Aversion to Learning in Development? A Global Field Experiment on Microfinance Institutions[1]

Matthew R. Brigham[1], Michael G. Findley[2], William T. Matthias[1], Chase M. Petrey[1], and Daniel L. Nielson[1]

[1]Department of Political Science
Brigham Young University

[2]Department of Government
University of Texas at Austin

18 March 2013

---

[1] The research design for this experiment was registered on July 27, 2012 with the Experiments in Governance and Politics Network (EGAP). See www.e-gap.org/design-registration/. We originally registered four conditions – control, positive, negative, and mixed. Before proceeding with the experiment we decided the mixed and control conditions captured the same basic idea. As such, we only employed the control, positive, and negative conditions in the experiment. We report on all three of them here. We originally stated that we would use cross-tabulations, including difference of proportions tests, as well as logit regressions to identify the effects. In this paper, we report the results of all of these tests. Although we did not register additional analysis, in this paper we also report additional robustness checks including a multinomial probit mode and a selection model. Institutional Review Board Clearance was received on 28 September 2011.

Abstract

Randomized evaluation has revolutionized the practice of international development by providing credible evidence on the causal effects of anti-poverty interventions. The new approach, however, relies on development organizations' willingness to subject their programs to rigorous evaluation and update based on what they learn. But development organizations, devoted to their methods, may be averse to such updating. With a global email field experiment using 1,419 micro-finance institutions (MFI) as subjects, we test the effects of scientific findings about microfinance on organizations' willingness to learn more about MFI effectiveness and pursue an offered partnership to randomly evaluate their programs. In the positive treatment subjects were randomly assigned to receive a summary of a study by prominent authors finding that microcredit is effective. The negative treatment provided information on research – by the same authors using a very similar design – reporting the *in*effectiveness of microcredit. We compare both conditions to a control in which no studies were cited. The positive treatment elicited twice as many responses as the negative treatment – and significantly more acceptances of our invitation to consider partnering on an evaluation of their program – thus suggesting significant confirmation bias among microfinance institutions. The randomization revolution thus faces real challenges in overcoming development organizations' apparent aversion to learning.

Introduction

Randomized evaluation has swept through the international development community, energizing anti-poverty scholarship and practice with the promise of learning the precise causal effects of interventions in foreign aid and private humanitarian efforts. Affiliates of MIT's Jameel Poverty Action Lab, led by economists Abhijit Banerjee and Esther Duflo, have completed and reported the results of 360 randomized control trials (RCTs) in development as of January 2013 (J-PAL 2013). They add more to the list every month. Official development organizations have joined the movement. Indeed, Banerjee and Duflo reported in 2009 that the World Bank had 67 RCTs under way out of a total of 89 program evaluations in the Africa region alone (152).

As field experimenters in anti-poverty and conflict resolution, we celebrate the success of randomized evaluation in motivating large improvements in learning what works in development (see Cohen and Easterly 2009, Banerjee and Duflo 2009). Because anti-poverty programs are interventions by their very nature, evaluators can test their effects rigorously with similar methods to those that have transformed medicine from quackery into a science that saves billions of lives. By assigning interventions to treatment and control groups, researchers can learn the causal effects of the projects and, by replication, accumulate knowledge of effective development practice in which we can place high confidence.

However, a next logical step in the research program requires that we rigorously test the willingness of practitioners to learn from the new knowledge acquired through RCTs. As in many human endeavors, development community members have great confidence in their current practices. Their methods make intuitive sense to them, and if their practices are generally followed by many others, the programs may seem "correct," "right," or even "moral" in a normative sense. Practitioners may resist or ignore evidence that contradicts their common sense and feelings of moral obligation. The irony here, of course, is that the goal of the development community is not the perpetuation of current practices but the relief of poverty. Thus, more than in many endeavors, people engaged in anti-poverty efforts ought to be open to information about ways to achieve their goal more effectively. But are they?

We currently have very little information about how open or averse development organizations might be to the new knowledge being generated by randomized evaluations. If the aversion to learning is significant, then the new wave of development scholarship faces the additional challenge of persuading a resistant target audience of the value of the new knowledge. The present study pursues this question with a field experiment in which development organizations serve as subjects.

We selected micro-finance institutions (MFIs) as subjects both because of the prominence of microfinance's boosters as well as the quality of the randomized control trials evaluating the effectiveness of micro-finance. Cheerleaders for microfinance, such as Nobel Peace Prize Winner and Grameen Bank founder Muhammad Yunus, have touted small loans to the very poor as the answer to many development problems, including missing labor markets, lack of women's empowerment, limited education opportunities, and poor public health (Yunus 2007).

High-quality randomized evaluations, however, suggest that microcredit can be very helpful in providing capital to entrepreneurs, but it can also induce high indebtedness and may have no effects on women's empowerment, education, or health (Banerjee et al. 2009). Thus, the disconnect between practitioners' beliefs and scholars' current findings creates an opportunity to probe the willingness of development practitioners to update.

We therefore sent sincere offers by email to 1,419 microfinance institutions listed on the *Mixmarket.org* directory. As affiliates with a development research lab, we are actively seeking partners in many areas of international development with which we might undertake randomized evaluations of their programs. The emails did not offer immediate partnership but instead emphasized current partnership commitments and the need for future funding premised on availability and mutual interest. The emails concluded with an invitation for the MFIs to receive additional information both about studies of microfinance and regarding a potential future partnership with our lab to perform a randomized evaluation. The offer was part of an active effort to recruit potential partners and thus involved no deception.

In the experiment we included two treatment conditions and a control. The control condition email introduced our academic organization and offered additional information about randomized evaluation and a potential partnership. The positive condition augmented the control email with a paragraph summarizing the findings from prominent development economists finding positive effects from microfinance. The negative condition also began with the control email language but added a paragraph summarizing findings from a different study by the same prominent authors finding that a microfinance program produced null effects. The positive condition elicited twice as many acceptances of our invitation for additional information about the potential evaluation relative to the negative condition, suggesting significant confirmation bias on the part of microfinance institutions and marking a major challenge for randomized evaluators in persuading development organizations to update their practices. In what follows we develop the context for this study, describe the experiment, report the results, and discuss the implications.

