Strategic Management Journal
Strat. Mgmt. J., **37**: 116–132 (2016)
Published online EarlyView in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2449
Received 25 June 2014; Final revision received 27 November 2014

FIELD EXPERIMENTS IN STRATEGY RESEARCH

AARON K. CHATTERJI,[1]* MICHAEL FINDLEY,[2] NATHAN M. JENSEN,[3] STEPHAN MEIER,[4] and DANIEL NIELSON[5]
[1] *Fuqua School of Business, Duke University, Durham, North Carolina, U.S.A.*
[2] *Department of Government, University of Texas at Austin, Austin, Texas, U.S.A.*
[3] *Department of International Business, The George Washington School of Business, Washington, District of Columbia, U.S.A.*
[4] *Graduate School of Business, Columbia University, New York, New York, U.S.A.*
[5] *Department of Political Science, Brigham Young University, Provo, Utah, U.S.A.*

Strategy research often aims to empirically establish a causal relationship between an independent variable and a dependent variable such as firm performance. For many important strategy research questions, however, traditional empirical techniques are not sufficient to establish causal effects with high confidence. We propose that field experiments have potential to be used more widely in strategy research, leveraging methodological innovations from other disciplines to address persistent puzzles in the literature. We first review the advantages and disadvantages of using field experiments to answer questions in strategy. We define two types of experiments, "strategy field experiments" and "process field experiments," and present an original example of each variety. The first study explores the liability of foreignness and the second study tests theories regarding corporate culture.

INTRODUCTION

Assessing the direction of causality correctly is crucial, not just for the advancement of strategy as a scholarly field, but also for the guidance we offer to practitioners. However, empirically establishing causality is challenging in strategy and many other fields (e.g., Hamilton and Nickerson, 2003). Some of the most interesting phenomena in strategy research that could be related to firm performance, such as the establishment of a strong corporate culture, the hiring of a new CEO, or the development of a new capability, do not occur randomly across firms. Other important attributes of organizations, such as founding conditions or country of origin, are likely time invariant and correlated with important unobserved factors.

Thus, although strategy scholars have made considerable progress on these topics, it remains difficult to conclude definitively that stronger culture drives greater profits, that insider CEOs do better than outsiders, or that capabilities are the source of heterogeneity in performance across firms. Similarly, while topics like imprinting and the liability of foreignness have been studied extensively, it has been challenging to disentangle founding conditions or country of origin from other important factors. Whereas various econometric techniques, such as employing an instrumental variable (e.g., Bascle, 2008), can sometimes be used to address

Keywords: field experiments; research methods; culture; liability of foreignness; foreign direct investment
*Correspondence to: Aaron K. Chatterji. The Fuqua School of Business, Duke University 100 Fuqua Drive, Box 90120 Durham, NC 27708. E-mail: ronnie@duke.edu
This paper is the result of a merger between two earlier working papers: "A Field Experiment on the Liability of Foreignness: Country of Origin and U.S. Inward Investment Promotion" (by Findley, Jensen, and Nielson) and "Field Experiments in Strategy Research" (by Chatterji and Meier). Each paper had an original field experiment: The experiment by Findley, Jensen and Nielson is "Experiment 1" in this version and the experiment by Chatterji and Meier is "Experiment 2". We thank the editors and four referees for very helpful comments.

these methodological concerns in strategy research, these hurdles persist. In this paper, we propose that field experiments are a promising methodology to address some of the limitations of prior work and shed light on many of the foundational questions in the field. We view this methodology as a complement, not a substitute, for existing strategy research methods, since field experiments have strengths and limitations that are often different from our traditional techniques.

Experiments have long been used in disciplines such as social psychology and marketing to establish causality. In an experiment, the researcher randomly assigns research subjects to a treatment and a control group—in the same way that patients are assigned to treatment and placebo groups in randomized controlled trials in medical research. Building on this tradition, field experiments, the application of experimental methods outside a traditional lab, have recently led to a research method revolution in economics, a field that had been historically resistant to experimental methods.[1] This trend is especially prevalent in development economics (for overviews, see Banerjee and Duflo, 2009; Duflo, Glennerster, and Kremer, 2006), behavioral economics (for overviews, e.g., Harrison and List, 2004; Levitt and List, 2009), public policy (Ludwig, Kling, and Mullainathan, 2011), and to a certain degree in organizational economics (see Bandiera, Barankay, and Rasul, 2011, for an overview). This revolution has not yet reached the field of strategy, however, for reasons we discuss below. The figure in Appendix A1.1 of File S1 shows that, although the number of papers published based on "field experiments" or "randomized controlled trials" has exploded in economics journals, no corresponding trend has occurred in strategy and management journals. In Appendix A1.2 of File S1, we provide a list of field experiments published and forthcoming at *SMJ* and provide a brief summary of how the results contribute to the field. The small number of field experiments in *SMJ* provides a window into the potential of this methodology in strategy research.

We contend that strategy scholars are especially well-placed to utilize field experiments, since many important research questions are unaddressed by other fields using this method.[2] In particular, strategy research places significant emphasis, especially relative to economics, in explaining firm heterogeneity. For example, many questions of interest to strategy researchers but traditionally less so for organizational economists, are related to firm capabilities (Helfat *et al.*, 2009; Teece *et al.*, 1997). As we describe below, this gap provides an important opportunity for strategy researchers to formulate and test original theory using field experiments. However, the field experiment methodology has specific advantages and disadvantages for strategy research in particular, which we also review below.

To guide future scholarship, we define two general types of experiments, "strategy field experiments" and "process field experiments," to explicate the different approaches strategy scholars can employ to address key questions in the field. We conclude that, although one might assert at first glance that field experiments have limited utility for strategy research, the opportunities to apply these methods are widespread.

Next, we detail two original field experiments, the first a "strategy field experiment" and the second a "process field experiment," that illustrate the great potential but also the drawbacks of using field experiments for strategy research. Lastly, we reflect on the path forward for field experiments in strategy research and discuss what kind of research agenda could be formulated to guide future scholarship.

METHODOLOGICAL INNOVATION: ANSWERING "BIG" QUESTIONS WITH "SMALL" EXPERIMENTS

Advantages of field experiments in strategy research

Field experiments as a method allow the design and implementation of creative treatments to answer relevant questions that are otherwise very hard to address. While the ideal context to study the effect of interest might not be available in observational data, researchers using field experiments can create their own exogenous variation to identity causal relationships cleanly. This feature allows

[1]Psychology, a discipline that has traditionally used lab experiments, is also seeing a surge in field experiments. See Shadish and Cook (2009) for an overview. Similarly, there has been increasing interest in field experiments in marketing research. As one example, *The Journal of Marketing Research* issued a Call for Papers for a *Special Issue on Field Experiments* in Spring 2014.

[2]We will focus on field experiments and not experiments conducted in the laboratory. Croson *et al.* (2007) argues that these laboratory experiments should also be part of the toolkit of strategy researchers. See Harrison and List (2004) for a taxonomy of the various kinds of field experiments.

DOI: 10.1002/smj

researchers to do question-driven research as opposed to being constrained by existing data, an important aspiration for the field. As one example, there might be little within-firm variation in national origin in the typical observational dataset, but in our first experiment we can manipulate perceptions of the audience about the origins of a company to test theory about the liability of foreignness.

Second, field experiments are ideally suited to assess specific processes and activities inside firms, where our traditional data sources are unable to provide much detail. As we discuss below, the outcome variables of interest in these field experiments need not be firm performance, but could be antecedents based on prior theory. Using this approach, the results of field experiments have great potential to reveal actionable insights for managers seeking to improve firm performance through the development of capabilities or, as we discuss in our second experiment, establishing a particular corporate culture.

