From field studies to field experiments:
Studying organizational behaviors in actual organizations

Oliver P. Hauser[1,2], Elizabeth Linos[3], Todd Rogers[2]

Abstract. Organizational scholarship centers on understanding organizational context, usually captured through field studies, as well as determining causality, typically with laboratory experiments. We argue that field experiments can bridge these approaches, bringing causality to field research and developing organizational theory in novel ways. We present a taxonomy that proposes when to use an audit field experiment (AFE), procedural field experiment (PFE) or innovation field experiment (IFE) in organizational research and argue that field experiments are more feasible than ever before. With advances in technology, behavioral data has become more available and randomized changes are easier to implement, allowing field experiments to more easily create value—and impact—for scholars and organizations alike.

Keywords: field experiment; organizational context; causality; organizational theory; behavior

---

[1] Harvard Business School, Boston, MA 02163
[2] Harvard Kennedy School, Cambridge, MA 02138
[3] Goldman School of Public Policy, University of California, Berkeley, Berkeley, CA 94720

Introduction

> "The general objectives of the Academy shall be therefore to foster: […] (b) greater understanding by executive leadership of the requirements for a sound application of the scientific method to the solution of managerial problems." (Editor's preface to first issue of *Academy of Management Journal*, 1958, 1(1): 5–6)

Organizational scholars have always sought to bring scientific and practical methodologies to the study and practice of management. And in the last 50 years, organizational researchers have made great strides towards uncovering organizational behavior in practice using a set of established tools. On the one hand, scholars have conducted field research: qualitative scholars work in organizations, observing and describing first-hand real behavior in a real organization. Quantitative researchers have applied advanced statistical models to empirical datasets, often in combination with longitudinal employee surveys. Some of the most important organizational constructs has arisen from field research, such as psychological safety (Edmondson, 1999), social identity theory (Ashforth & Mael, 1989; Tajfel, 1982), trust and psychological contracts (Malhotra & Lumineau, 2011; Robinson, 1996; Robinson, Kraatz, & Rousseau, 1994; Rousseau, 1989), newcomer socialization (Bauer, Bodner, Erdogan, Truxillo, & Tucker, 2007), and individual-organizational fit (OReilly, Chatman, & Caldwell, 1991). One reason these methods have had such an impact in the field is that they put emphasis on the value of organizational context (Cappelli & Sherer, 1991; Johns, 2006; Mowday & Sutton, 1993).

On the other hand, other organizational researchers—in an often non-overlapping set—

have begun to test causality of organizationally relevant phenomena. For example, how does power affect propensity to participate and take risks in negotiations (C. Anderson & Galinsky, 2006; Magee, Galinsky, & Gruenfeld, 2007)? What are the conditions that cause unethical behavior to arise (Bazerman & Gino, 2012; Gino & Pierce, 2009; Gino, Krupka, & Weber, 2013), and what effect do honor codes and signatures have on ethicality (Kettle, Hernandez, Sanders, Hauser, & Ruda, 2017; Shu, Gino, & Bazerman, 2011; Shu, Mazar, Gino, Ariely, & Bazerman, 2012; Social and Behavioral Sciences Team, 2015)? To what extent do binding and non-binding contracts affect cooperative behavior and trust in groups (Hauser, Rand, Peysakhovich, & Nowak, 2014; Malhotra & Murnighan, 2002)? How do unstructured interviews influence interviewer perceptions of candidates (Dana, Dawes, & Peterson, 2013)? These researchers have largely relied on university laboratories or online platforms to design experiments with tight control over the decision environment where they can exogenously vary a variable of interest (Weick, 1967). Laboratory research places a premium on the causal nature of those findings.

Here we highlight a research method that bridges the gap between field context and causality: field experiments. For the purposes of this review, and in keeping with previous scholars (Eden, 2017; Harrison & List, 2004; List, 2011; Shadish, Cook, & Campbell, 2001), we broadly define field experiments as:

> ***Studies that induce a change in a randomly selected subset of individuals (or teams, or units) within their natural organizational context, and compare outcomes to a randomly selected group for which the change was not introduced.***

We argue that field experiments are an excellent tool for researchers keen on both context and causality – this is because field experiments are a method of *causal* inquiry within *real*

organizational contexts. This not only allows researchers to ground their work within actual managerial practice but also learn "what works" based on causal inference.

This paper does not introduce field experiments as a new method. Indeed, field experiments have been around for several decades, and the call for implementing them widely became part of a more general push for evaluative practices in the 1960s when pressure was mounting on public authorities to provide evidence for the outcomes of social programs (Suchman, 1968). Until that point, experimental randomization with the aim of establishing causality was seen as the purview of the natural sciences or, within the social sciences, to be conducted in the laboratory. Thus Donald Campbell, an early pioneer of the field experimental method in the administrative sciences, argued in his famous essay "Reforms as Experiments": "Experiments with randomization tend to be limited to the laboratory and agricultural experiment station. But this certainly need not be so. (…) We need to develop the political postures and ideologies that make randomization at [individual, unit and higher] levels possible." (Campbell, 1969, p. 425). The constraint, it seemed, was whether randomization for experimentation would be acceptable on a political or ideological front, rather than whether it was a useful methodology for answering questions on organizations.

Campbell's vision of an "experimenting society" (Campbell, 1991) was echoed further in organizational scholarship: Barry Staw coined the analogical term "experimenting organization" (Staw, 1977), and Gerarld Salancik proposed that qualitative research ought to engage in "field simulations" (Salancik, 1979). And while field experiments have subsequently been used to study organizational behavior and advance theory, many scholars (e.g., Scandura & Williams, 2000; Shadish & Cook, 2009) have lamented the fact that field experiments remain underutilized in organizational scholarship relative to other field research methods and relative to other scholarly

fields. This remained largely true in the field of organizational behavior in spite of the many excellent introductions and review articles that have been written over the years. (For interested readers, we recommend Campbell (1969) and Staw (1977) for an introduction as to why organizations should randomize; Eden (2017) and King, Hebl, Botsford Morgan, & Ahmad (2013) for a thorough review, especially surrounding sensitive topics in organizations; and Harrison and List (2004) for describing the continuum between laboratory and natural field experimentation. Relatedly, we recommend readers interested in the "mechanics" and "how to's" of carrying out their own first field experiment to the many helpful guides that exist on the topic, such as Boruch and Wothke (1985), Eden (2017), Gerber and Green (2012), Glennerster and Takavarasha (2013), Hauser and Luca (2015a), Ibanez and Staats (2016), King et al. (2013) and List (2011).) So, if field experiments are not a recent invention, why are they not more prevalent in organizational behavior?

In this paper, we focus on making the case for why field experiments matter specifically for *researchers in organizational behavior* and why they matter *now*. Field experiments have a reputation for being "hard to pull off" – but with changes in technology, availability of digital data, and a shifting culture of experimentation in organizations, we believe that they are now easier than ever to carry out and create value—and impact—for organizational scholars and practitioners alike.

To aid organizational scholars in conducting more field experiments, we also provide what we hope is a useful taxonomy of field experiments based on their function in organizational scholarship. We believe that the "method needs to fit the question", both across different types of research methods (e.g., ethnographic, survey, lab experimentation, field experimentation, etc.) as

well as within the same type of research method, like field experimentation. We hope that this paper will help readers become familiar with the type of field experiment that is best suited to their research questions, whether they are covertly testing if things work as they are believed to; making changes to an organizational process; or innovating within an organization.

Field and laboratory research

Organizational research methods vary widely. For the purposes of this short review, we distinguish between two broad classes of commonly used methods: those grounded in empirical data in the external world; and those aimed at establishing causality within a controlled laboratory or online framework. (For a deeper treatment of field-based versus manipulation-based methods, see Chatman and Flynn (2005)'s full-cycle research model.)

Field-based researchers have worked with and in organizations to understand how incentives, leadership style, and organizational constraints affect the behavior of employees and managers in organizations. Their work may be primarily qualitative (Denzin & Lincoln, 1994; Patton, 1990) using inductive and descriptive methods, among others, or primarily quantitative, using existing or newly created datasets to answer similar questions (Neuman, 1991; Suen & Ary, 1989). Broadly, these methods share a common "groundedness" in real organizational context (Feldman & Orlikowski, 2011; Staw, 2016): they work with real decision-makers in real organizations to understand real organizationally relevant outcomes. These methods are therefore compelling for a number of reasons, not least of them because of their embeddedness within, and relevance to, organizational context (Cappelli & Sherer, 1991; Johns, 2006; Mowday & Sutton, 1993).