Background and Literature

This article probes the openness of development practitioners to updating their practices. The history of development evaluation suggests caution here. For decades aid agencies have claimed success rates for all projects ranging from two-thirds to four-fifths (Picciotto 2012, Faiola 2009). Traditional program evaluation involves monitoring the outputs of projects and comparing them to initial goals, which presents a particularly low bar. If program plans state objectives explicitly, say, constructing so many miles of paved road, it should be relatively unproblematic to provide the planned outputs. The new road can be observed and measured accordingly. What is more, the aid agency personnel who produce the monitoring data are often the same people who designed the project in the first place, and both career incentives and confirmation bias likely influence how they report results. Hence, very high success rates for projects naturally follow.

Aid donors reported this high rate of success despite the fact that voluminous scholarship has produced very limited evidence that foreign aid positively affects economic growth in the main (see, for example, Easterly, Levine and Roodman 2004; Rajan and Subramanian 2008). Indeed, the disconnect between high reported success rates for individual projects and limited evidence of economy-wide impact of aid has been lamented as the "micro-macro paradox" (Mosley 1986, Hansen and Tarp 2000). The inattention to the broader impacts of aid programs beyond the narrowly defined and measured project objectives perhaps helps account for the paradox.

The micro-macro paradox, along with a rise in policy research interest on foreign aid effectiveness, led to a serious discussion in development circles about the methodologies used in evaluating the actual impact of projects (Picciotto 2012). Many argued that while large-N observational studies and qualitative methods were useful in many settings, they could prove potentially imprecise and biased in estimating the causal effects of a program because they could not directly observe the counterfactuals (Ravallion 2001).

In the early 2000s, MIT's newly established Jameel Poverty Action Lab (J-PAL) led the charge in arguing that experimental methods provided the most effective way to approach impact evaluation. Esther Duflo, co-founder of J-PAL, stated at a World Bank Conference on evaluation and development effectiveness in 2003 that "Just as randomized trials for pharmaceuticals revolutionized medicine in the 20th century, randomized evaluations have the potential to revolutionize social policy during the 21st" (Duflo and Kremer 2004). Proponents tout the main virtue of randomized evaluations: due to the close collaboration between researchers and practitioners, RCTs allow the estimation of causal effects – the actual impact of projects – that would not otherwise be possible to evaluate (Duflo and Kremer 2004, Banerjee and Duflo 2009).

These claims have proven compelling to many, so randomized field experiments have become a popular tool in development economics research and have found increasing purchase in development practice. As noted above, in the Africa region alone the World Bank in 2009 was performing RCTs on 67 of 89 (or 75 percent of) program evaluations. The Development Impact Evaluation Initiative at the World Bank, which routinely employs randomization, covers 13 percent of the joint IBRD-IDA portfolio of the Bank (Legovini 2010). And this proportion appears to be growing.

In January of 2011 Rajiv Shaw, Director of the U.S. Agency for International Development (USAID) announced a major overhaul of the agency's monitoring and evaluation practices. The new policy mandates that all programs be evaluated by third parties reporting directly to USAID (not to project contractors) and requires that all "innovative" programs employing "untested" hypotheses undergo randomized impact evaluation (USAID 2010). These evaluation initiatives by the world's two largest aid organizations suggest that RCTs have broken out of the academic cloister and have captured the attention – and the resources – of important development practitioners.

But randomized evaluation has been met with skepticism in the academy. Prominent development economists have questioned both the external validity and theoretical grounding of randomized evaluations (Rodrik 2008, Deaton 2010). Others have openly worried about the perceived high cost of RCTs (Copestake et al. 2009). And yet others point out that RCTs cannot answer many critical questions, including some of the biggest. Writes Avril Subramanian, "What would be the effects of disbursing $1-1.5 billion of foreign aid to Pakistan? RCTs do not, and cannot, have anything to say on the matter – not only because of their narrow focus and applicability, and hence non-generalizability, but also because they cannot speak to macroeconomic effects. The larger developmental effects of aid may be good or bad but RCTs cannot help us distinguish them" (Subramanian 2011).

Advocates of randomization have generally acknowledged these issues. They have answered that problems of external validity can be addressed through systematic replication of experiments in diverse settings. They have granted that experiments should test discrete causal mechanisms derived from sound theory. They recommend that evaluation costs be built into development projects up front. And they admit that RCTs cannot answer many important questions in development (Karlan 2009). This back and forth between "randomistas" and their critics has proven generally helpful in focusing and refining the practice of randomized evaluation.

The present article, however, addresses an additional – and potentially bigger – problem faced by proponents of randomized evaluation: practitioners' potential unwillingness to accept the results of the studies and update their operations. The topic area of this study, microfinance, perhaps best illustrates the challenges involved in motivating development practitioners to open their minds to scientific findings and change their procedures accordingly. Some of the best designed and most persuasive RCTs in development economics have put microfinance to the test, and the results suggest that microfinance significantly improves entrepreneurs' access to credit and therefore provides an important tool in overcoming poverty (Banerjee et al. 2009, Karlan and Zinman 2010). Even where microfinance fails in its primary goals of income-generation or empowerment of women, it may have ancillary benefits in strengthening community ties, helping borrowers cope with risk, and improving informal credit access (Karlan and Zinman 2011). Scholars performing the studies clearly see microfinance as providing part of the answer to the development puzzle.

But part of the answer is insufficient for the advocates of microcredit. Rather, microcredit has been advanced as a panacea for a panoply of problems in developing countries. Most microfinance institutions organize (predominantly female) borrowers into solidarity groups, which meet together often to repay loans and apply for new financing. Access to small amounts of capital purportedly allows these groups of poor women to invest in their small businesses and generate new sources of income enabling them to lift themselves out of the poverty trap while addressing many other problems of poverty, including poor healthcare, lack of access to education, and discrimination against women. In the thirty-five years since Bangladeshi economist Muhammad Yunus started the Grameen Bank, thousands of MFIs around the world have been created to join in the effort to alleviate poverty through small loans to the very poor.