More broadly, experiments make it possible to vary one factor at a time and therefore provide "internally" valid estimates. In the spirit of Shadish, Cook, and Campbell (2002), field experiments allow the researcher to be more confident that any difference between the treatment and control means is due to the intervention. This feature creates the ability to use multiple conditions to disentangle forces that covary in the real world. Importantly, however, this method does not automatically confer external validity (Lynch, 1982, 1983), which we discuss in detail below.

Challenges for field experiments in strategy research

Despite these positive attributes, there are several drawbacks to the use of field experiments.

External validity

First, since this methodology promises strong internal validity through the random assignment of treatment and control conditions, it might be intuitive to think that there are advantages related to external validity as well. However, Lynch (1982) points out that no research methodology in and of itself can generate externally valid results. External validity can be thought of as whether the results from the experiment can be generalized to some population/context and/or can be generalized across populations/contexts (Cook and Campbell, 1979). Most importantly, "generalizability can only be judged on the basis of replication across settings, subjects, stimuli, and responses." (Dipboye, 1990: 26).

One of the fundamental challenges that prevents generalization across settings from a field experiment is that various background factors can substantially interact with the treatment, unbeknownst to the researcher (Lynch, 1983). This kind of oversight can create challenges to external validity via two channels. First, researchers might not test for the background factor and treatment interactions, committing an aggregation fallacy because of unobserved heterogeneity. As a result, researchers report the main effect of the treatment but the "simple effects" vary according to unobserved background factors (Hutchinson, Kamakura, and Lynch, 2000). In the second case, researchers may hold background factors fixed when they actually interact with the treatment, leading to treatment effects that would be different if the background variables were held constant at another level (Lynch, 1982).

While it is impossible to generate perfect external validity from field experiments, there are a few strategies to mitigate these concerns. First, the treatment should be replicated in different contexts. Second, process field experiments (see below) that test the underlying mechanisms more directly should be conducted, since insights about processes advance our understanding of firms more generally. Indeed, while typical field experiments excel at generating causal effects in which researchers can place high confidence, they are limited in their ability to indicate the causal mechanisms that link interventions to outcomes. Process field experiments, by randomly assigning the mechanisms themselves, provide a possible way forward (see also Ludwig *et al.*, 2011) for detailed discussion with regard to policy field experiments). Third, theoretical development can help anticipate the specific ways that an observed treatment effect might change if extrapolated to a specific, but different situation (see Lynch, 1999).[3]

[3]Closely related to the problem of external validity is a concern expressed by Heckman (1992) that the organizations that agree to participate in an experiment may be different from those that do not (see Allcott and Mullainathan ,2012, for an example). While there are some limited strategies to mitigate this challenge, it is still difficult for researchers to assess the extent of this bias in interpreting the results of their experiments.

Feasibility

There are also additional challenges to field experiments for strategy scholars in particular. Strategy scholars have traditionally sought to explain firm performance by variation in industry structure and firm capabilities. Very few field experiments could ever be large enough or sustained for long enough to generate appreciable differences in firm performance. Next, it is quite difficult to manipulate attributes of an entire industry or firm positions experimentally. Finally, while working inside a single firm may provide a promising vantage point to observe firm capabilities, prior work has found that competitive advantage relies on bundles of resources and capabilities that are endogenously linked to one another and possibly impossible to separate even in a well-designed study (e.g., Siggelkow, 2001).

As a result, many hypothetical research designs to answer important questions in the field are clearly impractical. For example, randomly assigning firms to different competitive positions in the market and then measuring their return on invested capital (ROIC) five years later is almost impossible. And randomly changing the culture of half of the business units in a firm and then measuring their subsequent performance compared to the other half of divisions, where the culture was presumably constant is also likely not feasible. More generally, if individual-level randomization is not feasible and randomization has to be done across larger units, like teams or business units, statistical power (or the lack thereof) might be a limitation. While there are unique opportunities to manipulate core attributes of firms experimentally (as we do in our first experiment), such instances may be rare. Further, as field experiments in strategy are usually conducted inside firms, they can be difficult to implement since enlisting employees to help manage the experiment and securing top management buy-in are not always easy.

The examples above constitute a reasonable critique for why field experiments cannot be applied universally in strategy research. However, although running experiments to estimate directly the effect of certain strategic choices on firm performance may be too impractical, running experiments that teach us about the activities that conceptually underlie how those strategic choices relate to firm performance is far more feasible. In other words, methodological innovations are unlikely to occur that will allow us to assign randomly corporate culture to firms and observe their subsequent performance. However, carefully designed experiments could manipulate one of the key processes that underpin culture and measure its effect on a process outcome that affects performance. This approach can help us answer the "big" questions in strategy with "small" experiments (as has been argued in development economics, e.g., Banerjee and Duflo, 2009.

We believe that field experiments can make headway in strategy in a similar spirit: through targeting processes identified by theory and assessing shorter-term outcomes that matter for firm performance and other longer-term outcomes of interest. To make this approach explicit, we categorize experiments into two broad categories (inspired by a discussion about field experiments in public policy in Ludwig *et al.*, 2011): "strategy field experiments" and "process field experiments."

Two types of field experiments in strategy research

The two types of field experiments shed light on a strategic question using different approaches. In general, managers are interested in understanding how a choice X—the effect of span of control (e.g., Rajan and Wulf, 2006)—affects an outcome Y; i.e., firm performance. However, X affects Y through a mechanism M; i.e., span of control affects communication channels between managers and subordinates (Bloom *et al.*, 2009). The two types of experiments either study the effect of X on Y directly or focus on the mechanism M in order to learn about $X \rightarrow Y$.[4]

Strategy field experiments

We call experiments "strategy field experiments" if they test the effect of significant strategic choices, X, on firm performance, Y. For example, in studying

[4]The two types of field experiments are not always mutually exclusive. It is possible to design a field experiment using different treatment arms that can test all aspects of the chain X → M → Y. For an interesting example on pricing, see Karlan and Zinman (2009). However, we believe that the distinction we draw between different kinds of experiments is still useful in determining appropriate research designs and in choosing which questions to address.

the effect of span of control, a strategy field experiment would reduce the span of control for a given firm's middle managers in a randomly assigned subset of divisions. Then the researcher would compare the performance of treated and control business units. The advantages of a strategy field experiment are twofold: first, it directly tests an important strategic choice and measures the effect on a key performance metric. Second, because it tests a strategic choice in its entirety, this research design can take into account the *total effect* of the strategic choice. For example, reducing span of control might not only affect the manager directly, but also influence the selection of potential managers, allowing the firm to hire higher quality managers. A strategy field experiment can capture such broader effects.

However, strategy field experiments have at least two major drawbacks: First, they are often difficult to conduct; second, strategy field experiments cannot typically provide insights into the mechanism for why a certain strategic choice has an effect on performance. There is a complicated intervention, X, producing a treatment effect on Y, but the researcher cannot always identify the "active ingredient." As such, any results would be especially vulnerable to criticisms about generalizability or external validity discussed above.[5] Our perspective is that there are some contexts where the merits of strategy field experiments will outweigh the drawbacks. In other cases, process field experiments (discussed below) and other methodologies such as lab experiments, regression analysis and qualitative work can combine to shed light on important research questions.

We see our first field experiment as an illustration of a "strategy field experiment," which manipulates a fundamental characteristic of the firm. We manipulate the implied country of origin of a confederate company and test the effects on investment incentives offered by U.S. cities as a test of the literature on the liability of foreignness. While we are not directly testing a firm's market entry strategy, we are testing for differences in the cost of a foreign firm's choice to enter via direct investment relative to that of a domestic firm. While a strategy field experiment is especially appealing for this research question, our second study highlights a line of inquiry where process field experiments are more appropriate.

