, Rebecca M.	The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from The Ame	Descriptive
1995	6	AER	304--321	Ayres, Ian and Siegelman, Peter	Race and Gender Discrimination in Bargaining for a New Car	Descriptive
1996	1	EMA	pp. 175-205	Ham, John C. and Lalonde, Robert J.	The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experime	Descriptive
1997	10	RES	605--654	Heckman, James J. and Ichimura, Hidehiko and To	Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Progr	Descriptive
1997	10	RES	537--553	Manski, Charles F.	The Mixing Problem in Programme Evaluation	Descriptive
1997	10	RES	487--535	Heckman, James J. and Smith, Jeffrey and Clemen	Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heteroge	Descriptive
1997	10	RES	655--682	Eberwein, Curtis and Ham, John C. and Lalonde, R	The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of D	Descriptive
1998	6	JPE	457--482	Camerer, Colin F.	Can Asset Markets Be Manipulated? A Field Experiment with Racetrack Betting	Descriptive
1999	12	AER	1063--1080	Lucking-Reiley, David	Using Field Experiments to Test Equivalence between Auction Formats: Magic on the Internet	Descriptive
1999	5	QJE	497--532	Krueger, Alan B.	Experimental Estimates of Education Production Functions	Descriptive
2000	9	AER	961--972	List, John A. and Lucking-Reiley, David	Demand Reduction in Multiunit Auctions: Evidence from a Sportscard Field Experiment	Descriptive
2000	5	QJE	651--694	Heckman, James and Hohmann, Neil and Smith, Je	Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment	Descriptive
2001	12	AER	1498--1507	List, John A.	Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Fie	Descriptive
2001	7	EMA	pp. 1099-111	Philipson, Tomas	Data Markets, Missing Data, and Incentive Pay	Descriptive
2001	5	QJE	607--654	Katz, Lawrence F. and Kling, Jeffrey R. and Liebma	Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment	Descriptive
2001	5	QJE	655--679	Ludwig, Jens and Duncan, Greg J. and Hirschfield,	Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment	Descriptive
2002	12	AER	1636--1643	List, John A.	Phenomenon	Descriptive
2002	12	AER	1535--1558	Angrist, Joshua and Bettinger, Eric and Bloom, Erik	Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment	Descriptive
2002	9	AER	850--873	Nagin, Daniel S. and Rebitzer, James B. and Sande	Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field E	Single Model
2002	1	EMA	pp. 91-117	Abadie, Alberto and Angrist, Joshua and Imbens, G	Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee I	Descriptive
2002	2	JPE	215--233	List, John A. and Lucking-Reiley, David	The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a Univ	Descriptive
2003	6	JPE	530--554	Grogger, Jeffrey and Michalopoulos, Charles	Welfare Dynamics under Time Limits	Descriptive
2003	8	QJE	815--842	Duflo, Esther and Saez, Emmanuel	The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Ra	Descriptive
2003	2	QJE	41--71	List, John A.	Does Market Experience Eliminate Market Anomalies?	Descriptive
2004	12	AER	1717--1722	Frey, Bruno S. and Meier, Stephan	in a Field Experiment	Descriptive
2004	9	AER	991--1013	Bertrand, Marianne and Mullainathan, Sendhil	Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Mark	Descriptive
2004	3	EMA	pp. 615-625	List, John A.	Neoclassical Theory versus Prospect Theory: Evidence from the Marketplace	Descriptive
2004	9	EMA	pp. 1409-144	Chattopadhyay, Raghabendra and Duflo, Esther	Women as Policy Makers: Evidence from a Randomized Policy Experiment in India	Competing Models
2004	1	EMA	159-217	Miguel, Edward and Kremer, Michael	Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities	Descriptive
2004	2	QJE	49--89	List, John A.	The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field	Descriptive
2004	4	RES	513--534	Shearer, Bruce	Piece Rates, Fixed Wages and Incentives: Evidence from a Field Experiment	Parameter Estim.

Appendix Table 1, Panel B. List of all Field Experiments Published in Top-5 Journals from 2005 to 2010