In 2006 the Norwegian Nobel Committee awarded the Nobel Peace Prize to Yunus and his Grameen Bank "for their efforts to create economic and social development from below"

(Mjøs 2006). In his presentation speech at the Nobel award ceremony, Nobel Committee Chairman Ole Danbolt Mjøs extolled the broad scope of microfinance, which clearly factored into the award decision: "The [female borrower] group meets regularly to sharpen each other's perceptions of borrowing, work, repayment and saving. The members undertake to work for food production, pure drinking water, hygiene, health, family planning, economy, discipline, community and motivation in the group and in their families. The groups form networks with other groups. At the grass-roots level the groups thus help to build up communities." In particular, Mjøs praised Yunus' and Grameen's focus on women: "Micro-credit has proved itself to be a liberating force in societies where women in particular have to struggle against repressive social and economic conditions. Economic growth and political democracy cannot achieve their full potential unless the female half of humanity on earth contributes on an equal footing with the male" (Mjøs 2006).

Yunus himself has done much to reinforce this impression of the broad impact of microfinance. For example, in Yunus' book, *Banker to the Poor*, he notes that "Grameen is a private-sector self-help bank, and as its members gain personal wealth they acquire water-pumps, latrines, housing, education, access to health care, and so on" (2007, 203). Later, he writes, "Grameen is committed to social objectives: eliminating poverty; providing education, health care, and employment opportunities to the poor; achieving gender equality through the empowerment of women; ensuring the well-being of the elderly" (2007, 209-210). Thus, the claims for the impact of microfinance are quite broad.

As noted above, development economists employing randomized evaluation put these claims of broad scope to the test in a series of studies. The findings were mixed. One study, which we used in our experimental intervention, found strong treatment effects across a wide range of positive outcomes for a microfinance program in South Africa. Access to microcredit caused improvements in economic self-sufficiency, consumption possibilities, and an index measuring subjects' self-reported perceptions of control and positive outlook – including women's sense of empowerment in their households (Karlan and Zinman 2010). But yet another study employing a similar design by the same authors, which we also used in the experiment, failed to replicate these findings in the Philippines, though as noted it did recover treatment effects for improving community trust, coping with risk, and access to informal credit (2011). Also as noted, a major study conducted by J-PAL scholars in India found that microcredit improved entrepreneurs' investment in durable goods and that the number of new businesses in treatment neighborhoods increased by one third. This is strong evidence that microfinance has positive effects. However, the study also showed that microfinance only increased consumption of non-durables (and therefore consumer debt) for people not inclined to business ownership, and it had no effect on health, education, or female empowerment (Banerjee et al. 2009).

After these results first became public, representatives of the six largest MFIs worldwide assumed that all the results would be negative (they were not) and reacted by producing six anecdotes of successful borrowers (Banerjee and Duflo 2011). Brigit Helms, CEO of Unitus, an international MFI, declared in a *Seattle Times* op-ed, "These studies are giving the inaccurate impression that increasing access to basic financial services has no real benefit…. Our worry is that if these studies can't empirically demonstrate significant economic impact in a short time period, the public will be left with the impression that microfinance has no value – especially dangerous at the exact moment microfinance is poised to do more than ever to alleviate global poverty." (Helms 2010). However, the randomized experimental studies do not show negative results, they show mixed results (Banerjee et al. 2009; Dupas and Robinson 2009; Karlan and

Zinman 2010; Pitt, Khandker, and Cartwright 2006; Holvoet 2005; Garikipati 2008; Rahman 1999). Taken as a whole, the studies merely suggest that microfinance may be overhyped.

In reaction, the Grameen Bank published an article in which it surveyed the evaluation literature relating to microfinance over the past 20 years. Its conclusions focus on the positive effect that microfinance has and downplays negative findings as limitations to current research methods. The report criticizes RCTs as being under-contextualized short-term evaluations that do not capture the robustness of microfinance's impact (Odell 2010). Rather than accept the evidence showing some positive results while providing a corrective to some of the grander claims of the microfinance movement, microfinance advocates counter-attacked, leaving the authors of the microfinance studies somewhat baffled. In their general-market book, *Poor Economics*, Banerjee and Duflo describe their experience in the aftermath of their landmark study.

> "As economists, we were quite pleased with these results: The main objective of microfinance seemed to have been achieved. It was not miraculous, but it was working. There needed to be more studies to make sure that this was not some fluke, and it would be important to see how things panned out in the long run, but so far, so good. In our minds, microcredit has earned its rightful place as one of the key instruments in the fight against poverty.
>
> "Interestingly, this is not how the main results played out in the media and the blogosphere. The results were mainly quoted for the negative findings and as proof that microfinance was not what it was made out to be. And though some MFIs accepted the results for what they were (chief among them, Padmaja Reddy [head of the MFI that partnered in the study], who said this was exactly what she had expected, and financed a second wave of the work to study the longer-term impacts), the big international players in microfinance decided to go on the offensive" (Banerjee and Duflo 2011).

Hypothesis

The present study follows up on this anecdote to learn if the reaction Banerjee and Duflo described is systematic and widespread in the microfinance community. Our intuition that MFIs may be slow to update and therefore averse to the possibility of negative results draws on important research from social psychology. Specifically, confirmation bias may encourage MFIs to accept what they already believe and resist what they do not believe in an effort to avoid or resolve cognitive dissonance (Festinger 1957).

Cognitive dissonance is the psychological discomfort an individual feels when presented with information that runs contrary to previously held beliefs (Stone and Cooper 2001; Steele and Liu 1983; Aronson 1969; Festinger 1957). Steele and Liu explain that previously held beliefs are an individual's "ideal self-image." For this reason, discomfort occurs when presented with information contrary to prior convictions. In the case of MFIs, the "ideal self-image" is that microcredit is a powerful poverty alleviation tool with general effects for a wide range of outcomes.

Confirmation bias clouds the judgment of human beings. Humans are wired to believe chiefly what they want to believe, and what they want to believe rests heavily on priors. For example, in one classic social psychology experiment, both pro-Arab and pro-Israeli citizens interpreted the same media broadcast as being biased against their side (Vallone et al. 1985). In a related experiment, a team of neuroscientists observed the neural responses of Republican and Democratic voters. Each group watched positive and negative campaign ads for candidates George W. Bush and John Kerry before the 2004 presidential election. The neuroscientists found that when subjects confronted a negative campaign ad for their preferred candidate, the region of the brain responsible for reasoning deactivated. When subjects viewed a positive ad, their emotional brains lit up (Westen et al. 2006).