Process field experiments

Process field experiments do not test a strategic choice or firm attributes directly, but rather test the activity or mechanism that is the theoretical underpinning of the causal relationship of interest. For example, in the case of span of control, before designing a process field experiment, the researcher would need to reflect on prior theory for why reducing span of control would affect performance. Perhaps a smaller span of control facilitates better communication across hierarchy levels. A researcher interested in this domain could randomize access to a particular communication channel used by managers and subordinates and measure the impact on employee productivity in tackling a short-term task such as the completion of a single project. This approach would be a "small" but tractable experiment that relies on the theoretical connection between communication and span of control and the correlation between performance on a particular project and long-term performance.[6]

The clear disadvantage of process field experiments is that they only test the effect of the strategic lever indirectly and, as such, might miss more general effects of the strategic choice. However, process field experiments have significant advantages as well. First, process field experiments are generally more feasible to implement due to their smaller scale. Moreover, as Ludwig *et al.* (2011) point out, even small experiments on testing mechanisms can inform larger policy or strategic choices. Second, they provide evidence on the process underlying why a certain strategic choice is effective, which can enhance opportunities to apply the lessons learned into different contexts and inform theory. To discriminate between different theories, it is often not enough to know whether a certain strategy works, but to understand why it worked. Understanding the process more deeply will also improve our advice for managers, who may devise better ways to affect the outcome of

[5] For a similar critique about randomized controlled trials in development economics, see, Deaton (2010).

[6] Process field experiments should not be confused with "process evaluations." Process evaluations are used in field experiments to study the degree to which the intervention was implemented correctly and can thus help in the interpretation of the results (for an example, see Oakley *et al.*, 2006).

interest (i.e., working on opening communication channels rather than simply changing the span of control in the example above). Our categorization has clear connections to the concept of mediating variables, where the relationship between two variables X and Y might be mediated by M (see Zhao, Lynch, and Chen, 2010). Process field experiments are ideal when the relationship of interest is $M \rightarrow Y$. In such a case, manipulating *M* directly might be preferable to exploring just $X \rightarrow Y$ (see Bullock, Green, and Shang, 2010; Green, Ha, and Bullock, 2010; Spencer, Zanna, and Fong, 2005).

We see our second field experiment as an illustration of a "process field experiment". We cannot manipulate corporate culture in a large sample of firms and explore performance patterns over time. Instead, we test an important link in the causal chain between identity and cooperation that speaks to the broader importance of culture in organizations. To extrapolate from this style of experiment, one needs to tie variation in employee identity to culture and cooperative behavior among employees to firm performance. As discussed below, these two sets of linkages are well established in the literature.

Our distinction between strategy and process field experiments is similar in spirit to Calder *et al.*'s (1982) distinction in consumer behavior research. On one hand, they present "effects application" as a test of a general theory, similar to our strategy field experiments. Alternatively, they argue that "theory application," similar to our process field experiments, is grounded in different philosophical foundations and presents a different set of methodological challenges. They advise that researchers clearly identify the primary goal of a given study and design it accordingly. Following Calder *et al.* (1982), we recommend that strategy field experiments prioritize realism, while process field experiments evaluate mechanisms with the cleanest possible test. As Ludwig *et al.* (2011) argue, "mechanism" experiments can still be very valuable even if they do not mirror policy interventions that would actually be implemented.

In the next section, we describe both of our field experiments. Due to space constraints, we only briefly review the prior literature on these well-studied topics and place many of the details about how we designed and implemented these two studies in File S1, which also includes suggestions for implementing experiments and managing ethical concerns.

FIELD EXPERIMENT 1: LIABILITY OF FOREIGNNESS AND INVESTMENT PROMOTION

A core argument in international business scholarship is that firms are often confronted with substantial barriers to operating abroad, and these barriers shape a firm's strategic choices on when, how, and if to enter a foreign market (Hymer, 1976). This "liability of foreignness" is a broad definition that encompasses a number of factors that can disadvantage firms operating in overseas markets—e.g., geographic distance, cultural distance, economic nationalism (Zaheer, 1995).[7] Some evidence suggests that this liability of foreignness has been shown to have a substantial impact on the performance of foreign firms operating in host countries (Zaheer, 1995; Zaheer, 2002; Zaheer and Mosakowski, 1997).[8] This topic has important implications for the field of strategy. One of the most cited works on the topic notes in the first paragraph that, "This liability of foreignness has been the fundamental assumption driving theories of the multinational enterprise … " (Zaheer, 1995: 341).

Prior work has built an empirical foundation for the plausibility of the liability of foreignness, yet these studies employ indirect tests using observational data. In these instances, unobserved variables may confound results, self-selection may introduce bias, and thus causality is difficult to demonstrate. For example, two different studies published in *SMJ* on the liability of foreignness in the financial services industry found different results with regard to domestic and foreign bank performance (Miller and Parkhe, 2002; Nachum, 2003). These differences could arise from using indirect measures of the liability of foreignness. Other studies have taken a more direct approach to measuring the liability of foreignness, but have not settled the debate on the underlying mechanisms driving patterns in the data. For example, Mezias's (2002) article tested the liability of foreignness using data on labor lawsuit judgments in the United States, finding that foreign firms were more likely to be the subject of labor lawsuits than a matched set of domestic firms.

Moreover, empirical studies of the liability of foreignness face problems of endogeneity, where the firms that are most likely to overcome the

[7]A related literature is the work on home-bias in financial asset markets (established by French and Poterba, 1991).

[8]See Miller and Eden (2006) for a discussion on this topic.

DOI: 10.1002/smj

liability of foreignness are the firms we are most likely to study and compare against domestic firms. For this reason, it is quite possible that endogeneity works against finding liability of foreignness; for example, if firms that can mitigate this liability are more likely to be active abroad. For this and other reasons, it is very difficult to estimate the exact magnitude of liability of foreignness based on observational studies. We can overcome this issue and related challenges by focusing on the narrower question and using a field experiment.

In this field experiment we examine how the firm's country of origin affects whether government officials are willing to provide firm-specific incentives to a potential investor.[9] The use of business incentives—ranging from grants, tax holidays, to low-cost loans—are quite common across countries. In the United States, states and municipalities offer these incentives to firms, often in bidding wars to attract new investment. We explore how country of origin affects a firm's propensity to receive these incentives from government officials.

Sample

Our subjects of study are the 3,117 U.S. municipalities with populations above 10,000 that list a mayor, city manager, or other executive online. We refer to these municipalities, ranging from small towns and counties to large cities, as "cities" in this paper. We focus on municipalities because U.S. cities and states are increasingly active in the promotion of investment, providing as much as $80 billion per year in financial incentives to firms.[10] There are also enough U.S. cities to ensure a sufficiently large sample size.

Experimental protocol

Our first step was to identify a "confederate" firm that was considering an investment in a new manufacturing facility. The firm was willing to provide concrete details on its future investment, including projected numbers on job creation and capital investment based on the operations of two existing plants. We signed a confidentiality agreement with the firm, assuring that the name of the company would not be used in the experiment. In return for collaborating as a confederate in the study, the firm was offered our analysis of incentive pledges. In Appendix A2 of File S1 we briefly outline practical challenges as well as ethical considerations in this field experiment.