Field research methods are also important for theory development in organizational scholarship (Eisenhardt, 1989; Feldman & Orlikowski, 2011). Real-world organizational context generates the framework and boundary conditions for all theories of organizational behavior. Theory development in organizational scholarship requires a deep understanding of context, because it requires an "effort to generate new theory about that context" (Staw, 2016, p. 11). Staw argues that contextualism "involves greater appreciation for the phenomenon under study" and as such he urges researchers "to understand the environment in which it takes place" (Staw, 2016, p. 10). For example, researchers would benefit from immersing themselves into the organization, using research methods that not only take into account the organizational context but also understand the central role the organization plays in affecting the behavior of interest (Staw, 2016). More generally, if the goal of theory is to explain and predict behavior in organizational settings, the interplay between field-based data and theory is inevitable (Glaser & Strauss, 1967). Put simply, field-based organizational scholars rely on, work with, and truly understand what the real-world data in organizations—be it qualitative or quantitative—tells them in testing theories and developing new ones.

While field-based organizational methods are often deeply insightful in the ongoing processes of individuals, units and the organization as whole, they typically cannot rigorously identify causal processes. Rather than trying to map out the complexity of factors that shape behavior in organizations, causal methods try to better capture how pulling any one lever impacts others, all else constant. The most common method used in organizational research for establishing causality is laboratory experimentation (Thau, Pitesa, & Pillutla, 2014), in which participants are brought into a physical laboratory or invited to an online platform where they take part in a

randomized experiment. We refer to "laboratory experiments" as any experimental setup in which the experimenter retains tight and exogenous control over the independent variable and in which the study participants willingly and knowingly take part in the experiment outside of their natural (organizational) environment. By this definition, we include experiments run in the physical laboratory on university campuses and on online platforms (e.g. Amazon Mechanical Turk; see Buhrmester, Kwang, and Gosling (2011), Paolacci and Chandler (2014), Paolacci, Chandler and Ipeirotis (2010)), as well as most "lab in the field" paradigms[1] (U. Gneezy & Imas, 2017; Harrison & List, 2004).

Laboratory experiments are particularly noteworthy for their tight control over confounding variables and the decision environment participants are in. Importantly, they bring credibility to causal claims using randomization. Randomly assigning individuals to groups ensures that, in expectation[2], the two (or more) groups are comparable in observable (e.g., gender, age, race, etc.) and unobservable (e.g., motivation, biases, beliefs, etc.) characteristics, such that the only difference between them is the researcher-induced change.

In contrast to field-based research, laboratory methods often lack real-world organizational

---

[1] The lab-in-the-field paradigm is a combination of laboratory and field experimentation involving elements from both methods: a lab-in-the-field experiment is conducted in a naturalistic environment (i.e., not a university laboratory) drawing study participants from a theoretically relevant population, but using validated laboratory paradigms to study the behavior of interest. For a recent review of the advantages and drawbacks of this method, we recommend (U. Gneezy & Imas, 2017).

[2] We note that randomization does not assure balance across observable and unobservable covariates (or control variables), but only does so in expectation (Deaton & Cartwright, 2016). We therefore advise researchers to always check whether randomization did actually lead to balance on theoretically important covariates. We refer interested readers to Glennerster and Takavarasha (2013) and Gerber and Green (2012) for a detailed discussion of how to deal with imbalance after randomization.

context. Some contextual variables simply cannot be reproduced in a laboratory environment, such as number of dependents and breadwinners in the household, complex interpersonal relationships, or subtle changes in the organizational climate conveyed through "tones at the top" (Brief & Smith Crowe, 2015). When the researchers believe that they *can* be reproduced, the independent and dependent measures used in a laboratory study must then be able to face up to the scrutiny of external (real-world) validity: have the researchers correctly identified the most relevant variable for the context they want to model? Does the operationalization of this variable tap into the same psychological mechanism that the researchers are interested in? Are the boundary conditions sufficiently understood? And can inferences be reasonably drawn from the study population (which, by definition, is not the same as the one in a work context) to justify the conclusions? Perhaps unsurprisingly, laboratory experiments published in the organizational literature have been criticized on the grounds that insights or "solutions" found in contrived experimental setups using college students as decision-makers can hardly be applied to the real world and real managers (Baumeister, Vohs, & Funder, 2007; Galizzi & Navarro-Martinez, 2015; Staw, 2010).

However, despite the challenges of contextuality and operationalization, laboratory experiments *do* enable causal inferences because they can test the impact of a single variable of interest on an outcome measure, a task that, in contrast, is challenging to do in empirical research where confounds with other variables exist. In short, when done correctly, laboratory experiments appeal to researchers committed to causality to drive theory and find out "what works."

What are field experiments, and why now?

We argue that field experiments bridge the gap between field research and causal-driven

laboratory studies by bringing *causality to field research*. Like other field research, they can capture behaviors and motivations that are almost impossible to capture in a lab (e.g., long-held beliefs about one's role on a team, complicated and multi-faceted relationships with bosses or coworkers, the high stakes of actually being hired or fired; see Ibanez and Staats (2016) for an in-depth discussion), but can also parse out these complexities to understand how changing one important variable affects others. In doing so, field experiments are not just "a more complicated lab experiment," nor are they yet another complex field method, but are able to answer a new set of theoretically-driven questions.

First, field experiments go beyond the possibilities of laboratory experiments because they are conducted within a natural organizational context. Several advantages derive from this setup: field experiments are typically developed from the ground up, often directly in collaboration with the organization in which the field experiment is later conducted. This enables the researcher to dive into and fully understand the context in which employees make decisions, and these are typically the same or similar employees on whom the experiment is later run.[3] Thus, unlike laboratory experiments who rely on a student population (or online subjects in the case of Amazon Mechanical Turk), field experiments are conducted on the theoretically relevant population of decision-makers. Finally, when an experiment has been concluded, the learnings can be applied first within the organization where they can be taken to scale. But importantly, organizational

---

[3] Care must be taken, of course, that employees are not aware of an imminent experiment being run, or else they might respond to the mere observation by researchers – a problem known as the "Hawthrone" effect (Parsons, 1974). Ways to mitigate these problems is to interview, discuss and work with a (smaller) group of employees to understand the context, but conduct the main experiment (based on the insights gained from the qualitative work) with the (larger) set of employees in the same organization.

scholars are often able to extrapolate their newly gained knowledge to outside the particular organization, assuming the field experiment was designed in a way to allow for more generalizable conclusions (which we will discuss in more detail below).

Second, field experiments also provide additional benefits over other field research methods which, by definition, are contextualized in the organization. We focus on one core advantage of field experiments: establishing causal inference. Field experiments are considered the "gold standard" of evaluative research practices (Shadish et al., 2001) because of the unconfounded variation of independent variables which allows scholars to determine the causal effect that this variation has on a dependent variable of interest. Most other field research methods must establish causality indirectly. For example, a quasi-experiment (Cook & Campbell, 1979) differs from a field experiment in the way participants (e.g., employees, teams, or units) are assigned to treatment: while participants in a field experiment are assigned randomly by the experimenter or the organization, participants in a quasi-experiment are, by definition, assigned non-randomly, usually by the organization. This implies that the researcher cannot be sure that control and treatment groups are not systematically different from another, a concern that is greatly reduced in a truly randomized field experiment. Other quantitative field methods that attempt to estimate a causal effect include regression discontinuity designs, propensity score matching, and difference-in-difference studies; however, the degree to which causality can be confidently inferred depends on the details of the dataset available (Campbell, 1969; Cook & Campbell, 1979; Shadish et al., 2001).

All this has, of course, always been true about field experiments, and yet organizational scholars have in the past not adopted field experiments as one of their standard tools. So why do

we think our call for more field experiments will be any different now?

The reason we argue in favor of field experiments today is that, now, field experiments have become much easier, cheaper and more feasible to conduct than in the past. First, with the advent of more technology and digitalization of so many aspects of work and life, the availability of theoretically-relevant data is becoming much less of a constraint. Most companies now collect and store more data than they can possibly work with. And even when they do want to analyze the most valuable data, they often do not know how to go about it. In fact, the few notable exceptions that do make great use of their data are also performing at the top of their industries, suggesting that there is much value to be captured. For example, data-driven companies such as Google, Facebook and Amazon, to name but a few, routinely hire social scientists who analyze, predict and recommend organizational, strategic, and human resource decisions based on the data they collect on their employees. In some of these companies, new employee hiring practices and training schemes are experimentally trialed before roll-out; performance evaluations are both quantitative and qualitative, and correlated with later success on the job; and product development as well as interactions with customers are often refined and rolled out through "staged innovations" (Campbell, 1969). However, the share of companies that use data and field experiments to this extent is still relatively small, which leaves much room for improvement and requires most companies to still learn how to make effective use of data and experimentation.