When people are presented with information consistent with prior beliefs, no cognitive dissonance occurs and confirmation bias causes them to readily accept the new information. However, when presented with information inconsistent with prior beliefs, cognitive dissonance occurs and confirmation bias causes them to ignore or minimize the new information. Aronson (1969) describes cognitive dissonance by using Festinger's (1957) real-world example of smoking:

> If a person believes that cigarette smoking causes cancer and simultaneously knows the he himself smoke cigarettes, he experiences dissonance. Assuming that the person would rather not have cancer, his cognition "I smoke cigarettes" is psychologically inconsistent with his cognition "Cigarette smoking produces cancer." Perhaps the most efficient way to reduce dissonance in such a situation is to stop smoking. But, as many of us have discovered, this is by no means easy. Thus, a person will usually work on the other cognition. There are several ways in which a person can make cigarette smoking seem less absurd. He might belittle the evidence linking cigarette smoking to cancer ... or he might associate with other cigarette smokers ... or he might smoke filter-tipped cigarettes and delude himself that the filter traps the cancer-producing materials; or he might convince himself that smoking is an important ... activity…. All of these behaviors reduce dissonance. (Aronson 1969).

Although confirmation bias is a well-documented shortcoming in human decision-making, its presence in non-profit organizations, such as MFIs, is not yet known. One might hope that anti-poverty organizations have developed organizational routines to maximize learning and minimize bias. After all, charitable organizations focus on poverty relief as their primary goal, and any information that might help them achieve that objective ought to be privileged.

We fear, however, that organizational routines are created by the same individuals prone to cognitive biases in the first place, so confirmation bias may be built into or even reinforced by organizational structures. Thus, we hypothesize that MFIs confronted with negative evidence of microfinance's effectiveness should be less willing to request scientific material on microfinance effectiveness or pursue additional information about a possible partnership in a randomized evaluation.

Research Design

We executed the experiment in November and December of 2011 on 1,419 microfinance institutions worldwide. Because the experiment was conducted by email, we carried out the entire experiment from our the Political Economy and Development Lab (PEDL) as detailed below.

Subject Pool and Randomization

Thousands of microfinance institutions spend billions of dollars annually and, while they are generating enormous amounts of attention, unfortunately no standard sampling frame exists for such institutions. We identified a data source, *Mixmarket.org*, which captures a very large number of MFIs, however. *Mixmarket.org* collects data on MFIs for several purposes, including for research and analysis. In their global database, they had just over 1,400 MFIs listed at the outset of the study. The listings included email addresses and demographic information. But we note that because we are using email to apply our treatments, we are limited to MFIs who have access to the Internet. This should bias the sample towards larger, more established groups that likely carry out the majority of microfinance work. While we cannot be certain, the sheer number of MFIs in *Mixmarket.org*, along with the inclusion criteria of Internet access, suggests that it captures the largest and most influential MFIs.

This leads to another potential source of bias in the underlying sample. *Mixmarket.org* likely captures MFIs that are already transparent. Because these MFIs have already provided financial information and agreed to be listed on a website designed to be analyzed, they may have an underlying propensity for transparency and accountability. This potential limitation is actually useful for our analysis. If we find that already transparent MFIs suffer from confirmation bias, then we would expect other MFIs to be that much worse.

The sample we obtain from *Mixmarket.org* represents a variety of regions worldwide as Table 1 below demonstrates.

Table 1: Regional Distribution of MFIs

Region	Freq.	Percent
Africa	290	20
East Asia and Pacific	177	13
Eastern Europe and Central Asia	294	21
Latin America and The Caribbean	372	26
Middle East and North Africa	73	5
South Asia	213	15
Total	1,419	100

Despite some potential bias in the sample, randomization should allow us to uncover unbiased causal effects. We employed block randomization of subjects using region of the world and size of the MFI loan portfolio to demarcate the blocking strata.

Experimental Conditions

Within each stratum we randomly assigned each of the 1,419 subjects to either the control or to one of two treatment conditions. The premise of the experiment was to provide information about the effectiveness or ineffectiveness of microfinance and then address whether MFIs are more or less likely to internalize and react to such information. Rather than simply ask them whether they agree with the information, we hoped to elicit a revealed preference. To accomplish this, we embedded the information about the success or failure of microfinance into an invitation to consider partnering on a randomized evaluation for one of their projects. We expected that if they received positive information, they would be more likely to accept the invitation; if they received negative information, then we expected them to be less likely to accept the invitation.

All email invitations to partner on the evaluation were fully identical in wording, except for a single paragraph in which we introduce the two information variants for treatments 1 and 2. See the appendix for the complete language included in the experimental conditions. Since our research lab is actively recruiting partners with which to undertake randomized evaluations, the invitation was sincere and therefore involved no deception. We have followed up with all organizations that answered our invitation and provided additional information as promised. The emails began with a short introduction to the Political Economy and Development Lab as well as an invitation worded as follows, including a short statement confirming the country in which they operate:

> "We are seeking to assess the interest of qualified microfinance institutions in possible partnerships to perform impact evaluations. We understand that you provide microcredit loans in <country>."

The positive treatment included a priming statement claiming that scientific studies indicate that microfinance is effective:

> "Academic research suggests that microfinance is effective. The results of a recent scientific study show that microcredit loans have a positive effect on economic self-sufficiency and subjective well-being of borrowers, including the decision making power women have in the home (Karlan and Zinman 2009, "Expanding Credit Access," *Review of Financial Studies*). These results are compelling to us, and we wish to learn more so we can further assist those in need."

This statement was designed to signal that our organization subscribes to the idea that microfinance is effective. By citing a published study from prominent authors in the area of microfinance, we also tried to signal that the results and possibly consequences of impact evaluation were not trivial.