Next, we legally incorporated a consulting company that mimicked existing U.S.-based investment-promotion and incentive-management companies that often act as brokers between city economic development officials and firms exploring investments. These companies are generally small operations and often do not publish their client lists on their websites. We incorporated our company, Globeus Consulting, as an LLC in Delaware in 2013. We created a company website and a board of consultants (all academics willing to lend their names for the purposes of the experiment) and listed a company president. Three research assistants, using their actual names, served as "Associates" who directly contacted cities through email addresses registered through our website.[11] Our consulting company was built for the purpose of academic research and thus the authors received no compensation nor were they directly engaged in any further negotiations with the investor and city governments.[12] We installed web-tracking software that aided the analysis of our experiment.[13]

[9]While perhaps counterintuitive, there are examples where firms strategically choose to change their home market, and thus their "foreignness." Lenovo's acquisition of IBM's PC division or the recent spate of "tax inversions" can be viewed this way, though other less dramatic corporate decisions are also germane to our study. The narrowest interpretation of our experiment is a relatively clean estimate of how much national origin matters in a particularly significant corporate decision, investing in a new facility.

[10]Please see "The United States of Subsidies." http://www.nytimes.com/interactive/2012/12/01/us/government-incentives.html (accessed 23 June 2014).

[11]We were careful to minimize any personal interaction between our RAs and city representatives. Our interactions were largely limited to email communications with a clear protocol. Individual inquiries were handled by our lead research assistant or the authors. We highlighted that our confidentiality agreement between our consulting firm and client barred us from providing any additional information on the potential investor.

[12]We purposely designed our experiment to minimize the amount of time public officials would dedicate to our inquiry. While a back-and-forth negotiation would have increased the realism of our experiment, we believe that this would lead to a more extensive time investment by public officials.

[13]We included two forms of tracking. For each experimental condition we provided a unique link that routed the respondent to our website. Thus while all respondents viewed the same websites, we could track subjects from each group based on the routing link. Second, our website allows for IP tracking, enabling us to track both unique users and the location of the visitor. This allows us to compare the web activity (pages viewed, time on the site, etc.) by treatment group and individual.

DOI: 10.1002/smj

Next, we built a dataset of the mayor/executive and any listed economic development official for all 3,117 municipalities.[14] Then we emailed all of the cities details on the proposed investment, which would include $2 million in capital investment and 19 full-time jobs. These numbers came directly from our client and are near the lower bound of the type of investments that would be of interest to most cities. It is thus possible that our study is underreporting the general amount of interest in attracting investment.

In our email (see Appendix A3.1, File S1) we directed the client to a web form (via Qualtrics) to fill out specific details on the types of incentives offered. This impersonal interaction was designed to ensure that all cities were treated with the same message and to help limit the amount of time city officials dedicated to our inquiry. We include all of the Qualtrics questions in Appendix A3.2 of File S1. The main question we examine in this paper is whether or not a city offers grant dollars or loans to investors, although we were careful to highlight that these were nonbinding offers. Our approach contained two types of experimental interventions, only one of which we will present in detail in this article.[15]

While we were not authorized by our client to change any of the details of the investment, we were authorized to vary "the pitch" of our consulting company. We did this in two ways. First, we began each email with the following paragraph randomizing across the three conditions (treatment conditions highlighted below but not in the original email):

> I am an associate with GLOBEUS Consulting (see our website here Globeus.com). GLOBEUS is a new consulting firm that specializes in matching cities with prospective firms. I work in the GLOBEUS group focusing on investors based in **[TREATMENT 1: the United States/TREATMENT 2: Japan/TREATMENT 3: China]** and am contacting you to see if your city would be a good match for a client I am representing.

Second, we signed each email with the name of one of our three associates (research assistants) with a group email address and country team. Below is the email signature.

> Associate Name
> **[us/japan/china]**_client_team@globeusconsulting.com
> Selection & Incentives Associate Globeus Consulting—**[U.S./Japan/China]** Client Team www.globeusconsulting.com

We specifically focus on differences between U.S., Japan, and China due to previous work documenting public perceptions of Japan and China. In a public opinion study by the Pew Foundation, only 42 percent of Americans had a favorable view of China as compared to a 77 percent favorability of Japan (Pew Foundation, 2007). Jensen and Lindstädt (2013) find that the country of origin has a major impact on foreign direct investment (FDI) preferences. Using survey experiments in the United States and the United Kingdom they found very little difference in support for Japanese investment relative to the control condition of "foreign investment" without a country name. In contrast, they found considerable skepticism towards Chinese investment.

With this background in mind, we first test whether the city respondents will favor U.S. firms over foreign firms in general and then test whether city respondents are less likely to respond to Chinese investors compared to Japanese investors.[16] Prior to random assignment we stratified the subject pool on a set of covariates that might affect the outcome and thus implemented the experiment using block randomization on the criteria of state, population size (above or below median), government type (council-manager or other), elected vs.

[14] Our universe of cities is all cities with a population of 10,000 or more listed by ICMA. As part of our data collection efforts, we also coded the date of the mayoral/council elections. Due to the lack of information on the web we directly called a number of municipalities.

[15] In our first treatment, we varied the timing of when the investor was willing to announce their intent to invest, randomizing between two months before or one month after the mayor or city councilors' next reelection month (without mentioning the reelection). Note that our full-factorial randomization strategy ensures that our first treatment will not produce any bias in our analysis of our liability of foreignness treatment because the two treatments were randomly assigned independently of one another.

[16] The term "investor" is widely used in this context to refer to the management team of the operating company seeking to make an investment. While this word could potentially be misinterpreted by some participants, we believe that this term of art is well understood by city officials involved in economic development projects.

DOI: 10.1002/smj

Table 1. Treatment differences in Experiment 1

Treatment country	**US**	**Japan**	**China**
No of observations	1,047	1,034	1,048
Panel A: Response rates			
N responded	184	152	165
Response rate	0.176	0.147	0.158
Test of differences			
USA vs. foreign	−0.024 *	[-0.051, 0.004]; p = 0.091	
USA vs. Japan	−0.029 *	[-0.060, 0.003]; p = 0.0749	
Japan vs. China	−0.011	[-0.042, 0.020]; p = 0.5017	
Panel B: Proportion of cities offering incentives			
Offered incentives	103	101	87
Proportion	0.098	0.098	0.083
Test of differences			
USA vs. foreign	0.008	[-0.013, 0.030]; p = 0.461	
USA vs. Japan	0.001	[-0.025, 0.026]; p = 0.957	
Japan vs. China	0.015	[-0.010, 0.039]; p = 0.243	
Panel C: Means of natural logs of grant dollars offered			
Mean ln(dollars)	0.157	0.156	0.165
Test of differences			
USA vs. foreign	−0.004	[-0.090, 0.082]; p = 0.930	
USA vs. Japan	0.001	[-0.097, 0.098]; p = 0.987	
Japan vs. China	0.009	[-0.110, 0.091]; p = 0.853	
Panel D: Means of natural logs of grant dollars offered			
Mean ln(dollars)	1.953	2.270	2.088
Test of differences			
USA vs. foreign	0.219	[-0.768, 1.206]; p = 0.663	
USA vs. Japan	0.317	[-0.843, 1.477]; p = 0.590	
Japan vs. China	0.182	[-1.022, 1.386]; p = 0.765	

*p < 0.1; **p < 0.05; ***p < 0.01.
Table shows the outcome variables for the three treatments in Panel A–D. Each panel also shows the differences between the various treatments, 95% confidence intervals and p-values. Non-response is coded as 0 in Panel C and excluded in Panel D.

appointed executives, and the quarter of the year for the next election. Block randomization ensures that the blocking covariates are balanced across experimental conditions and provides a stronger methodological basis for simple difference-in-means analysis.[17] Also prior to fielding our experiment, we preregistered both our research design and analysis plan as a means of precommitting the study to experimental conditions and analysis strategies. This step is used to limit the temptation to "fish" for statistical associations after the fact.