The lack of an organization's clear and strategic use of data suggests a value-creating opportunity for organizational scholars: researchers can bring (*i*) expertise and theoretical understanding of the organizational behavior in question and (*ii*) methodological and statistical expertise to test those theoretically driven questions, which feed back into the organizations and

help them solve their core concerns. In our experience, many organizations would gladly share important data in exchange for a better understanding what they mean. To be sure, in many cases this means that researchers have to go beyond their original research questions so as to be helpful to their partner organizations. While this requires balance and prioritization, we see this as more than just a "necessary evil." Rather, when researchers help organizational leaders grapple with the managerial questions with which they are struggling, researchers often discover interesting and potentially theoretically important research directions for the future. This highlights the "field research" component of field experiments: through exploration, discussions, and openness in working with partner organizations, researchers deepen their understanding of the context and can discover new insights that are hard to spot from within their offices. Indeed, observing and studying organizational struggles can provide a valuable learning experience for scholars—which has been dubbed "phenomenon-driven research" (Staw, 2016)—and research that makes contributions to such problems meets one of the fundamental aims of organizational scholarship: to provide solutions to the managerial challenges of today using scientific methods.

Second, companies are becoming more open to partnering and sharing data with organizational researchers. To foster these productive partnerships, researchers will often need to assure their partner organizations of the highest data-storage and data-security standards they can afford – which, in most cases, is guaranteed through the institutional review process (i.e., IRB) that researchers have to undergo before working with human subjects data. In fact, we have found that these academic standards are often higher than what the organizations themselves would have expected – a fact that tends to build trust. Moreover, technology again provides a series of tools (e.g., de-identifying data remotely, or easy-to-use randomization and measurement tools such as

Google Analytics) to study outcomes without ever seeing sensitive, personally identifiable information. As such, while data privacy is often raised as an obstacle to conducting field experiments, in our experience it is rarely the underlying reason why a field experiment does not move forward.

More broadly, the rise of A/B experimentation—an alternative term for "randomized controlled trial", often used by tech startups in Silicon Valley to refer to experimentation on their websites or in emails—in the technology sector, as well as continuing calls for innovation, experimentation and evidence-based management within organizations by popular practitioners' outlets such as *Harvard Business Review* (e.g., Beshears & Gino, 2015; Kohavi & Thomke, 2017; Luca & Hauser, 2016; Pfeffer & Sutton, 2006) suggests that many organizations now view their ability to harness their data as central to their business strategies. The "political postures and ideologies," that Donald Campbell lamented in 1969 needed to be reformed, have in many ways undergone the very reform he called for: while organizational scholars had long advocated that data—and, crucially, understanding it—is one of the most valuable resources a company has, business leaders from Silicon Valley to civil servants in the British and Australian governments are now beginning to understand and explore the opportunities of big (and small) data (George, Haas, & Pentland, 2014; G.-H. Kim, Trimi, & Chung, 2014; McAfee & Brynjolfsson, 2012). Researchers with the toolkit to work with data, and the theory to understand it, can help organizations of all sizes capitalize on those insights.

Field experiments for organizational research

Having an organization that is willing to work with scholars is a valuable step towards

implementing a field experiment. However, collaboration and buy-in from an organization is not always necessary to run a field experiment (Salancik, 1979). This is because not all field experiments are the same, and choosing the right type of field experiment is essential in studying organizational behavior.

A field experiment can be conducted in many different ways. To help organize the myriad of ways in which field experiments differ, we propose what we hope will be a useful taxonomy: we suggest that organizational field experiments can be categorized as **Audit Field Experiments (AFE)**, **Procedural Field Experiments (PFE)**, and **Innovation Field Experiments (IFE)**. They can be distinguished from one another by the variation of three independent dimensions: input variation into an existing process, parameter variation of an existing process, and innovation of a new process, respectively. **Table 1** provides an overview of our proposed taxonomy.[4]

We broadly define "process" as the organizational structure in place for certain purposes in the firm (e.g., hiring, onboarding, leadership trainings, employee communication, etc.). One reason to run field experiments is to understand whether an existing process produces the desired outcomes consistently. That is, when relevant or often seemingly irrelevant "inputs" are changed (e.g., the name of a customer, the gender of a manager, the status of the messenger), does the output remain the same or is it different? Audit Field Experiments (AFE) systematically vary inputs into a process and measure whether these changes cause unexpected or unintended variability in output.[5] In many cases, the goal of the AFE is to understand if the process itself

---

[4] In our definition, natural experiments where the independent variable is fully randomized due to an (unintended) design feature in the company—which has also been referred to as an "incidental experiment" (Hauser & Luca, 2015b)—are a subset of PFEs.

[5] An AFE can be viewed as form of evaluation of the process (Suchman, 1968): instead of using an evaluation technique that looks, for example, at historic data, an AFE evaluates the

works as intended, while keeping the existing process untouched.

In contrast, other field experiments keep inputs constant but vary the process itself. When an existing component of a process is varied, we refer to this as a Procedural Field Experiment (PFE). Furthermore, we call Innovation Field Experiment (IFE) the introduction of an entirely new process, going beyond modifying aspects of an existing one.

In the following sections, we discuss each type of field experiment separately; we outline its defining characteristics and highlight a few exemplary field experiments in recently published papers in leading organizational journals.

--- Table 1 about here ---

Audit Field Experiment (AFE)

We begin by looking at Audit Field Experiments (AFE) which aim to probe an existing process to find out whether it works as intended. To do so, scholars vary the input into a process systematically, such that they can compare the outcomes they obtain with the ones they (or the company) is hoping and expecting to observe, while holding everything else (including the existing process) constant.

A classic use of an AFE is in the domain of discrimination (L. Anderson, Fryer, & Holt, 2006), or in studies of any sensitive behaviors which would suffer from selective reporting bias if employees were asked directly (King et al., 2013). Bertrand and Mullainathan (2004) are among

---

program of interest "prospectively" by systematically looking at the causal effects of input variation.

the most well-known (though not the first ones; for example, see Bendick, Jackson, & Reinoso, 1994) to use a large-scale AFE to test whether racial discrimination exists in the U.S. labor market. They did so by introducing exogenous and systematic variation of race and gender in (fictitious) applicants' resumes sent to real jobs, holding constant all else but the race and gender of the applicant. If there were no discrimination, the same call-back rates for white and black applicants would be expected. However, they found that applicants with an African-American sounding name received significantly fewer callbacks than white applicants. Similar findings have been reported for other ethnic minority groups, suggesting that white applicants are substantially more likely to receive a response, callback or job offer than members of a minority group (Booth, Leigh, & Varganova, 2011; Milkman, Akinola, & Chugh, 2012; Pager, Western, & Bonikowski, 2009).

Given this discriminatory behavior on the side of the firm, how do minority applicants cope? In a recent paper in *Administrative Science Quarterly*, Kang, DeCelles, Tilcsik and Jun (2016) explore a strategy referred to as "resume whitening" among minority applicants – the idea that applicants who fear that they will get discriminated against based on their race reduce or remove any signals about their race on their CV. Indeed, using qualitative and lab-based studies, Kang et al. demonstrate that minority applicants engage in a number of "whitening" strategies, but they do so less when the company presents itself as valuing diversity. To explore whether the whitening strategy works and to find out whether minority applicants are correct in assuming that they need to engage in less resume-whitening when applying to pro-diversity companies, Kang et al. turned to running a large-scale AFE. The authors sent 1,600 resumes of a fictitious job applicant to companies advertising in 16 metropolitan areas in the U.S. By varying the degree to which the fictional job applicant whitened his or her resume, they could explore the causal consequences of

whitening in an actual labor market context. They find that employers are indeed less likely to respond to non-whitened resumes. Importantly, however, callback rates were just as low for non-whitened minority applicants from companies that included a pro-diversity statement in the job advertisement, suggesting that minority applicants cannot expect lower levels of whitening to be successful among companies that position themselves as pro-diversity.

AFEs are not, of course, limited to discrimination-based research. The same approach can be used to study other variations in input that are not directly linked to known discrimination (e.g., varying the messenger on an email between a supervisor and a peer may change how an employee responds; varying the company name of a potential partner organization between well-known or anonymous may lead to differently quoted prices).

Another way an AFE can vary the input into a process, is to change how an input is framed. For example, Desai and Kouchaki (2015) hypothesized that framing a price request in terms of "units" (versus just an overall cost estimate) increases accountability to report more honestly, ultimately leading to lower (and more accurate) quoted prices. To test this hypothesis, research assistants (who were blind to the hypotheses) called up car garages to ask for an estimate for a repair on their car: when the request was framed by asking first about the costs for labor and parts and second about a good-faith estimate, the quoted price was significantly lower compared to a control condition, in which the request order was reversed – first asking a good-faith estimate, then for parts and labor.

In all such cases, an AFE is distinguished by the fact that inputs, not the process itself, are varied to measure whether an existing process is consistently operating as expected. In short, AFEs are a powerful tool to study existing organizational processes—without changing the parameters

of the underlying process. Notably, AFEs therefore do not necessarily require buy-in from an organization for a researcher to carry out this research. That said, organizations might be well-advised, and possibly self-motivated, in carrying out such experiments to test the efficacy of a wide range of existing organizational processes. If the AFE finds unexpected inefficiencies or biases, scholars and organizations can go about fixing them. That may be the point where researchers and organizations alike would turn to a Procedural Field Experiment or Innovation Field Experiment.