The negative treatment is identical to the positive treatment but suggests that microfinance is not effective:

> "Academic research suggests that microfinance is ineffective. The results of a recent scientific study show that microcredit loans have no effect on business growth and subjective well-being, nor are there disproportionate benefits in targeting women with microcredit loans (Karlan and Zinman 2011, "Microcredit in Theory and Practice," *Science*). These results are compelling to us, and we wish to learn more so we can further assist those in need."

The overall email and much of the treatment language is identical across all conditions. We even cite the same authors who report different conclusions in two separate studies. Of course, the control email left out any treatment language entirely.

Additional Email Protocol

For all emails, we attempted to write the subject line strategically to maximize the likelihood of response. First round emails were given the subject “Potential Partnership”; second and third round emails were given the subject “Potential Partnership Reminder”.

Emails were addressed to the MFI’s legal name or abbreviation. We sent all emails from an email address associated with a university. By sending it from an educational address, we hoped to increase the validity of our invitation. In addition to the email address, we signed all of the emails in the name of the director of PEDL; we anticipated that the emails were most likely to be opened if the sender was an individual. The emails were signed accordingly with a full electronic signature including the director’s professional title, address, and a link to the organization’s website (see the Appendix for the full text of the emails).

Because we did not always receive responses to the first inquiry, we followed up only with those providers that did not respond. For example, if Acción responded to the first email, we did not send it a reminder email. We followed up only two times with each provider before coding them as non-respondents. The reminder emails were prefaced with the following text:

> “I sent an email <date email was sent> regarding a potential partnership. Below is a copy of the email I sent that day. I just wanted to confirm you received the invitation.”

To enable more controlled execution of the experiment, we sent all the emails from a proxy server. This approach allowed the timing to be consistent and also enabled us to receive confirmation that all of the emails had been sent. For copies of all the emails see the Appendix.

Estimation Strategy

Because we randomized the assignment of experimental conditions, we can employ simple difference-in-means tests. In expectation, all other covariates should be equally balanced across conditions, thus allowing us to identify the treatment effect of the information we provide in the emails referenced earlier. Three outcomes are possible: non-response, decline, and accept. Of course, the latter two possibilities are contingent on receiving a response.

Although the difference-in-means tests should be sufficient to probe whether one or both of the treatments alter the effect relative to the control, we nonetheless estimate a multinomial probit model, separate logit models, as well as a selection model. The multinomial probit model sets response as the base category and then estimates the likelihood of declining or accepting the invitation. The logit models set up a series of dichotomies between the treatment 1 and control, treatment 2, and control, and treatment 1 vs. treatment 2. We consider each of the possibilities on the response, decline, and accept outcomes. And finally the selection model allows us to incorporate response and outcome into the same category. Because we do not have additional information with which to identify the separate stages, we use the model designed by Sartori (2003). As we will highlight below, the results are consistent across model specifications.

Results

Table 2 reports the results of the basic comparisons across treatment and control conditions for the response, decline, and accept outcomes. We report the numbers of observations in each category, the percentages, as well as p-values representing the confidence in the difference between different sets of conditions. The p-values for statistically significant differences are bolded.

On basic response rates, there is no statistically meaningful difference between the positive prompt and the control. Although there was a higher response rate based on the positive message (10.44% vs. 8.37%), the result is not statistically significant. Similarly, there is no statistical difference between the negative prompt and the control (6.06% vs. 8.37%). It is interesting that fewer MFIs responded when faced with negative information. The differences between the treatment and control are not significant statistically, however. Interestingly, there is a strong and statistically meaningful difference between the positive and negative treatments (10.44% and 6.06%; $p = 0.015$).

The next column shows that there were very few MFIs that declined the invitation to receive additional information on a partnership for an impact evaluation. This results holds regardless of the condition (control = 4, positive = 3, negative = 5). None of these differences across experimental conditions even come close to standard levels of statistical significance. The results that begin to emerge from the response category are thus not reflected in the decline outcome, but rather in the acceptance.

The numbers and proportions of acceptances do appear to change in response to the experimental condition. The difference between the positive treatment and the control (9.81% vs. 7.53%) is not statistically significant ($p = 0.210$), but the differences between the negative treatment and control (4.98% vs. 7.53%; $p = 0.107$), as well as the positive and negative treatments (9.81% vs. 4.98%; $p = 0.005$), are significant below or close to conventional levels.

Table 2: Contingency Table of Outcomes across Conditions

Condition	*N*	Response	Decline	Accept
Control	478	40	4	36
Proportion		8.37%	0.84%	7.53%
Positive	479	50	3	47
Proportion		10.44%	0.63%	9.81%
P-value vs. Control		0.273	0.703	0.210
Negative	462	28	5	23
Proportion		6.06%	1.08%	4.98%
P-value vs. Control		0.173	0.700	0.107
P-value vs. Positive		**0.015**	0.447	**0.005**
Total	1,419	118	12	106
		8.32%	0.85%	7.47%

These basic results are telling. When MFIs received an email indicating that microfinance work has been evaluated positively, they are more likely to respond and more likely to request additional information on partnering on an impact evaluation of their work than if they received information indicating that microfinance work may be ineffective. This offers some support for the conclusion that organizations experience confirmation bias. Indeed, only about 20% of the acceptances (23 of 106) occurred in response to the negative prompt. Nearly half of the acceptances (47 of 106) occurred in response to positive information. The difference is striking. But does it hold up when subject to additional scrutiny.

We conducted a randomization check to consider whether other observable factors were balanced equally across conditions. This check demonstrates that the randomization occurred as expected. Only one of nine variables (age of organization) was related to treatment assignment. To account for the possibility that this is affecting the results, we conduct additional analysis.

First, we estimated separate multinomial probit models for each of the experimental conditions. Table 3 displays these results. The findings confirm what we learn in the basic difference-in-means tests showing that receiving the negative prompt makes MFIs on average less likely to request additional information on the offered partnership for an impact evaluation than when receiving the control ($p < 0.1$). It also shows a very strong difference between the positive and negative conditions ($p < 0.01$).