RESULTS

For the purpose of this study we focus on three main indicators: (1) response rates to our inquiry, (2) cities' offering some form of financial incentive, and (3) the natural log of the mean dollars offered in grants per new job for this investment.[18] We supplement the analysis of the main outcomes with a description of the website traffic by treatment group (see Appendix A3.3, File S1).

In Table 1, Panel A, we document the response rate to our inquiries on behalf of our client. Our response rates range from 14.7 percent (Japan Treatment) to 15.8 percent (China) to 17.6 percent (U.S. treatment). These relatively small differences in response rates are significant at the 90 percent level for U.S. vs. foreign (Japan and China pooled) and U.S. vs. Japan. This provides modest evidence that cities tend to respond more to inquiries on behalf of U.S. firms over foreign firms in general and Japanese firms (though not Chinese firms) in particular. This rather weak evidence of treatment effects

[17]As a robustness check, blocking covariates should also be used in regression analysis to control for the stratification variables alongside the treatment conditions. In this experiment, regression analysis with the blocking covariates as controls produces substantively similar findings to those reported in the difference-in-means analysis.

[18]We phrased this per job since many incentives programs allocate dollars on a per-job basis. We included the total number of jobs (25) in the email proposal.

DOI: 10.1002/smj

may undermine the conventional wisdom that foreign firms are treated differently from domestic firms in the allocation of discretionary benefits. The liability of foreignness argument would expect that cities would respond much more frequently to U.S. firms, especially over Chinese companies. This was not the case in our study. In the case of response rates to U.S. vs. Chinese firms, our results suggest that any differences that might be detected with greater statistical power would be modest.

More telling still was the lack of differences in the number of respondents offering incentives to our client firm, as shown in Table 1, Panel B. The client firm in the U.S. condition received incentive offers from 9.84 percent of cities. The firm in the Japan condition received offers from a nearly identical 9.77 percent. The firm in the China condition received a slightly lower proportion of offers at 8.30 percent, but this difference was not significant statistically. The relatively tight 95 percent confidence intervals around the difference in treatment means suggests that any statistical differences undetected by this test are likely quite small substantively, especially in the U.S. vs. Japan comparison. The null hypothesis that there is zero difference between incentives for U.S. vs foreign firms cannot be rejected at any conventional level of confidence. Likewise, there is no statistically significant difference between incentive offers for the firm in the Japan vs. China conditions, providing no meaningful evidence for our second prediction. These results are robust to regression analysis with the blocking criteria employed as control variables. See Table A3.4.1 in File S1.

This pattern persists when we examine the size of the incentive offers measured in the natural log of grant dollars per job. We employed the natural log because the dollar values of the offers are highly skewed, but the results are substantively similar if unlogged dollar amounts are employed. As shown in Table 1, Panel C, the logged dollar values of offered grants are very closely matched across conditions when we code nonresponses as zero, again providing no significant evidence of discrimination against foreign firms. In Table 1, Panel D, where we dropped all nonresponses, shows a similar pattern.

Our empirical analysis thus far provides modest evidence for treatment effects on response rates but no significant evidence for treatment effects on actual incentive offers. We find no significant differences between the offers made to the U.S. firm compared to the putative Japanese or Chinese firm.[19, 20, 21]

In summary, our results provide limited evidence of discrimination against foreign firms in the allocation of incentives. City officials were just as likely to respond to emails when they believed they were interacting with Japanese and Chinese investors, though there was some modest evidence that U.S. firms were more likely to receive a response. However, this slight response-rate bias did not extend to actual incentives offers. Rather, the rate of offers to the U.S.-based client was statistically indistinguishable from rates for the Japan- and China-based clients.[22]

We suggest caution in interpreting our results, given that our field experiment was designed to test precisely just one mechanism underpinning the liability of foreignness. As a result, it is not straightforward to extrapolate this result to other theorized and practical barriers to operating abroad, including cultural distance, linguistic challenges, and other issues. For ethical reasons, we only made initial contact with the cities, and it is plausible that further negotiations between a given city and investors could have different potential outcomes based on the country of origin. However, our design was informed by discussions with economic development professionals to mimic closely this common interaction between firms and governments. In sum, we thus see our study as a complement to existing work on the liability of foreignness. Our study has the advantage of using the narrow lens of experimentation to compare how identical proposals are

[19] All of the results reported above are substantively similar when robustness is checked using regression analysis and covariates.

[20] As a supplemental analysis we provide documentation of activity on our website, from August 1, 2013 (the start of our experiment) to August 29, 2013. We installed tracking software on our website that allows for the tracing of location and individual IP addresses. More important for this project, in emails to the different groups, we provided a different routing address that all forwarded to our main website. Our tracking software thus allows us to examine the activities on our website by treatment group (U.S. vs. Japan vs. China). We document our results in Appendix A3.3 of File S1, but note here that similar to our findings above, we find very weak evidence for the liability of foreignness.

[21] We also tested city responses, incentive offers, and log of dollars using logit regressions in Appendix A3.4 of File S1. We include a model with our blocking covariates as control variables and a model that interacts our treatments with each blocking covariate. Our results remain unchanged.

[22] The largest contrast between the treatment of U.S. and foreign firms were found in the interest in our "U.S. Clients" versus "International Clients" tabs on our webpage.

treated prior to investment. But the clear limitation is that field experiment methodology does limit what we can claim about how cities would react to different types of investments or how foreign firms are treated relative to domestic firms after the investment has been made.

FIELD EXPERIMENT 2: CULTURE, IDENTITY AND PROSOCIAL BEHAVIOR

Our second field experiment can be categorized as a process field experiment and explores the canonical question of whether organizational culture influences firm performance. Prior strategy literature has argued and shown empirically that a key driver of firm performance is organizational culture (Barney, 1986; Sørensen, 2002), where strong organizational cultures are often described as having widely shared norms and values that motivate employees (Gibbons and Henderson, 2012; Goldberg, 2011; OReilly and Chatman, 1996; Sørensen, 2002). Using traditional methods, it is challenging to identify a causal effect, because, for example, strong performance could, in turn, drive widespread agreement on norms and values instead of the other way around. Resolving this empirical issue is not only difficult, but also has important lessons for practice.

In our parlance, a process field experiment aimed at this question would take the following approach. Such an experiment tests the process underlying why we think culture has an effect on performance. For example, prior work has found that individuals strive to be connected to one another, to organizations, or to ideas in an effort to construct identity (Ashforth and Mael, 1989; Chen *et al.*, 2014; Shih *et al.*, 1999). Strong cultures are built by articulating norms and values with which employees identify. Not only can stronger identity influence individual performance (e.g., Shih *et al.*, 1999), but it can also facilitate cooperation and has been posited to be a crucial factor in explaining firm performance (Kogut and Zander, 1996). Thus one potential process underlying how culture affects firm performance is through the influence of employee identity on the likelihood of cooperation. A process field experiment would aim to answer the broader question of how culture influences performance by manipulating employee identity and exploring the impact on cooperation on a particular task (see Cable, Gino, and Staats, 2013 for another field experiment investigating employee identity and behavior). This is the approach we will use in our experiment.

In addition to investigating the effect of firm identity, we will also analyze which identity (firm or business unit) is more important. A trade-off could exist in strengthening identities at different organizational levels. Identity at the lowest level of the organization could have the most significant impact on prosocial behavior and cooperation between members of that small group (e.g., Bohnet and Frey, 1999; Goette *et al.*, 2012). However, such an intervention might reduce cross-group cooperation, which could harm the firm as a whole. Competition between business units could potentially pronounce the possible negative effect of strengthening business-unit identity on cross-business-unit cooperation. We consider these various trade-offs explicitly in our design.