Procedural Field Experiment (PFE)

Sometimes scholars or organizations want to make changes to an existing component of a process to find out whether altering it can significantly impact an expected outcome. Processes are defined by a wide range of parameters. Here, we use "parameters" to capture all structural elements of the process—such as the time, place, incentives, people, etc.—involved. When existing components of a process are changed, we refer to these experiments as "Procedural Field Experiments."

Returning to the hiring example above, a company might be aware that they are not hiring enough female or minority applicants. (They might have carried out their own AFE to figure this out systematically and scientifically.) Let's assume that the scholar believes that part of the reason that women and minority applicants do not get invited to interview at the same rate is because of (explicit or subconscious) biases of the hirers. The scholar aims to test whether blinding CVs has an effect on interview rates by gender. Blinding CVs (i.e., not including the gender, name or photo of the applicant) is an excellent example of changing a small but important parameter. Because of

our reliance on surrounding cues (e.g., particularly visual, see Tsay (2013)), blinding strategies such as blind auditions have been shown to significantly reduce gender bias in hiring and identifying the best applicants (Bohnet, 2016; Goldin & Rouse, 2000). We would refer to this type of experiment as a "Procedural Field Experiment."

In another PFE, Grant (2012) tested a variation of the standard procedure to motivate employees and increase productivity. Building on a past work on transformational leadership (Dvir, Eden, Avolio, & Shamir, 2002) and beneficiary contact (Grant et al., 2007), Grant theorized that employees' performance improves through transformational leadership because it motivates employees to transcend their self-interest, but that it may be most effective in interaction with beneficiary contact. That is, he predicted that performance would improve when employees interact with beneficiaries of their work, which heightens the sense that they are making a meaningful impact on other people's lives. To causally change perceptions of his independent variables (transformational leadership and social impact), Grant ran a PFE using four randomly assigned conditions: at the training sessions for newcomers, he randomly assigned a manager of the firm (the "transformational leadership" condition), a customer who benefited from the employees' work (the "beneficiary contact" condition), both a manager and a customer (interaction condition), or neither (control condition) to give a speech to the newcomers. He found that the interaction of transformational leadership and beneficiary contact causally led to the theorized boost in performance.

PFEs can be conducted to change and improve any organizational process. For example, Cable, Gino and Staats (2013) modified an existing socialization process during the training phase of newcomer employees in a call center. Drawing on socialization theory, the authors compared

competing hypotheses that derived from an original theoretical framework proposed by Van Maanen and Schein (1979), in which the authors placed different socialization strategies on a continuum from "individualized socialization" to "institutionalized socialization." The latter had become the de facto accepted socialization strategy, frequently adopted by organizations and studied by scholars. However, Cable, Gino and Staats proposed that individualized socialization, highlighting employees' individual and unique strengths that they can contribute to the organization, would bring out under-theorized benefits to the organization. By assigning newcomer staff to one of three randomly assigned conditions—control, individualized socialization, and institutionalized socialization—the authors demonstrated that turnover rates are lowest and customer satisfaction highest in the individualized socialization treatment. Importantly, the authors were able to causally attribute these changes in outcomes not just to a generally accepted form of socialization (institutionalized), but to the specific role of *individualized* socialization.

Innovation Field Experiment (IFE)

While organizational innovation is still seen as the product of the lone genius who is talented rather than hard-working (Montuori & Purser, 1995; Tsay, 2016; Tsay & Banaji, 2011), it actually requires a systematic and robust experimental methodology (Camuffo, Cordova, & Gambardella, 2017). We propose that field experiments that go beyond changing components of existing processes and instead bring a new process to an organizational problem can be described as "Innovation Field Experiments" (IFE). Here we broadly speak of innovation as any new process that an organization introduces, in order to solve a problem that an existing process is not

sufficiently addressing.

With a view towards discrimination in hiring mentioned above, an IFE might introduce a completely new method to removing bias. For example, instead of altering an existing evaluation process (e.g., by removing gender and race from CVs), an IFE might test the use of advanced algorithms to identify suitable candidates before they even apply and inviting them to apply at equal rates across gender and race. IFEs aim to bring a new process into the organization and test it systematically before it is rolled out – as such, they are often the analytically robust analogue to what organizations may call a "pilot project."

Buell, Kim and Tsay (2017) ran an IFE to test whether increasing transparency and reciprocity can increase productivity. Building on past work demonstrating that transparency can promote perceptions of reciprocity (Buell & Norton, 2011; Hauser, Kraft-Todd, Rand, Nowak, & Norton, 2016; Hauser, Hendriks, Rand, & Nowak, 2016), Buell and colleagues theorized that introducing customer transparency in an organizational context can enable performance reciprocity. In a restaurant and service context, they randomly assigned chefs to a transparency condition, in which they could view customers picking up the food they prepared for them through a video live stream on an iPad in the kitchen, or a control condition, in which chefs could not observe customers reactions. The authors found that customer transparency had a large impact on throughput times and affect the chefs' perceived impact of their work on others, consistent with earlier theorizations of job design, task identification and feedback (e.g., see Hackman & Lawler, 1971).

In another example of an IFE, A. Gneezy, Gneezy, Nelson and Brown (2010) propose that organizations ought to shift from corporate social responsibility (CSR) to shared social

responsibility (SSR). To appeal to customers and potential employees, organizations typically aim to demonstrate their commitment to socially responsible practices (Burbano, 2016; McWilliams, n.d.). Yet, Gneezy and colleagues hypothesized that part of this responsibility could be shifted to customers who also care about social causes. Through a novel price strategy, Gneezy et al. randomly assigned customers in their field experiment to one of four conditions which varied across two dimensions: the variation along the first dimension is that customers either pay the regular price for a product or can choose how much to pay for it (including $0, known as "pay-what-you-want" pricing, see A. Gneezy, Gneezy, & Riener, (2012)). Meanwhile, along the second dimension, the revenue either all goes to the organization or 50% of the price paid is donated to charity. In line with the authors' SSR predictions, customers paid significantly more for the product in the condition where customers choose their own price and where 50% of profits are donated to charity, leading to the highest firm profits despite the donations deducted for charity. Using an IFE, the authors were able to demonstrate the causal effect of this innovative pricing strategy in a real organizational context.

Criticisms and limitations

Practical concerns

Field experiments have a reputation for being "hard to pull off." We believe this is only partly true. Field experiments usually take a long time for a number of reasons, such as finding and convincing an organization to carry out the experiment. While we argue that organizations are becoming increasingly more comfortable with the idea of running experiments, the most successful experiments are ones where the timeline of the organization matches that of the research team.

This may mean that while an organization is broadly interested in the idea, it may take them months for a given change in their process to become salient enough for a new experiment. The time lag between initial conversations and launch is highly variable and unpredictable. On top, field experiments also involve meticulous planning, which often depends on getting to know the field site, working with existing data, and adjusting the experiment accordingly. Specifically, once a field experiment is designed, changing elements of the intervention can significantly impact the validity of results and so this initial planning stage is particularly important. Understanding how and when the organization itself collects data can play an important role in determining how outcomes will be understood. Indeed, the way a technology (e.g., in this case, data collection or delivery of interventions) is initially implemented has significant effects on its likely adoption and impact (Bergman & Rogers, 2017). This too—depending on the relative experience of the data collection team and the previous uses of existing data—can be an arduous process.

Field experiments in organizations can also be difficult to execute because organizations are complex entities. First, seeking to understand the concerns of other stakeholders (e.g., executives and site-level managers, ethics boards, etc.) beyond those that are directly involved in the field experiment not only determines whether a project will launch, but importantly, whether a second or third project can be launched with the same company. Often initial buy-in from other stakeholders also determines whether the promise of a successful field experiment (e.g., the ability to scale-up an impactful intervention) can in fact be accomplished. Getting pre-approval or a general agreement that a successful "pilot" phase should lead to scale up is, at its core, a process of stakeholder management. Second, interventions and outcome measures are complex too; often both are longitudinal in nature, increasing exposure to unexpected complications. While careful

design and planning can mitigate execution problems to some extent, scholars should remember that errors can happen. If they do, scholars are reminded that even experiments that encounter problems or were changed unexpectedly can still lead to valuable insights (e.g., see Eden (2017) for a discussion on "quasification"). Conversely, the complexity of running field experiments should also provide sensible boundary conditions to what reviewers at journals should ask for: while all scholars agree that an impeccable, smoothly-run experiment is preferable to one that encountered real-world issues along the way, reviewers should consider whether it is feasible to ask the authors for a full repetition of a large-scale field experiment involving several hundreds or thousands of additional real-world employees and potentially a new field site altogether. That said, reviewers might want to see replication in other ways, such as complementary analysis from other data sources or different methodological angles (Chatman & Flynn, 2005; Jick, 1979). (For an in-depth discussion of field experimental challenges—with, in particular, recommendations on how reviewers might evaluate them—see Ibanez and Staats, 2016.)