Table 3: Multinomial Probit Table of Outcomes across Conditions

Treatments	**Response**	**Decline**	**Accept**	***N***
Positive	Base	-0.104	0.198	945
	Base	(0.368)	(0.164)	
Constant	Base	-3.076***	-1.953***	
	Base	(0.287)	(0.143)	
Negative	Base	0.064	**-0.329***	930
	Base	(0.338)	**(0.186)**	
Constant	Base	-3.119***	-1.944***	
	Base	(0.293)	(0.147)	
Pos. vs. Neg.	Base	-0.167	**0.522*****	929
	Base	(0.354)	**(0.179)**	
Constant	Base	-3.125***	-2.365***	
	Base	(0.260)	(0.155)	

Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
These models include the age variable, but the results are not reported here. It is insignificant in all of the regression models.

We also considered the comparisons as a set of logit models on the outcome variables separately. Like the multinomial model, we compared the negative prompt to control, positive prompt to control, and positive prompt to negative prompt, but in the basic logit models we conduct each of these regressions separately. Table 4 displays the results of these analyses.

Table 4: Logit Results for Accept, Reject, Response

Variables	***Response***	***Response***	***Response***
Positive vs. Control	0.236		
	(0.225)		
Negative vs. Control		-0.407	
		(0.263)	
Positive vs. Negative			**0.632****
			(0.249)
Age	-0.203	-0.246	0.131
	(0.232)	(0.269)	(0.244)
Constant	-2.296***	-2.278***	-2.817***
	(0.198)	(0.205)	(0.218)
Variables	***Accept Offer***	***Accept Offer***	***Accept Offer***
Positive vs. Control	0.284		
	(0.234)		
Negative vs. Control		**-0.508***	
		(0.284)	
Positive vs. Negative			**0.780*****
			(0.267)
Age	-0.166	-0.235	0.142
	(0.240)	(0.287)	(0.257)
Constant	-2.425***	-2.397***	-3.038***
	(0.207)	(0.216)	(0.239)
Variables	***Decline Offer***	***Decline Offer***	***Decline Offer***
Positive vs. Control	-0.323		
	(0.775)		
Negative vs. Control		0.236	
		(0.690)	
Positive vs. Negative			-0.548
			(0.735)
Age	-0.593	-0.288	0.028
	(0.849)	(0.727)	(0.735)
Constant	-4.541***	-4.646***	-4.514***
	(0.589)	(0.613)	(0.512)
Observations	945	930	929

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results are consistent with the previous findings reported in Tables 2 and 3. The significant results are again bolded and show robustness across specifications. As one final check, we consider the possibility that accepting or declining is a two-stage process. We therefore estimate a selection model in which assignment to treatment is the independent variable generating response as well as outcome (decline or accept). Because we do not have enough information to satisfy an exclusion restriction, we use the selection model developed by Sartori (2003). These results (reported in Table 5) are largely consistent with those reported earlier. The effect of the negative condition on acceptance is in the same direction but only marginally significant ($p = 0.107$). That result is consistent with the difference-in-means results reported above.

Table 5: Selection Model of Response and Accept/Decline Outcome

Treatments	Response	Accept	Resp. Constant	Comp. Constant	N
Pos. v. Control	0.124	0.145	-1.381***	-1.437***	957
	(0.113)	(0.115)	(0.082)	(0.085)	
Neg. v. Control	-0.169	-0.210	-1.381***	-1.437***	940
	(0.124)	(0.130)	(0.082)	(0.085)	
Pos. v. Neg.	0.293**	0.355***	-1.550***	-1.647***	941
	(0.120)	(0.126)	(0.092)	(0.098)	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Conclusion

We began with the question of whether development institutions learn and update from the wealth of evaluation information currently coming online. This experiment tested the possibility that MFIs, one of the most prolific types of development organizations, were more willing to accept information that confirmed the efficacy of their work and reject information that called it into question. Specifically, we hypothesized that MFIs would be less willing to accept our invitation in response to the treatment emails providing negative information about micro-finance when compared to the positive treatment. Although we do not explicitly test the causal mechanisms in this study, the results are at least consistent with the conjecture that organizations engage in significant confirmation bias when confronted with new information about their field.

Giving MFIs information on the effectiveness of microfinance could have reinforced their belief in the industry, raised questions about MFI efficacy, or had no effect. Since we assume MFIs believe in their cause, providing an MFI with positive scientific information on microfinance appears to reinforce previously held beliefs. The MFIs on average seemed to engage in confirmation bias by agreeing with the content of the new information; they were more likely to respond favorably to receiving additional information on a possible impact evaluation partnership. The opposite is true for the MFIs that received negative scientific information on microfinance. The evidence is consistent with the proposition that the information that microfinance is ineffective ran contrary to the MFIs staff members' previously held beliefs in a way that induced cognitive dissonance.

When representatives of an organization experience cognitive dissonance, they could either be open to updating their methods or they could rationalize their organizational behavior. The results of this study suggest that MFI staffers assigned to the negative treatment may be significantly less interested in updating. It seems that MFI representatives in the negative condition were prone to ignore the new information. They may have rationalized that their current methods of alleviating poverty are effective and thus ignored our invitation for a possible impact evaluation.

While we are fascinated by the results, we are also disappointed by their implications. Multiple studies in social psychology show that humans are susceptible to confirmation bias (Westen et al. 2006; Vallone et al. 1985). However, we hoped that MFIs' organizational structure would transcend this human tendency, especially given that MFIs' chief purpose is poverty alleviation. We also hoped that the MFIs that received the negative treatment would have more of a desire to at least explore the idea of an impact evaluation. We thought that if MFIs were shown some scientific evidence suggesting that current methods may not be effective, they would want to discover if their specific practices could be improved.

A more optimistic interpretation, on the other hand, would point to the five percent of subjects in the negative treatment condition that accepted the invitation for additional information about a partnership to perform a randomized evaluation. They accepted the invitation despite the fact that they received information suggesting that a negative result might be found. The invitation may have also signaled that the researchers proposing the partnership may have themselves been biased against microfinance. Yet a non-trivial share of MFIs was still willing to work with the team to learn their own organizations' effectiveness. This provides some grounds for optimism about the willingness of some development organizations to update.