Experimental design

The field site for this experiment is a Fortune 500 medical device company with more than 45,000 employees dispersed across the world. Through a former student, we established a relationship with the company's "innovation" team.[23] One of the key responsibilities of this team is to manage the company's internal knowledge-sharing platform. This platform allows employees to participate in "Facebook-like" social networks across the firm and ask and answer questions of their fellow employees in a similar way to the popular site Yahoo! Answers.[24] Corporate management and the "innovation team" were very interested in increasing engagement on this platform to foster increased collaboration and strengthen corporate culture more generally. In our study we focus on 4,185 engineers at the firm across six business units. Our experiment involved manipulating employee identity using three different kinds of email prompts and observing subsequent cooperative behavior using data from the internal ideation platform.

We measure cooperation in the context of knowledge transfer. Like many companies, the firm uses their intranet platform to facilitate knowledge transfer across employees and to encourage novel ideas

[23]To protect the anonymity of the firm, we will sometimes use more general descriptions of business units, products, and corporate attributes.

[24]See answers.yahoo.com as an example.

by "breaking down silos" within and across business units and geographies. Employees can ask "the company" questions related to work. These questions range from requests to help solve a particular technical problem to queries about possible interactions between various materials to broader and more abstract scientific and technical concerns. The company sends out a bi-monthly email from the same email account to all employees with a list of featured questions selected by the innovation team after culling through a larger list provided by employees across the firm. All employees receive the same featured questions. We measure whether employees open this email, whether they click on a link in the email, and whether they answer any of the questions. This approach provides us with an objective measure of whether the employee engages in cooperative behavior within the firm. We can observe three mailings (1.5 months) before and three mailings (1.5 months) after our two intervention mailings (1 month).

We opted for a subtle intervention that only marginally manipulated the text of the emails the innovation team was sending out.[25] In doing so, we rely on standard research methods in social psychology to manipulate identity. In the email that went out to employees of our firm, we changed the text in ways that make particular identities salient. We compare this firm-identity treatment to a control in which the identity of the firm is not directly made salient.

To test experimentally the effect of identity on different organizational levels, we not only made the firm identity salient, but also implemented three treatments. In addition to having a control group ("T0: Control Group"): (1) making firm identity salient ("T1: Firm Identity"), (2) making business-unit (BU) identity salient ("T2: BU Identity"), and (3) making business-unit identity salient plus introducing competition between the business units ("T3: BU Competition"). Please see Appendix A4 of File S1 for the text of all the treatments, summary statistics and confirmation that our randomization (we stratified according to business unit) succeeded in creating groups with similar observable characteristics.

Comparing the different treatments allows us to test (1) whether making identity salient (at either the firm or the BU level) increases cooperation (comparing T0 to either T1, T2, or T3); (2) whether firm or BU identity is more powerful in increasing cooperation (comparing T1 to T2); (3) whether competition between BU increases cooperation even further and/or has detrimental effects (comparing T2 to T3).

We will interpret differences between the treatments as being driven by identity.[26] However, as with all priming experiments, it is not entirely clear whether the prime actually affects identity or not. Consider our comparison of T0 to T1/T2: In this instance, our intervention could either make the identity of the firm or the BU salient or, alternatively, highlight the culture of prosocial behavior in general, as T1/T2 contain prosocial language more explicitly than T0. To partially remedy this concern, we compare T1 and T2, where the only difference is the mention of the firm or the business unit and all of the other language remains constant.

In addition to testing whether the treatments have a main effect on cooperation, we also want to test whether heterogeneous treatment effects occur; that is, whether certain groups react differently to the treatment (see Appendix A2 of File S1 for a discussion about the pitfalls of conducting subgroup analyses). In our field experiment, we will test whether employees who are actively engaged with the firm prior to the experiment are affected by the treatments differently from employees who are not engaged. Increasing the engagement of employees who were not as active in providing public goods within the organization was a key objective of our internal team. Interestingly, this is an important consideration for future academic work on this topic since it could be that typical initiatives to strengthen corporate culture only impact a subset of employees.

We use prior activity on the internal platform as an objective measure of engagement by creating an

[25]It is important to think about how subtle a given treatment intervention will be. The trade-off in this decision is that if the treatment is not very subtle or even implausible, subjects in the experiment may become aware of the intervention, and their behavioral reactions might be influenced, comprising internal validity. One advantage of this specific subtle intervention was that it was extremely inexpensive to execute, which would positively influence any cost-benefit analysis.

[26]The different messages also differ in terms of length. While T1 and T2 have the same length, the control email is shorter and T3 is longer. It is possible that message length affects behavior independent of the content. If so, comparing T1 vs. T2 should help mitigate this concern.

DOI: 10.1002/smj

Table 2. Effect of identity treatments on opening of email

	(1) All	(2) All	(3) Low engagement	(4) High engagement	(5) Not HQ	(6) HQ
Firm identity × treatment periods	0.022 *	0.029**	0.043***	0.008	0.019	0.044**
	(0.013)	(0.012)	(0.015)	(0.018)	(0.015)	(0.019)
BU identity × treatment periods	0.011	0.026**	0.031**	0.018	0.026 *	0.027
	(0.013)	(0.012)	(0.015)	(0.018)	(0.015)	(0.019)
BU competition × treatment periods	−0.000	0.019	0.015	0.025	0.017	0.023
	(0.014)	(0.012)	(0.015)	(0.019)	(0.015)	(0.020)
Treatment periods (=1)	0.022 *	0.011	0.008	0.013	0.019 *	−0.003
	(0.012)	(0.008)	(0.010)	(0.013)	(0.010)	(0.014)
Time trend	−0.010***	−0.009***	−0.010***	−0.008***	−0.009***	−0.010***
	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Constant	0.717***	0.716***	0.709***	0.726***	0.732***	0.690***
	(0.006)	(0.004)	(0.006)	(0.006)	(0.005)	(0.007)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of mailings	1–5	1–8	1–8	1–8	1–8	1–8
R squared	0.000	0.006	0.008	0.005	0.007	0.006
No. of observations	20,925	33,480	19,888	13,592	20,480	13,000
No. of clusters	4,185	4,185	2,486	1,699	2,560	1,625

Significance level: *p < 0.1; **p < 0.05; ***p < 0.01
Coefficients of OLS regressions with individual fixed effects. Robust standard errors in parentheses clustered at the individual level. Dependent variable: 1 if opened email and 0 otherwise. Column (1) excludes post-intervention periods. Columns (3) and (4) split the sample at the median level of engagement index. Columns (5) and (6) show results for employees who work either in the Headquarters (HQ) or not.

"Engagement Index." The index takes values from 0, "not engaged," to 4, "most engaged."[27] A second possible method for thinking about engagement is the distance from headquarters (HQ). The medical device company has multiple business units spread out across the United States and the world, so we also tested for heterogeneous treatment effects on employees who work in HQ as opposed to other locations.

RESULTS

In principle, the analysis of the results of a field experiment is straightforward, because the randomization allows a simple comparison of means between the control group and the treatment groups.[28] However, individual data before and/or after the intervention period allows us to take into account slight differences between the treatment groups to estimate the treatment effects more precisely. Figure A4.3.1 in File S1 which presents the raw data (Figure A4.3.2 of File S1 for above or below median engaged) illustrates that it can be important to look at changes in behavior in the intervention period while taking pretreatment period differences into account.