It should be noted that all these concerns exist for other methods used in organizational research as well. Empirical methods, such as qualitative interviews, face similar challenges in terms of stakeholder buy-in and navigating the complex decision and hierarchical environments of organizations (Ibanez & Staats, 2016). The complexity of obtaining and working with existing data is a well-known challenge for quantitative field researchers – but it is both a challenge and an opportunity: as we have argued above, getting into an organization and helping them figure out, address and ideally solve their problems is a value-creating opportunity for both the organization and the organizational scholar. Likewise, carefully conducted laboratory studies can also be complicated in their execution, as one tries to simulate a precise decision context that arises in the

external world, as well as the very context around which a decision is made. Given that, by definition, a laboratory context must be constructed by the experimenter, careful planning and an in-depth understanding of context is just as much a part of lab experimentation as it is part of field experimentation.

Therefore, we suggest that while some field experiments can be particularly costly to execute, they are not necessarily more difficult to execute than other methods of research. However, even initial start-up costs are very well worth the effort, given that a fuller and more comprehensive view of organizational life is obtained. Finally, though it is a truism, quality research takes time and can be difficult; this is true of conventional organizational research methods as well as field experimental methods.

Ethical considerations

As with all research activity involving human subjects, a field experiment is subject to the rules and regulations of institutional review boards (IRB) and must comply with the university's ethical standards. However, scholars should take extra care: field experiments, by definition, affect real people's lives in real work situations. An employee's chances for a promotion might be improved, but also reduced, by a field experiment intervention to innovate with performance evaluations; an employee's take-home pay, and their family's household income, might rise, or suffer, from the compensation structure that may be varied as part of an experiment; or an employee's relationship with their boss might be improved, or strained, by novel design choices in feedback and review meetings. Designing and running a field experiment is therefore a delicate task that can have as much positive as well as negative impact for the people who are part of the

experiment.

Some particular ethical considerations that are different from other methods are worth noting. First, in many field experiments people don't sign up to be part of the experiment. In some cases, an email from a manager is the only notification they receive about research taking place and their decision not to participate requires opt-out; other times, they are not informed at all of the ongoing research activity. Understanding the risks (and benefits) of not informing the participants in your field experiment is an important consideration upfront. That said, the often subtle nature of field experiments is also one of its strengths: by varying elements of the actual work environment without informing employees about those changes in advance, organizational scholars learn about the effects of the change without needing to worry about demand effects or experimenter effects, which helps increase external validity of the results obtained.

Second, field experiments can have spill-over effects into behaviors and on outcomes not explicitly considered in the planning of the experiment (Dolan & Galizzi, 2015). Spill-overs can occur both for the individual or unit that was treated in an experiment as well as between individuals or units. To reduce unintended side effects, scholars are advised to carefully consider all potential outcomes of the trial, not just the most likely or the most desirable.

Finally, while random assignment of treatment and control conditions is typically considered "fair" in academic circles, it may not be perceived so by all employees in the organizations, especially those who did not receive the treatment. When objections are raised by managers or employers around randomization, we most frequently hear this question: how can the organization justify withholding a "desirable treatment" from all of employees? The answer we propose is twofold. First, organizations typically run "pilots" on some (but not all) units within the

organization before rolling out the change widely. Few executives oppose the idea of a pilot and a staged roll-out. So, a field experiment can be viewed as the first part of a multi-stage rollout process where the randomly chosen treatment group receives the pilot first and the control group receives it after the evaluation of results has been completed – a process, which Campbell (1969) referred to as "staged innovation." Second, we caution scholars and organizations alike that an untested treatment is never sure to deliver the desired outcome; that is, without a field experiment that provides causal evidence for a treatment effect, the treatment may have all sorts of effects – from positive to negative to null. Due to the ambiguity of the actual treatment effect in advance of an experiment, scholars should not promise an organization a quick "fix" or a "desirable treatment" for their problems; instead treatment effects are unknown until tested. Thus, no employee is getting preferential treatment. But to find out if the treatment works as desired—and, more generally, to find out "what works"—randomization is the fairest and most effective way.

Generalizability

Some have also argued that field experiments are particularly limited in their ability to generalize an observed phenomenon because they are tested only within the scope of specific organizational contexts (Deaton & Cartwright, 2016). That is, field experiments provide rigorous evidence of causality, but tells us little about whether the causal linkage presented would apply in other contexts, if the experiment itself were to be replicated in another place or time. Furthermore, small changes can sometimes have big effects, and they may be varied by a scholar or an organization without realizing that they might matter. For example, a scholar who is trying to test the impact of changing language in an employee-wide email will have to make choices about font,

grammar, length of email, time of day sent, and a myriad of other factors that could all play into the success of the intervention. Field experiments are designed for strong internal validity, but not necessarily external validity; they can be a "black box" according to this argument. Importantly, Deaton (2010) and others have argued that attempting to tease further analyses out of a field experiment, to understand, for example, what works for whom or to better explore mechanisms ex-post is similar to data-mining. After all, if you run 20 regressions, a p-value of 0.05 (the current threshold for statistical significance—though see Benjamin et al. (2017) for a recommendation and discussion that the threshold be lowered to 0.005 to mitigate some of these problems) would allow for one of those regressions to show a significant effect even if there is no underlying impact. Many of these arguments are common to most field-based methods. However, organizational theory can be helpful when addressing the challenge of assessing the external validity of field experiments. As we have argued above, field experiments can be excellent tools to sharpen and develop theory. However, at the same time, organizational theory is necessary for conducting a field experiment that contributes to generalizable knowledge (for in-depth discussions of this topic, see Card, DellaVigna, & Malmendier (2011), Sutton & Staw (1995) and Weick (1995)). Only when a field experiment is grounded in a theoretical framework can we understand its findings: what does it mean when an independent variable has a predicted effect? Does it advance our understanding of the behavior studied in a meaningful way that is not dependent on the institutional structure of the organization? And, when a field experiment does not "work" (assuming sufficient statistical power and no experimental shortcomings), what would be a theoretically relevant interpretation that might explain the finding – or does it help advance the theory in a new

direction?[6] Ideally these questions can and should be answered before an experiment is launched and certainly before any analysis is conducted. By committing to a theoretical interpretation of the results before analysis, experimenters can avoid criticism of data mining and can ensure that their research is both theoretically grounded as well as clearly explored. To that end, there is a growing movement of experimental researchers who aim to make pre-registration of experimental and analysis protocols the norm, so as to avoid both data mining and to push researchers to explain in advance what specific theoretical mechanism they are hoping to test and how they will go about doing so (Munafò et al., 2017). Pre-registration is a powerful commitment device and strengthens empirical evidence for theoretical claims (but their value may depend on the type of research conducted; see Coffman & Niederle, 2015). For interested readers and those interested in pre-registering their research, we recommend the tutorials and free, easy-to-use pre-registration services offered by AsPredicted.org, the Open Science Framework (www.osf.io), or Evidence in Governance and Politics (www.egap.org).

Turning results from a field experiment into theoretically relevant insights is often challenging. As many others have argued before us (Abdul Latif Jameel Poverty Action Lab, n.d.; Munafò et al., 2017; Nosek et al., 2015), good experimentation requires careful planning, starting with building on strong theory, developing testable predictions, designing a clear test of the

---

[6] We want to emphasize the importance of publishing null results, as others have advocated before us (e.g., see Easterbrook, Berlin, Gopalan, & Matthews, 1991; Landis & Rogelberg, 2013; Landis, James, Lance, Pierce, & Rogelberg, 2014, and Simmons, Nelson, & Simonsohn, 2011), to inform and advance our understanding of interventions that do not change intended outcomes. Of course, null results are only as insightful as the underlying theoretical motivation for conducting the experiment in the first place: the more theoretically compelling the reasons for testing the interventions were (and the better the methodological execution, e.g. see also pre-registered reports: Bouwmeester et al., 2017 and Simons, Holcombe, & Spellman, 2014), the more learning can be gained from a published null result.

hypotheses, pre-registering an analysis plan before the experiment is analyzed, and carrying out the experiment and analysis as planned to make valid theoretical inferences. (For more explorative research projects, adjustments—but no less planning—are needed to derive meaningful theoretical insights; see Center for Open Science (n.d.).)