Year	Month	Journa	Pages	Authors	Title	Classification
2005	11	EMA	pp. 1723-177	Card, David and Hyslop, R.	Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers	Descriptive
2005	2	QJE	87–130	Jeffrey R. Kling, Jens Ludwig, Lawrence F. Katz	Neighborhood Effects on Crime for Female and Male Youth: Evidence From a Randomized Housi	Descriptive
2006	9	AER	988--1012	Bitler, Marianne P. and Gelbach, Jonah B. and Hoy	What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments	Descriptive
2006	12	AER	1384--1417	Todd, Petra E. and Wolpin, Kenneth I.	Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Valid	Parameter Estim.
2006	9	EMA	pp. 1365-138	Gneezy, Uri and List, John A.	Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Ex	Descriptive
2006	2	JPE	1--37	List, John A.	The Behavioralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actu	Descriptive
2006	11	QJE	1311–1346	Esther Duflo, William Gale, Jeffrey Liebman, Peter	Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H	Descriptive
2006	5	QJE	635–672	Nava Ashraf, Dean Karlan, Wesley Yin	Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines	Descriptive
2006	5	QJE	747–782	Craig E. Landry, Andreas Lange, John A. List, Mich	Toward an Understanding of the Economics of Charity: Evidence from a Field Experiment	Single Model
2007	12	AER	1774--1793	Karlan, Dean and List, John A.	Does Price Matter in Charitable Giving? Evidence from a Large-Scale Natural Field Experiment	Descriptive
2007	3	AER	298--317	Fehr, Ernst and Goette, Lorenz	Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment	Competing Models
2007	1	EMA	pp. 83-119	Kling, Jeffrey R. and Liebman, Jeffrey B. and Katz,	Experimental Analysis of Neighborhood Effects	Descriptive
2007	9	EMA	pp. 1501-151	Falk, Armin	Gift Exchange in the Field	Descriptive
2007	4	JPE	200--249	Olken, Benjamin A.	Monitoring Corruption: Evidence from a Field Experiment in Indonesia	Descriptive
2007	8	QJE	1235–1264	Abhijit V. Banerjee, Shawn Cole, Esther Duflo, Leig	Remedying Education: Evidence from Two Randomized Experiments	Descriptive
2007	8	QJE	1007–1065	Michael Kremer, Edward Miguel	The Illusion of Sustainability	Single Model
2007	11	QJE	1639–1676	Marianne Bertrand, Simeon Djankov, Rema Hanna,	Obtaining a Driver's License in India: An Experimental Approach to Studying Corruption	Descriptive
2008	12	AER	1829--1863	Thornton, Rebecca L.	The Demandfor, and Impact of, Learning HIV Status	Descriptive
2008	6	AER	1040-1068	Dean S. Karlan and Jonathan Zinman	Credit Elasticities in Less-Developed Economies: Implications for Microfinance	Descriptive
2008	9	AER	1553--1577	Robert T. Jensen and Nolan H. Miller	Giffen Behavior and Subsistence Consumption	Single Model
2008	12	AER	1887--1921	Schochet, Peter Z, Burghardt, John and McConnell,	Does Job Corps Work? Impact Findings from the National Job Corps Study	Descriptive
2008	5	EMA	643-660	Graham, Bryan S.	Identifying Social Interactions through Conditional Variance Restrictions	Descriptive
2008	11	QJE	1329–1372	Suresh de Mel, David McKenzie, Christopher Wood	Returns to Capital in Microenterprises: Evidence from a Field Experiment	Single Model
2008	11	QJE	1373-1414	Justine S. Hastings, Jeffrey M. Weinstein	Information, School Choice, and Academic Achievement: Evidence from Two Experiments	Descriptive
2008	1	RES	117--132	Fisman, Raymond and Iyengar, Sheena S. and Kan	Racial Preferences in Dating	Descriptive
2009	9	AER	1384-1414	Angrist, Joshua and Lavy, Victor	The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial	Descriptive
2009	6	AER	864--882	Cai, Hongbin, Chen, Yuyu and Fang, Hanming	Observational Learning: Evidence from a Randomized Natural Field Experiment	Descriptive
2009	3	AER	486--508	Angelucci, Manuela and Giorgi, Giacomo De	Indirect Effects of an Aid Program: How Do Cash Transfers Affect Ineligibles' Consumption?	Single Model
2009	11	EMA	pp. 1993-200	Karlan, Dean and Zinman, Jonathon	Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Exp	Competing Models
2009	5	EMA	pp. 909-931	Charness, Gary and Gneezy, Uri	Incentives to Exercise	Descriptive
2009	8	JPE	668--700	Brown, Jennifer and Morgan, John	How Much Is a Dollar Worth? Tipping versus Equilibrium Coexistence on Competing Online Auctic	Competing Models
2009	6	JPE	453--503	Bobonis, Gustavo J.	Is the Allocation of Resources within the Household Efficient? New Evidence from a Randomized I	Competing Models
2009	5	QJE	735-769	Martina Björkman, Jakob Svensson	Power to the People: Evidence from a Randomized Field Experiment on Community-Based Monit	Descriptive
2009	11	QJE	1815-1851	Leider, Stephen and Mobius, Markus M. and Rosen	Directed Altruism and Enforced Reciprocity in Social Networks	Descriptive
2009	4	RES	451--469	Ariely, Dan and Gneezy, Uri and Loewenstein, Geo	Large Stakes and Big Mistakes	Descriptive
2009	7	RES	1071--1102	Lee, David S.	Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects	Descriptive
2010	12	AER	2383--2413	Ashraf, Nava and Berry, James and Shapiro, Jesse	Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia	Competing Models
2010	9	AER	1358-98	Chen, Yan and Harper, F. Maxwell and Konstan, Jo	Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens	Single Model
2010	6	AER	958-83	Landry, Craig E. and Lange, Andreas and List, Johr	Is a Donor in Hand Better Than Two in the Bush? Evidence from a Natural Field Experiment	Single Model
2010	4	JPE	274--299	Levav, Jonathan and Heitmann, Mark and Herrman	Order in Product Customization Decisions: Evidence from Field Experiments	Descriptive
2010	5	QJE	515-548	Jensen, Robert	The (Perceived) Returns to Education and the Demand for Schooling	Descriptive
2010	2	QJE	263-305	Bertrand, Marianne and Karlan, Dean and Mullainat	What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment	Single Model
2010	5	QJE	729-765	Anderson, Eric T. and Simester, Duncan I.	Price Stickiness and Customer Antagonism	Descriptive
2010	5	QJE	859-876	Brown, Jennifer and Hossain, Tanjim and Morgan, J	Shrouded Attributes and Information Suppression: Evidence from the Field	Descriptive
2010	2	QJE	Jan-45	Cohen, Jessica and Dupas Pascaline	Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment	Single Model

Notes: List of all papers published in top-5 journals from 1975 to 2010 classified as field experiments. For the categorization into 4 types by the role of theory, see text