We estimate panel regressions with a proxy for cooperation (opening of the email in Table 2 and clicking on a link or answering a question from a colleague in Appendix A4, File S1) as the dependent variable. For independent variables, the dummy "Treatment Periods (=1)" captures the effect on the two treatment periods for the control group. The dummies for the different treatments are 1 in the two periods in which we implemented our intervention and 0 otherwise; i.e., "T1/T2/T3 × Treatment Periods" shows the difference of the treatments compared to the control group for the treatment periods. As all the regressions include individual fixed-effects,

[27] For each possible activity on the platform, whether the employee (1) has created a user account on the internal platform, (2) ever posted a comment, (3) joined more than two employee groups, or (4) ever asked a question to the firm, we assigned 1 point. We then summed them up to create the Engagement index.

[28] This is only true if we look at the intent-to-treat effect; i.e. at behavior of all individuals who get assigned to control and treatment. If we condition on behavior—i.e. whether an individual opened the email—or look at subgroups, we are working with non-random samples and the empirical approach must be adjusted accordingly.

we estimate difference-in-differences. We also add a variable that captures any time trend.[29] All the regressions include individual fixed effects to capture time-invariant individual effects. We also cluster the standard errors at the individual level, to adjust the standard error for the fact that we have eight observations per individual. The regressions either include or exclude the posttreatment periods as it is unclear whether behavior after the treatment intervention is over (i.e., email text reverts to normal) should be seen as a continuation of the level before the intervention or was affected by the treatment intervention. In most of the models we include the posttreatment period and look at whether the interventions "break" the time trend.[30]

Table 2 reveals a number of results about the effect of our identity manipulation on opening of the email. The results mainly support the observations from the Figures in File S1. Whether we exclude the postintervention periods (column 1) or not (column 2), making firm identity salient increases the proportion of opening by between 2.2 percentage points ($p < 0.1$) and 2.9 percentage points ($p < 0.05$). Making BU identity salient has a small effect when excluding the postintervention period. The effect is about half as big as making firm identity salient. However, the difference between the two treatments is not statistically significant. When taking the postintervention pattern into account, both firm and BU identity increases opening 2.9 or 2.6 percentage points, respectively. Adding competition has no statistically significant effect on opening the email either way. None of the differences between the treatments are statistically significant at the 10-percent level (except for the difference between firm identity and BU competition in the model shown in column 3).

Comparing the reaction of employees who were engaged or not engaged on the platform shows that our effects almost entirely come from individuals who were not previously engaged at the firm. Firm identity increases opening by 4.3 percentage points for individuals less engaged prior to the intervention ($p < 0.01$). BU identity increases opening by less, but still by 3.1 percentage points ($p < 0.05$). By contrast, competition has no statistically significant effect on opening and is half as strong as BU identity alone based on the point estimate. However, the difference between individuals who exhibited low or high previous engagement is not statistically significant at any conventional level.

Comparing the effect of the intervention for employees working in the HQ or not also reveals interesting patterns. Judging from both the size of the effect and the significance level, the results suggest that for employees who don't work in the HQ, making BU identity salient is more powerful than appealing to firm identity. BU identity increases the proportion opening the email by 2.6 percentage points ($p < 0.1$). This effect is reversed for employees in the HQ: the firm identity intervention increases opening by 4.4 percentage points ($p < 0.05$) versus 2.7 percentage points (BU identity) ($p = 0.152$). Again, the coefficients are not different between the subgroups "Not HQ" and "HQ" at any conventional significance level.

In sum, we sought to test an important link in the chain between strong corporate culture and performance by manipulating employee identity and testing the impact on cooperative behavior. Our results show some effects of our intervention on the opening of emails containing questions from colleagues, the first step in cooperation. However, there is no equivalent effect of the intervention on whether an employee clicked on any of the links in the email (Table A4.3.1 in File S1) or whether they actually answered any of the questions from their colleagues (Table A4.3.2 in File S1). We did find interesting differences in cooperative behavior according to whether the employee was previously engaged and whether they worked at HQ, pointing to interesting opportunities for future analysis. These results are in the spirit of Kogut and Zander (1996), who discussed the role of identity inside organizations. For many firms that are highly acquisitive, integrating new employees and building collective identities are a significant challenge in the postacquisition process. Future process experiments might further inform the best strategies to navigate these issues.

[29] While there is no pre-treatment time trend differences between the treatments (analysis is available on request), there is a general decreasing time trend—even for the control group. Our discussions with corporate management indicate that this is likely attributable to seasonality coinciding with the end of the fiscal quarter, during which employees are often busier than usual.

[30] Further analyzing employee behavior in the post-treatment periods might reveal interesting insights. In Appendix A4.3 of File S1, we provide some analysis on post-treatment behavior and find differences between the different treatments. Exploring these differences more systematically is an interesting topic for future work.

DOI: 10.1002/smj

DISCUSSION

One of the most unique features of strategy research is the potential to make prescriptions that guide behavior in the field, often managerial actions with significant economic value at stake. These questions are invariably the most difficult to answer, in large part because the underlying causal models are not well understood. Although rigor and relevance are often posed as being at odds in our research, we think otherwise. Using field experiments is one way to provide rigorous answers to relevant questions, building on a long tradition of influential strategy research. In fact, given that so many strategy scholars work closely with firms already, our community has a particular advantage in formulating and executing useful field experiments.

The two experiments presented in this paper apply a new research methodology to canonical questions in strategy research and reveal useful empirical implications. More broadly, we use these studies to illustrate the wider applicability of this method to the field. In doing so, we identify both important drawbacks and advantages of using field experiments in strategy research. Neither kind of field experiment we present allows researchers to account for all of the impact of mediating and background variables. External validity cannot be guaranteed. However, field experiments are especially useful to establish internal validity—a challenge in much of the literature that links various firm characteristics to performance. Field experiments also may provide a high level of "naturalism" that has appealing properties for researchers. Conducting research in an actual business environment also ought to endow the results with greater relevance for managers, which may increase the chance that the findings have an impact on managerial decision-making.

As a result, we argue that using "strategy field experiments" and "process field experiments" in concert can yield valuable insights for strategy researchers and practice. In general, the opportunities to do strategy field experiments will be few, but the results are more likely to be provocative and significant for future research and practice. This approach will work best when the firm attribute under study operates through the perceptions of outside stakeholders: government officials, the media, or competitors. For their part, process field experiments rely on the development and refinement of strong theory to direct researchers to the relationships that matter. This attribute will drive continued interest in theoretical development alongside the iterative implementation of new field experiments. Especially as the field moves toward studying activities inside firms, the process field experiment approach has considerable merit, especially since it allows us to create new data rather relying on existing datasets. Taken together, we believe that the field experiment method, encompassing the two approaches discussed above, can be a key driver in the development of strategy research going forward.

ACKNOWLEDGMENTS

This paper is the result of a merger between two earlier working papers: "A Field Experiment on the Liability of Foreignness: Country of Origin and U.S. Inward Investment Promotion" (by Findley, Jensen, and Nielson) and "Field Experiments in Strategy Research" (by Chatterji and Meier). Each paper had an original field experiment: The experiment by Findley, Jensen, and Nielson is "Experiment 1" in this version and the experiment by Chatterji and Meier is "Experiment 2." We thank the editors and four referees for very helpful comments.

REFERENCES

Allcott H, Mullainathan S 2012. External validity and partner selection bias. Working Paper. NBER Working Paper 18373: Cambridge, MA.

Ashforth BE, Mael F. 1989. Social identity theory and the organization. *Academy of Management Review* **14**(1): 20–39.