Consider the role of experimental design in deriving generalizable insights. Some of the most impactful field experiments focus on testing "mechanisms:" by focusing not just on outcomes but also on underlying mechanisms, findings in one context are more likely to be generalizable to other contexts (Bates & Glennerster, 2017; Chatterji, Findley, Jensen, Meier, & Nielson, 2016; Ludwig, Kling, & Mullainathan, 2011). We thus encourage organizational scholars to combine their AFE, PFE and IFE with "mechanism" tests whenever possible. Practically speaking, scholars might include, not just an experimental condition that produces the desired treatment effect (as predicted by the theory), but *also* include a condition that turns off this effect by blocking a theorized pathway responsible for the effect. That is, by also testing a version of the intervention where the treatment should not show the same impact, if the theorized mechanism is correct, field experiments can operate as a test for both mechanism and broader causality. For helpful illustrations of organizational field experiments designed to test underlying mechanisms see the followings two examples. First, Cable, Gino and Staats (2013) include two experimental treatment conditions so they can distinguish between individualized and institutionalized socialization strategies (instead of just one generic socialization condition). Second, Gilchrist, Luca and Malhotra (2016) document a productivity increase in response to a "surprise bonus" for employees – they are able to attribute the increase to the "surprise" element causally by including two control conditions: a first control condition that pays the same baseline payment as the treatment condition

without a bonus and a second control condition that pays as a baseline the combined amount that the employee receives in the treatment condition from both the baseline and the bonus payments, thus ruling out the possibility that higher effort is simply the result of higher payment. For excellent discussions of the importance of mechanisms to promote of generalizability of field experiments across contexts, see Bates & Glennerster (2017), Congdon, Kling, Ludwig, & Mullainathan (2016) and Green, Ha, & Bullock (2010).

Another approach to addressing the challenge of generalizability is to use a multi-method approach that includes both field experiments as well as other research methods, cycling between the different methods to provide a richer description of the phenomenon at hand. Although field experiments are the gold standard for causal investigations in the field, fully understanding the depth and breadth of a research problem often requires "multiple operationalism" (Campbell & Fiske, 1959). Jick (1979) illustrates that "triangulation" of methods (Denzin & Lincoln, 1994; Webb, Campbell, Schwartz, & Sechrest, 1966) combining both qualitative and quantitative approach is a powerful way to fully understand the same phenomenon. The combination of methods provides for a richer description, greater exploration, and deeper theoretical insight than using just one of the many methods available to organizational scholars. Chatman and Flynn (2005) go one step further by arguing that research methods should not only be combined once but field, lab, and observational methods should be applied in a continuous cyclical fashion to close the gap that any single method by itself would leave exposed. Indeed, Staw (2016) sends a warning and a reminder to all organizational scholars that, "given that all research is flawed in some fundamental way (McGrath et al. 1982), the only route to achieving a better understanding of a phenomenon is through the use of multiple methodologies" (p. 11).

Conclusion

Field experiments are a powerful method with particular appeal to organizational scholars: they enable organizational scholars to explore theory-driven research questions in an organizational setting with a degree of confidence about causality that other methods cannot provide. While field experiments are not right for every research question or research opportunity, we argue that they are more feasible now than in the past and that they are well suited to two aspects of organizational behavior about which most scholars care: *organizational context* and *causality*.

To help scholars develop experimental designs that suit their research questions, we have proposed an easy-to-use taxonomy: field experiments which test whether organizational processes work as they should (AFE); field experiments which change an existing component of an organizational process (PFE); or field experiments which innovate new processes (IFE).

But what about the widely bemoaned difficulty of successfully executing a field experiment with an organization? We believe this difficulty tends to be over-stated in contemporary contexts, or at least over-generalized. While it can be difficult to run field experiments in organizations (as it can be difficult to carry out qualitative field studies or laboratory experiments), conducting field experiments has never been easier: scholars will find common ground with many organizations willing to share data and to experiment to gain insights and competitive advantage. Moreover, technological progress means that many field experiments can be conducted at a very low cost, ranging from easy and fast online communications channels (e.g., email, text, and mobile apps), easy-to-use online data collection tools (e.g., Qualtrics or

SurveyMonkey surveys) as well as new forms of automatically captured data (e.g., click-throughs on links, timestamps of all computer-based actions, or entire social networks between individuals within and outside a company). Many existing websites, survey platforms, and communication software have built-in randomization or "A/B testing" functions (e.g., Optimizely for website pages and click-throughs; Google Ads and Facebook Ads for running field experiments involving advertising; and Qualtrics for surveys), thus simplifying the back-end of randomizing and making the field experiment itself compatible with platforms that organization are more accustomed to using (and allowing for easy sharing of data).

Finally, and most importantly, field experiments create value for both organizational scholars and organizations alike. They advance theoretical scholarship and they help organizations become more effective. While the importance of contributing to theory is well-recognized in organizational scholarship (Gephart, 2004; Suddaby, 2006; Sutton & Staw, 1995; Weick, 1995), the relevance of having an impact on management and organizational practice is not as commonly highlighted. However, most academic scholars also care about how organizations work and function in the real world, and they want to make a real and lasting impact on organizational practice – and many would like their scholarly contributions to have an impact they can see and measure. Field experiments often do just that: they can help organizations address real problems and, thanks to the causal inference made possible through the randomization process, scholars can observe and measure the differences they are making in organizations after running field experiments. Scholars who conduct field experiments in organizations have the opportunity to have an organizational impact, as their partner organizations can learn more about "what works" (and often, what does not work). Well-designed field experiments further contribute to

generalizable knowledge that many organizations can subsequently use to improve and innovate in their organizations. In short, field experiments serve the cause we are hoping to achieve – by more fully bringing scientific thinking to management practice.

References

Abdul Latif Jameel Poverty Action Lab. (n.d.). Transparency & Reproducibility. Retrieved June 23, 2017, from https://www.povertyactionlab.org/research-resources/transparency-and-reproducibility

Anderson, C., & Galinsky, A. D. (2006). Power, optimism, and risk-taking. *European Journal of Social Psychology*, *36*(4), 511–536. http://doi.org/10.1002/ejsp.324

Anderson, L., Fryer, R., & Holt, C. (2006). Discrimination: experimental evidence from psychology and economics. In *Handbook on the Economics of Discrimination* (pp. 97–118).

Ashforth, B. E., & Mael, F. (1989). Social Identity Theory and the Organization. *Academy of Management Review*, *14*(1), 1–20.

Bates, M. A., & Glennerster, R. (2017). The Generalizability Puzzle. Retrieved June 1, 2017, from https://ssir.org/articles/entry/the_generalizability_puzzle

Bauer, T. N., Bodner, T., Erdogan, B., Truxillo, D. M., & Tucker, J. S. (2007). Newcomer adjustment during organizational socialization: A meta-analytic review of antecedents, outcomes, and methods. *The Journal of Applied Psychology*, *92*(3), 707–721. http://doi.org/10.1037/0021-9010.92.3.707

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements. *Perspectives on Psychological Science*, *2*(4), 396–403.

Bazerman, M. H., & Gino, F. (2012). Behavioral Ethics: Toward a Deeper Understanding of Moral Judgment and Dishonesty. *Annual Review of Law and Social Science*, *8*(1), 85–104. http://doi.org/10.1146/annurev-lawsocsci-102811-173815

Bendick, M., Jackson, C. W., & Reinoso, V. A. (1994). Measuring employment discrimination through controlled experiments. *The Review of Black Political Economy*, 23(1), 25–48. http://doi.org/10.1007/BF02895739

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2017). Redefine statistical significance. *Nature Human Behaviour*, 1–0. http://doi.org/10.1038/s41562-017-0189-z

Bergman, P., & Rogers, T. (2017). Is this Technology Useless? How Seemingly Irrelevant Factors Affect Adoption and Efficacy. *Working Paper*, 1–33.

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination. *The American Economic Review*. http://doi.org/10.3386/w9873

Beshears, J., & Gino, F. (2015). Leaders as Decision Architects. *Harvard Business Review*, 1–11.

Bohnet, I. (2016). What works: Gender equality by design. Harvard University Press.

Booth, A. L., Leigh, A., & Varganova, E. (2011). Does Ethnic Discrimination Vary Across Minority Groups? Evidence from a Field Experiment. *Oxford Bulletin of Economics and Statistics*, *74*(4), 547–573. http://doi.org/10.1111/j.1468-0084.2011.00664.x

Boruch, R. F., & Wothke, W. (1985). Seven Kinds of Randomiation Plans for Designing Field Experiments. *New Directions for Program Evaluation*, 1–19.

Bouwmeester, S., Verkoeijen, P., Aczel, B., Barbosa, F., Begue, L., Branas-Garza, P., et al. (2017). Registered Replication Report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, 1–16.

Brief, A., & Smith Crowe, K. (2015). Organizations Matter. In *The Social Psychology of Good*

and Evil.

Buell, R. W., & Norton, M. I. (2011). The Labor Illusion: How Operational Transparency Increases Perceived Value. *Management Science*, *57*(9), 1564–1579. http://doi.org/10.1287/mnsc.1110.1376

Buell, R. W., Kim, T., & Tsay, C.-J. (2017). Creating Reciprocal Value Through Operational Transparency. *Management Science*, *63*(6), 1673–1695. http://doi.org/10.1287/mnsc.2015.2411

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science*, *6*(1), 3–5. http://doi.org/10.1177/1745691610393980

Burbano, V. C. (2016). Social Responsibility Messages and Worker Wage Requirements: Field Experimental Evidence from Online Labor Marketplaces. *Organization Science*, *27*(4), 1010–1028. http://doi.org/10.1287/orsc.2016.1066

Cable, D. M., Gino, F., & Staats, B. R. (2013). Breaking Them in or Eliciting Their Best? Reframing Socialization around Newcomers' Authentic Self-expression. *Administrative Science Quarterly*, *58*(1), 1–36. http://doi.org/10.1177/0001839213477098

Campbell, D. (1969). Reforms as Experiments. *American Psychologist*, *24*(4), 409–429.