However, if these results correctly apply to learning in development more generally, on balance they are not good news. With the recent evaluation revolution in development, there is substantial hope that practices will be updated based on the findings and that development activities will subsequently become more effective. But a missing step has been overlooked between the execution of impact evaluations and the planning of new interventions: the willingness of organizations to update based on scientific information has been assumed and not established. If organizations continue to seek confirmation of priors, then moving from evaluation to better interventions may take much longer than expected.

Of course, further research is necessary to determine whether other NGOs and development organizations behave consistently with MFIs. This industry may be unique. We suspect, however, that the extensive findings from social psychology and neuroscience on confirmation bias will extend to additional organizations involved in poverty relief. But additional research will need to establish the scope of the problem. On balance, however, it appears that even if all of the other stipulated problems with randomized evaluations can be addressed (and we happen to believe they can be), the willingness of organizations to update based on the findings from RCTs may still attenuate the effectiveness of field experiments in development. Future research should therefore also explore the conditions that enable organizational openness to new information and willingness to update established practices accordingly.

References

A short history of Grameen Bank. Grameen Communications. http://www.grameen-info.org/index.php?option=com_content&task=view&id=19&Itemid=114 (accessed November 12, 2011).

Armendáriz, Beatriz, and Jonathan Morduch. 2010. Measuring impacts. In *Economics of microfinance*, 267. Cambridge, Massachusetts: MIT Press.

Aronson, Elliot. 1969. The theory of cognitive dissonance: A current perspective. In *Advances in experimental social psychology* 4, no. 2-32. New York: Academic Press Inc.

Banerjee, Abhijit V., and Esther Duflo. 2009. The Experimental Approach to Development Economics. *Annual Review of Economics* 1, 1: 151-178.

Banerjee, Abhijit V., and Esther Duflo. 2010. Giving credit where it is due. *Journal of Economic Perspectives* 24, no. 3: 61-80.

Banerjee, Abhijit V., and Esther Duflo. 2011. *Poor economics: A radical rethinking of the way to fight global poverty.* New York: Public Affairs.

Banerjee, Abhijit V., Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2009. The Miracle of Microfinance?: Evidence from a Randomized Evaluation. Massachusetts Institute of Technology.

Copestake, James, Nathanael Goldberg and Dean Karlan. 2009. Randomized control trials are the best way to measure impact of microfinance programs and improve microfinance product designs. *Enterprise Development and Microfinance* 20, no. 3 (September): 167-176.

Deaton, Angus S. 2010. Instruments, Randomization, and Learning about Development. *Journal of Economic Literature* 48(2): 424-55.

Duflo, Esther and Michael Kremer. 2004. "Use of Randomization in the Evaluation of Development Effectiveness." In *Evaluating Development Effectiveness*. Edited by George K. Pitman, Osvaldo N. Feinstein, and Gregory K, Ingram. New York: Transaction Publishers.

Faiola, Anthony. 2009. "Some World Bank Health Programs Ineffective, Report Says." Washington Post 1 May 2009, accessed 15 March 2013 athttp://articles.washingtonpost.com/2009-05-01/business/36842038_1_world-bank-emergency-programs-nutrition-and-population.

Festinger, Leon. 1957. *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Garikipati, Supriya. 2008. The impact of lending to women on household vulnerability and women's empowerment: Evidence from India. *World Development* 36, no. 12:2620-2642.

Hansen, Henrik and Finn Tarp. "Aid Effectiveness Disputed." *Journal of International Development* 12, 3: 375-398.

Helms, Brigit. 2010. Microfinancing changes lives around the world—Measurably. *Seattle Times*. 7 April 2010. Accessed 17 March 2013 at http://seattletimes.com/html/opinion/2011545639_guest08helms.html.

Holvoet, Nathalie. 2005. The impact of microfinance on decision-making agency: Evidence from south india. *Development and Change* 36, no. 1:75-102.

Karlan, Dean S. 2009. Cairo Evaluation Clinic: Thoughts on Randomized Trials for Evaluation of Development (June 26). Yale University Economic Growth Center Discussion Paper No. 973; Yale Economics Department Working Paper No. 65. Accessed 17 March 2013 at http://ssrn.com/abstract=1426130

Karlan, Dean, and Jonathan Zinman. 2011. Microcredit in theory and practice: Using randomized credit scoring for impact evaluation. *Science* 332, no. 6035:1278-1284.

Karlan, Dean and Jonathan Zinman. 2010. Expanding credit access: Using randomized supply decisions to estimate the impacts. *The Review of Financial Studies* 23, no. 1:433-464.

Legovini, Arianna. 2010. "Development Impact Evaluation Initiative: A World Bank-Wide Strategic Approach to Enhance Developmental Effectiveness." Washington, D.C.: World Bank. Accessed 17 March 2013 at http://siteresources.worldbank.org/INTDEVIMPEVAINI/Resources/3998199-1286546178578/74657781291306572028/Legovini_dime_paper_ext.pdf.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology* 37, no. 11:2098-2109.

Morduch, Jonathan. 1998. Does Microfinance Really Help the Poor: New Evidence from Flagship Programs in Bangledesh. Department of Economics and HIID Harvard University and Hoover Institution Stanford University.

Mosley, Paul. 1986. "Aid Effectiveness: The Micro-Macro Paradox." *The Institute of Development Studies Bulletin* 17, 2 (April): 22-27.

Picciotto, Robert. 2012. Experimentalism and development evaluation: Will the bubble burst? *Evaluation* 18, no. 2 (April): 213-229.

Pitt, Mark M., Shahidur R. Khandker, and Jennifer Cartwright. 2006. Empowering women with micro finance: Evidence from Bangladesh. *Economic Development and Cultural Change* 54, no. 4:791-831.

Rahman, Aminur. 1999. Microcredit initiatives for equitable and sustainable development: Who pays? *World Development* 27, no. 1:67-82.

Rajan, Raghuram G. and Arvind Subramanian. 2008. "Aid and Growth: What Does the Cross-Country Evidence Really Show?," *The Review of Economics and Statistics* 90, 4: 643-665.

Ravallion, Martin. 2001. The mystery of the vanishing benefits: An introduction to impact evaluation. *World Bank Econ Review* 15, no. 1: 115-140.