Bandiera O, Barankay I, Rasul I. 2011. Field experiments with firms. *Journal of Economic Perspectives* **25**(3): 63–82.

Banerjee AV, Duflo E. 2009. The experimental approach to development economics. *Annual Review of Economics* **1**: 151–178.

Barney JB. 1986. Organizational culture: can it be a source of sustained competitive advantage? *Academy of Management Review* **11**(3): 656–665.

Bascle G. 2008. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization* **6**(3): 285–327.

Bloom N *et al.* 2009. The distinct effects of information technology and communication technology on firm organization. NBER Working paper 14975: Cambridge, MA.

Bohnet I, Frey BS. 1999. Social distance and other-regarding behavior in dictator games: comment. *American Economic Review* **89**(1): 335–339.

Bullock JG, Green DP, Ha SE. 2010. Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology* **98**(4): 550.

Cable D, Gino F, Staats B. 2013. Breaking them in or eliciting their best? Reframing socialization around newcomers' self-expression. *Administrative Science Quarterly* **58**(1): 1–36.

Calder BJ, Phillips LW, Tybout AM. 1982. The concept of external validity. *Journal of Consumer Research* **9**: 240–244.

Chen Y, Li SX, Liu TX, Shih M. 2014. *Games and Economic Behavior* **84**: 58–86.

Cook TD, Campbell DT. 1979. *Quasi-Experimentation: Design and Analysis for Field Setting*. Houghton Mifflin: Boston, MA.

Croson R *et al*. 2007. Using experiments in corporate strategy research. *European Management Review* **4**(3): 173–181.

Deaton A. 2010. Instruments, randomization, and learning about development. *Journal of Economic Literature* **48**(2): 424–455.

Dipboye RL. 1990. Laboratory vs field research in industrial and organizational psychology. *International Review of Industrial and Organizational Psychology* **5**: 1–34.

Duflo E, Glennerster R, Kremer M. 2006. Using randomization in development economics: a toolkit, NBER Working paper 333. Cambridge, MA.

French K, Poterba J. 1991. Investor diversification and international equity markets. *American Economic Review* **81**(2): 222–226.

Gibbons R, Henderson R. 2012. Relational contracts and organizational capabilities. *Organization Science* **23**(5): 1350–1364.

Goette L, Huffman D, Meier S, Sutter M. 2012. Competition between organizational groups: its impact on altruistic and antisocial motivations. *Management Science* **58**(5): 948–960.

Goldberg A. 2011. Mapping shared understandings using relational class analysis: the case of the cultural omnivore reexamined1. *American Journal of Sociology* **116**(5): 1397–1436.

Green DP, Ha SE, Bullock JG. 2010. Enough already about "black box" experiments: studying mediation is more difficult than most scholars suppose. *Annals of the American Academy of Political and Social Science* **628**(1): 200–208.

Hamilton BH, Nickerson JA. 2003. Correcting for endogeneity in strategic management research. *Strategic Organization* **1**(1): 51–78.

Harrison GW, List JA. 2004. Field experiments. *Journal of Economic Literature* **42**(4): 1009–1055.

Heckman J. 1992. Randomization and social policy evaluation. In *Evaluating Welfare and Training Programs*, Manski C, Garfinkel I (eds). Harvard University Press: Cambridge, MA; 201–230.

Helfat CE *et al.* 2009. *Dynamic Capabilities: Understanding Strategic Change in Organizations*. Blackwell Publishing: Oxford, UK.

Hutchinson JW,, Kamakura WA, Jr. Lynch JG. 2000. Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research. *Journal of Consumer Research* **27**(3): 324–344.

Hymer SH. 1976. *The International Operations of National Firms: A Study of Foreign Direct Investment*. MIT Press: Cambridge, MA.

Jensen NM, Lindstädt R. 2013. Globalization with whom: context-dependent foreign direct investment preference. Working Paper. Washington University in St. Louis: St. Louis, MO.

Karlan D, Zinman J. 2009. Observing unobservables: identifying information asymmetries with a consumer credit field experiment. *Econometrica* **77**(6): 1993–2008.

Kogut B, Zander U. 1996. What firms do? Coordination, identity, and learning. *Organization Science* **7**(5): 502–518.

Levitt SD, List JA. 2009. Field experiments in economics: the past, the present, and the future. *European Economic Review* **53**(1): 1–18.

Ludwig J, Kling JR, Mullainathan S. 2011. Mechanism experiments and policy evaluations. *Journal of Economic Perspectives* **25**(3): 17–38.

Lynch JG Jr. 1982. On the external validity of experiments in consumer research. *Journal of Consumer Research* **9**(3): 225–239.

Lynch JG. 1983. The role of external validity in theoretical research. *Journal of Consumer Research* **10**(1): 109–111.

Lynch JG. 1999. Theory and external validity. *Journal of the Academy of Marketing Science* **27**(3): 367–376.

Mezias J. 2002. Identifying liabilities of foreignness and strategies to minimize their effects: the case of labor lawsuits judgments in the United States. *Strategic Management Journal* **23**(2): 229–244.

Miller S, Eden L. 2006. Local density and foreign subsidiary performance. *Academy of Management Journal* **49**(2): 341–355.

Miller S, Parkhe A. 2002. Is there a liability of foreignness in global banking? An empirical test of banks' X-efficiency. *Strategic Management Journal* **23**(1): 55–75.

Nachum L. 2003. Liability of foreignness in global competition? Financial service affiliates in the city of London. *Strategic Management Journal* **24**(12): 1187–1208.

Oakley A, Strange V, Bonell C, Allen E, Judith Stephenson J, and RIPPLE Study Team 2006. Process evaluation in randomised controlled trials of complex interventions. *British Medical Journal* **332**(7538): 413–416.

OReilly CA, Chatman JA. 1996. Culture as social control: corporations, cults, and commitment. *Research in Organizational Behavior* **18**: 157–200.

Pew Foundation. 2007. Pew global attitudes survey. Available at http://tinyurl.com/6tp2oka. 13 December 2015.

Rajan RG, Wulf J. 2006. The flattening firm: evidence from panel data on the changing nature of corporate hierarchies. *Review of Economics and Statistics* **88**(4): 759–773.

Shadish WR, Cook TD. 2009. The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology* **60**: 607–629.

Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Design for Generalized Causal Inference*. Houghton-Mifflin: Boston, MA.

Shih M, Pittinsky T, Ambady N. 1999. Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychological Science* **10**(1): 80–83.
Siggelkow N. 2001. Change in the presence of fit: the rise, the fall, and the renaissance of Liz Claiborne. *Academy of Management Journal* **44**(4): 838–857.
Sørensen JB. 2002. The strength of corporate culture and the reliability of firm performance. *Administrative Science Quarterly* **47**(1): 70–91.
Spencer SJ, Zanna MP, Fong GT. 2005. Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology* **89**(6): 845.
Teece D, Pisano G, Shuen A. 1997. Dynamic capabilities and strategic management. *Strategic Management Journal* **18**(7): 509–533.
Zaheer S. 1995. Overcoming the liability of foreignness. *Academy of Management Journal* **38**: 341–363.
Zaheer S. 2002. The liability of foreignness, redux: a commentary. *Journal of International Management* **8**(3): 351–358.
Zaheer S, Mosakowski E. 1997. The dynamics of the liability of foreignness: a global study of survival in financial services. *Strategic Management Journal* **18**: 439–464.
Zhao X, Lynch JG, Chen Q. 2010. Reconsidering Baron and Kenny: myths and truths about mediation analysis. *Journal of Consumer Research* **37**(2): 197–206.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

File S1. Field experiments in strategy research.