Campbell, D. T. (1991). Methods for the Experimenting Society. *Evaluation Practice*, *12*(3), 223–260.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 1–25.

Camuffo, A., Cordova, A., & Gambardella, A. (2017). A scientific approach to entrepreneurial experimentation: Evidence from a randomized control trial. *Working Paper*, 1–34.

Cappelli, P., & Sherer, P. D. (1991). The missing role of context in OB: The need for a meso-level approach. *Research in Organizational Behavior*, *13*, 55–110.

Card, D., DellaVigna, S., & Malmendier, U. (2011). The Role of Theory in Field Experiments. *Journal of Economic Perspectives*, *25*(3), 39–62. http://doi.org/10.1257/jep.25.3.39

Center for Open Science. (n.d.). Preregistration makes your science better by increasing the credibility of your results. Retrieved June 23, 2017, from https://cos.io/prereg/

Chatman, J. A., & Flynn, F. J. (2005). Full-Cycle Micro-Organizational Behavior Research. *Organization Science*, *16*(4), 434–447. http://doi.org/10.1287/orsc.1050.0136

Chatterji, A. K., Findley, M., Jensen, N. M., Meier, S., & Nielson, D. (2016). Field experiments in strategy research. *Strategic Management Journal*, *37*(1), 116–132. http://doi.org/10.1002/smj.2449

Coffman, L. C., & Niederle, M. (2015). Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible. *Journal of Economic Perspectives*, *29*(3), 81–98. http://doi.org/10.1257/jep.29.3.81

Congdon, W. J., Kling, J. R., Ludwig, J., & Mullainathan, S. (2016). Social Policy: Mechanism experiments & policy evaluations. In A. Banerjee & E. Duflo (Eds.), (pp. 1–60).

Cook, T. D., & Campbell, D. T. (1979). Quasi-Experimentation: Design & Analysis Issues for Field Settings. Houghton Mifflin Company.

Dana, J., Dawes, R., & Peterson, N. (2013). Belief in the unstructured interview: The persistence of an illusion. *Judgement and Decision Making*, *8*(5), 512–520.

Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of*

Economic Literature, *48*(2), 424–455. http://doi.org/10.1257/jel.48.2.424

Deaton, A., & Cartwright, N. (2016). *Understanding and Misunderstanding Randomized Controlled Trials* (pp. 1–69).

Denzin, N. K., & Lincoln, Y. S. (1994). Handbook of qualitative research. SAGE Publications.

Dolan, P., & Galizzi, M. M. (2015). Like ripples on a pond: behavioral spillovers and their implications for research and policy. *Journal of Economic Psychology*, 47, 1-16.

Desai, S. D., & Kouchaki, M. (2015). Work-report formats and overbilling: How unit-reporting vs. cost-reporting increases accountability and decreases overbilling. *Organizational Behavior and Human Decision Processes*, *130*(C), 79–88. http://doi.org/10.1016/j.obhdp.2015.06.007

Dvir, T., Eden, D., Avolio, B. J., & Shamir, B. (2002). Impact of Transformational Leadership on Follower Development and Performance: A Field Experiment. *Academy of Management Journal*, *45*(4), 735–744.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. (1991). Publication bias in clinical research. *The Lancet*, *337*(8746).

Eden, D. (2017). Field Experiments in Organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*(1), 91–122. http://doi.org/10.1146/annurev-orgpsych-041015-062400

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*, 350–383.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *Academy of Management Review*, *14*(4), 532–550.

Feldman, M. S., & Orlikowski, W. J. (2011). Theorizing Practice and Practicing Theory. *Organization Science*, *22*(5), 1240–1253. http://doi.org/10.1287/orsc.1100.0612

Galizzi, M. M., & Navarro-Martinez, D. (2015). On the External Validity of Social- Preference Games: A Systematic Lab-Field Study. *Working Paper*, 1–38.

George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, *57*(2), 321–326. http://doi.org/10.5465/amj.2014.4002

Gephart, R. P. (2004). Qualitative Research and the Academy of Management Journal. *Academy of Management Journal*, 1–10.

Gerber, A. S., & Green, D. P. (2012). Field experiments: Design, analysis, and interpretation. WW Norton.

Gilchrist, D. S., Luca, M., & Malhotra, D. (2016). When 3 + 1 > 4: Gift Structure and Reciprocity in the Field. *Management Science*, *62*(9), 2639–2650. http://doi.org/10.1287/mnsc.2015.2275

Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science*.

Gino, F., Krupka, E. L., & Weber, R. A. (2013). License to Cheat: Voluntary Regulation and Ethical Behavior. *Management Science*, *59*(10), 2187–2203. http://doi.org/10.1287/mnsc.1120.1699

Glaser, B. G., & Strauss, A. L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago: Aldine Publishing Company.

Glennerster, R., & Takavarasha, K. (2013). Running randomized evaluations: A practical guide. Princeton University Press.

Gneezy, A., Gneezy, U., & Riener, G. (2012). Pay-what-you-want, identity, and self-signaling in

markets. *Proceedings of the National Academy of Sciences*, *109*(19). http://doi.org/10.1073/pnas.1120893109/-/DCSupplemental

Gneezy, A., Gneezy, U., Nelson, L. D., & Brown, A. (2010). Shared Social Responsibility: A Field Experiment in Pay-What-You-Want Pricing and Charitable Giving. *Science*, *329*(5989), 325–327. http://doi.org/10.1126/science.1186744

Gneezy, U., & Imas, A. (2017). Lab in the Field: Measuring Preferences in the Wild. In *Handbook of Economic Field Experiments* (pp. 439–464).

Goldin, C., & Rouse, C. (2000). Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians. *American Economic Review*, *90*(4), 715–741.

Grant, A. M. (2012). Leading with Meaning: Beneficiary Contact, Prosocial Impact, and the Performance Effects of Transformational Leadership. *Academy of Management Journal*, *55*(2), 458–476. http://doi.org/10.5465/amj.2010.0588

Grant, A. M., Campbell, E. M., Chen, G., Cottone, K., Lapedis, D., & Lee, K. (2007). Impact and the art of motivation maintenance: The effects of contact with beneficiaries on persistence behavior. *Organizational Behavior and Human Decision Processes*, *103*(1), 53–67. http://doi.org/10.1016/j.obhdp.2006.05.004

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough Already about "Black Box" Experiments: Studying Mediation Is More Difficult than Most Scholars Suppose. *The ANNALS of the American Academy of Political and Social Science*, *628*(1), 200–208. http://doi.org/10.1177/0002716209351526

Hackman, R., & Lawler, E. E. (1971). Employee reactions to job characteristics. *Journal of Applied Psychology*, *55*(3).

Harrison, G. W., & List, J. A. (2004). Field Experiments. *Journal of Economic Literature*, *42*(4), 1009–1055.

Hauser, O. P., Kraft-Todd, G., Rand, D. G., Nowak, M. A., Norton, M. I. (2016). Invisible Inequality Leads to Punishing the Poor and Rewarding the Rich. *Academy of Management Proceedings*.

Hauser, O. P., Hendriks, A., Rand, D. G., & Nowak, M. A. (2016). Think Global, Act Local: Preserving the Global Commons. *Scientific Reports*, *6*.

Hauser, O. P., Rand, D. G., Peysakhovich, A., & Nowak, M. A. (2014). Cooperating with the future. *Nature*, *511*(7508), 220–223. http://doi.org/10.1038/nature13530

Hauser, O., & Luca, M. (2015a). How to Design (and Analyze) a Business Experiment. *Harvard Business Review*.

Hauser, O., & Luca, M. (2015b). Your Company Is Full of Good Experiments (You Just Have to Recognize Them). *Harvard Business Review*.

Ibanez, M. R., & Staats, B. R. (2016). Behavioral Empirics and Field Experiments. In *The Handbook of Behavioral Operations* (K. L. Donohue, E. Katok and S. Leider, Eds.). New York: Wiley.

Jick, T. D. (1979). Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, *24*(4), 602–611.

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review*.

Kang, S. K., DeCelles, K. A., Tilcsik, A., & Jun, S. (2016). *Whitened Résumés: Race and Self-Presentation in the Labor Market. Administrative Science Quarterly* (pp. 1–55).