Rodrik, Dani. 2008. "The New Development Economics: We Shall Experiment, but How Shall We Learn?" Harvard Kennedy School Working Paper RWP08-055.

Sartori, Anne E. 2003. An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions. *Political Analysis* 11 (2): 111-38.

Steele, Claude M., and Thomas J. Liu. 1983. Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology* 45, no. 1:5-19.

Stone, Jeff, and Joel Cooper. 2001. A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology* 37, no. 3.

Subramanian, Arvind. 2011. "Nicholas Kristof and Aid." Washington, D.C.: Center for Global Development. Accessed 17 March 2013 at http://blogs.cgdev.org/globaldevelopment/2011/05/nicholas-kristof-and-aid.php.

Vallone, Robert P., Lee Ross, and Mark R. Lepper. 1985. The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology* 49, no. 3:577-585.

United States Agency for International Development. 2011. *USAID Evaluation Policy*. Washington, D.C.: USAID. Accessed 15 March 2013 at http://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf.

Westen, Drew, Pavel S. Blagov, Keith Harenski, Clint Kilts, and Stephan Hamann. 2006. Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 U.S. presidential election. *Journal of Cognitive Neuroscience* 18, no. 11:1947-1958.

Wesselink, Ben. 2010. Directory of development organizations. http://www.devdir.org/ (accessed February 15, 2012).

Yunus, Muhammad. 2007. Banker To The Poor: Micro-lending and the Battle Against World Poverty. New York: Perseus Books Group.

Appendix

Negative Treatment Email

<MFI Name>,

I am contacting you as director of the Political Economy and Development Lab (PEDL) of Brigham Young University. Founded in 2008, we study the relationship between politics and economics with a special focus on global development, including the impact microfinance institutions have on the poor. We are seeking to assess the interest of qualified microfinance institutions in possible partnerships to perform impact evaluations. We understand that you provide microcredit loans in <country>.

Academic research suggests that microfinance is ineffective. The results of a recent scientific study show that microcredit loans have no effect on business growth and subjective well-being, nor are there disproportionate benefits in targeting women with microcredit loans (Karlan and Zinman 2011, "Microcredit in Theory and Practice," *Science*). These results are compelling to us, and we wish to learn more so we can further assist those in need.

As I am sure you understand, in order to improve MFI processes we must carefully evaluate impact. This is best accomplished through scientific evaluations using random assignment. Should grant funding, balance of prior commitments, and mutual interest allow, would your organization be interested in receiving more information about potentially partnering with PEDL on a future impact evaluation?

Please understand that this is not an invitation for immediate partnership. We have several other commitments to partners currently and thus can pursue only a few new joint projects going forward – and those will, of course, depend on future grant funding. But we are hoping to gauge your possible interest.

Due to numerous research commitments, we would prefer to communicate – at least through this initial phase – through email. In order that we can keep better track of your response, please reply directly to this email. We hope to hear from you soon.

Thank you very much for attention to this inquiry.

Sincerely,

Daniel L. Nielson
Associate Professor & Director, Political Economy and Development Lab
Political Science Department
Brigham Young University
Provo, UT 84602-5545
Political Economy and Development Lab

Positive Treatment Email

<MFI Name>,

I am contacting you as director of the Political Economy and Development Lab (PEDL) of Brigham Young University. Founded in 2008, we study the relationship between politics and economics with a special focus on global development, including the impact microfinance institutions have on the poor. We are seeking to assess the interest of qualified microfinance institutions in possible partnerships to perform impact evaluations. We understand that you provide microcredit loans in <country>.

Academic research suggests that microfinance is effective. The results of a recent scientific study show that microcredit loans have a positive effect on economic self-sufficiency and subjective well-being of borrowers, including the decision making power women have in the home (Karlan and Zinman 2010, "Expanding Credit Access," *Review of Financial Studies*). These results are compelling to us, and we wish to learn more so we can further assist those in need.

As I am sure you understand, in order to improve MFI processes we must carefully evaluate impact. This is best accomplished through scientific evaluations using random assignment. Should grant funding, balance of prior commitments, and mutual interest allow, would your organization be interested in receiving more information about potentially partnering with PEDL on a future impact evaluation?

Please understand that this is not an invitation for immediate partnership. We have several other commitments to partners currently and thus can pursue only a few new joint projects going forward – and those will, of course, depend on future grant funding. But we are hoping to gauge your possible interest.

Due to numerous research commitments, we would prefer to communicate – at least through this initial phase – through email. In order that we can keep better track of your response, please reply directly to this email. We hope to hear from you soon.

Thank you very much for attention to this inquiry.

Sincerely,

Daniel L. Nielson
Associate Professor & Director, Political Economy and Development Lab
Political Science Department
Brigham Young University
Provo, UT 84602-5545
Political Economy and Development Lab

Control Treatment Email

<MFI Name>,

I am contacting you as director of the Political Economy and Development Lab (PEDL) of Brigham Young University. Founded in 2008, we study the relationship between politics and economics with a special focus on global development, including the impact microfinance institutions have on the poor. We are seeking to assess the interest of qualified microfinance institutions in possible partnerships to perform impact evaluations. We understand that you provide microcredit loans in <country>.

As I am sure you understand, in order to improve MFI processes we must carefully evaluate impact. This is best accomplished through scientific evaluations using random assignment. Should grant funding, balance of prior commitments, and mutual interest allow, would your organization be interested in receiving more information about potentially partnering with PEDL on a future impact evaluation?

Please understand that this is not an invitation for immediate partnership. We have several other commitments to partners currently and thus can pursue only a few new joint projects going forward – and those will, of course, depend on future grant funding. But we are hoping to gauge your possible interest.

Due to numerous research commitments, we would prefer to communicate – at least through this initial phase – through email. In order that we can keep better track of your response, please reply directly to this email. We hope to hear from you soon.

Thank you very much for attention to this inquiry.

Sincerely,

Daniel L. Nielson
Associate Professor & Director, Political Economy and Development Lab
Political Science Department
Brigham Young University
Provo, UT 84602-5545
Political Economy and Development Lab