Kettle, S., Hernandez, M., Sanders, M., Hauser, O., & Ruda, S. (2017). Failure to CAPTCHA Attention: Null Results from an Honesty Priming Experiment in Guatemala. *Behavioral Sciences*, *7*(28). http://doi.org/10.3390/bs7020028

Kim, G.-H., Trimi, S., & Chung, J.-H. (2014). Big-data applications in the government sector. *Communications of the ACM*, *57*(3), 78–85. http://doi.org/10.1145/2500873

King, E. B., Hebl, M. R., Botsford Morgan, W., & Ahmad, A. S. (2013). Field Experiments on Sensitive Organizational Topics. *Organizational Research Methods*, *16*(4), 501–521. http://doi.org/10.1177/1094428112462608

Kohavi, R., & Thomke, S. (2017). The Surprising Power of Online Experiments. *Harvard Business Review*.

Landis, R. S., & Rogelberg, S. G. (2013). Our Scholarly Practices Are Derailing Our Progress: The Importance of "'Nothing'" in the Organizational Sciences. *Industrial and Organizational Psychology*, *6*(3), 299–302.

Landis, R. S., James, L. R., Lance, C. E., Pierce, C. A., & Rogelberg, S. G. (2014). When is Nothing Something? Editorial for the Null Results Special Issue of Journal of Business and Psychology. *Journal of Business and Psychology*, *29*(2), 163–167. http://doi.org/10.1007/s10869-014-9347-8

List, J. A. (2011). Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *Journal of Economic Perspectives*, *25*(3), 3–16. http://doi.org/10.1257/jep.25.3.3

Luca, M., & Hauser, O. (2016, February 4). Good Communication Requires Experimenting with Your Language. Retrieved June 22, 2017, from https://pdf.printfriendly.com/pdfs/make

Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism experiments and policy evaluations. *Journal of Economic Perspectives*, *25*(3).

Magee, J. C., Galinsky, A. D., & Gruenfeld, D. H. (2007). Power, Propensity to Negotiate, and Moving First in Competitive Interactions. *Personality and Social Psychology Bulletin*, *33*(2), 200–212. http://doi.org/10.1177/0146167206294413

Malhotra, D., & Lumineau, F. (2011). Trust and Collaboration in the Aftermath of Conflict: The Effects of Contract Structure. *Academy of Management Journal*, *54*(5), 981–998. http://doi.org/10.5465/amj.2009.0683

Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*.

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 1–9.

McWilliams, A. (n.d.). Corporate Social Responsibility. In *Wiley Encyclopedia of Management* (Vol. 12).

Milkman, K. L., Akinola, M., & Chugh, D. (2012). Temporal Distance and Discrimination. *Psychological Science*, *23*(7), 710–717. http://doi.org/10.1177/0956797611434539

Montuori, A., & Purser, R. E. (1995). Deconstructing the Lone Genius Myth: Toward a Contextual View of Creativity. *Journal of Humanistic Psychology*, *35*(3), 1–44.

Mowday, R. T., & Sutton, R. I. (1993). Organizational Behavior: Linking Individuals and Groups to Organizational Contexts. *Annual Review of Psychology*, *44*, 195–229.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Sert, du, N. P., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 1–9. http://doi.org/10.1038/s41562-016-0021

Neuman, W. L. (1991). Social research methods: Qualitative and quantitative approaches. Allyn & Bacon.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. http://doi.org/10.1126/science.aab2374

OReilly, C. A., III, Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of Management Journal*, *34*(3), 487–516.

Pager, D., Western, B., & Bonikowski, B. (2009). Discrimination in a Low-Wage Labor Market: A Field Experiment. *American Sociological Review*, *74*, 777–799.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, *23*(3), 184–188. http://doi.org/10.1177/0963721414531598

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419. Retrieved from http://repub.eur.nl/pub/31983

Parsons, H. M. (1974). What happened at Hawthorne? *Science*, *183*(4128), 922–932.

Patton, M. Q. (1990). Qualitative evaluation and research methods.

Pfeffer, J., & Sutton, R. I. (2006). Evidence-Based Management. *Harvard Business Review*, 1–13.

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly*, *41*(4), 574–599.

Robinson, S. L., Kraatz, M. S., & Rousseau, D. M. (1994). Changing Obligations and the Psychological Contract: A Longitudinal Study. *Academy of Management Journal*, *37*(1), 137–152.

Rousseau, D. (1989). Psychological and Implied Contracts in Organizations. *Employee Resonsibilities and Rights Journal*, *2*(2).

Salancik, G. R. (1979). FIeld Stimulations for Organizational Behavior Research. *Administrative Science Quarterly*, *24*(4), 638–649.

Scandura, T. A., & Williams, E. A. (2000). Research Methodology in Management: Current Practices, Trends, and Implications for Future Research. *Academy of Management Journal*, *43*(6), 1248–1264.

Shadish, W. R., & Cook, T. D. (2009). The Renaissance of Field Experimentation in Evaluating Interventions. *Annual Review of Psychology*, *60*(1), 607–629. http://doi.org/10.1146/annurev.psych.60.110707.163544

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Cengage Learning. http://doi.org/10.1198/jasa.2005.s22

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest Deed, Clear Conscience: When Cheating Leads to Moral Disengagement and Motivated Forgetting. *Personality and Social Psychology Bulletin*, *37*(3), 330–349. http://doi.org/10.1177/0146167211398138

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, *109*(38), 15197–15200.

http://doi.org/10.1073/pnas.1209746109

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. http://doi.org/10.1177/0956797611417632

Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An Introduction to Registered Replication Reports at Perspectives on Psychological Science. *Perspectives on Psychological Science*, *9*(5), 552–555. http://doi.org/10.1177/1745691614543974

Social and Behavioral Sciences Team. (2015). 2015 Annual Report. *White Paper*, 1–64.

Staw, B. M. (1977). The Experimenting Organization. *Organizational Dynamics*, 3–18.

Staw, B. M. (2010). The Trouble With JDM: Some Limitations to the Influence of JDM on Organizational Research. *Industrial and Organizational Psychology*, *3*, 411–416.

Staw, B. M. (2016). Stumbling Toward a Social Psychology of Organizations: An Autobiographical Look at the Direction of Organizational Research. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*(1), 1–19.

Suchman, E. (1968). Evaluative Research: Principles and Practice in Public Service and Social Action Programs. New York: Russell Sage Foundation.

Suddaby, R. (2006). From the editors: What grounded theory is not. *Academy of Management Journal*, 1–11.

Suen, H. K., & Ary, D. (1989). Analyzing quantitative behavioral observation data. Psychology Press.

Sutton, R. I., & Staw, B. M. (1995). What Theory is Not. *Administrative Science Quarterly*, *40*, 371–384.

Tajfel, H. (1982). Social Psychology of Intergroup Relations. *Annual Review of Psychology*, *33*, 1–39.

Thau, S., Pitesa, M., & Pillutla, M. (2014). Experiments in Organizational Behavior. In *Laboratory Experiments in the Social Sciences* (pp. 433–447). Elsevier. http://doi.org/10.1016/B978-0-12-404681-8.00019-4

The Editor. (1958). Front Matter. *Academy of Management Journal*, *1*(1), 5–6.

Tsay, C. J. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences*. http://doi.org/10.1073/pnas.1221454110/-/DCSupplemental

Tsay, C.-J. (2016). Privileging Naturals Over Strivers. *Personality and Social Psychology Bulletin*, *42*(1), 40–53. http://doi.org/10.1177/0146167215611638

Tsay, C.-J., & Banaji, M. R. (2011). Naturals and strivers: Preferences and beliefs about sources of achievement. *Journal of Experimental Social Psychology*, *47*(2), 460–465. http://doi.org/10.1016/j.jesp.2010.12.010

Van Maanen, J., & Schein, E. H. (1979). Toward a theory of organizational socialization. *Research in Organizational Behavior*, *1*, 209–264.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). Unobtrusive measures: Nonreactive research in the social sciences. Rand McNally.

Weick, K. E. (1967). Promise and limitations of laboratory experiments in the development of attitude change theory. In M. Sherif & C. W. Sherif (Eds.), *Attitude, ego involvement, and change*.

Weick, K. E. (1995). What Theory Is Not, Theorizing Is. *Administrative Science Quarterly*, *40*(3), 385–390.

Table 1. Proposed taxonomy of field experiments in organizational research. An **Audit Field Experiments (AFE)** varies the inputs that probe an existing organizational process, testing whether the process functions as expected and uncovering potential shortcomings or biases. A **Procedural Field Experiments (PFE)** varies parameters of the process while keeping the input constant: PFEs allow scholars to understand the effect of independently varying one component of an existing organizational process on an outcome variable of interest. An **Innovation Field Experiments (IFE)** is a systematic methodology to introduce an entirely new process to the organizational setup and test its effect of an outcome of interest.

	Audit Field Experiment (AFE)	**Procedural Field Experiment (PFE)**	**Innovation Field Experiment (IFE)**
Variation of *inputs* into process?	**Yes**	No	No
Variation of *existing components* of a process?	No	**Yes**	No
Introduction of *new* process?	No	No	**Yes**