Primer for Experimental Methods in Organization Theory

Sheen S. Levine,[a,]* Oliver Schilke,[b] Olenka Kacperczyk,[c] Lynne G. Zucker[d]

[a] The University of Texas at Dallas, Richardson, Texas 75080 and Columbia University, New York, New York 10027; [b] The University of Arizona, Tucson, Arizona 85721; [c] London Business School, London NW1 4SA ; [d] The University of California, Los Angeles 90095
*Corresponding author

Experiments have long played a crucial role in various scientific disciplines and have been gaining ground in organization theory, where they add unique value by establishing causality and uncovering theoretical mechanisms. This essay provides an overview of the merits and procedures of the experimental methodology, with an emphasis on its application to organization theory. Drawing on the historical roots of experiments and their impact across science, we argue the method holds immense potential for furthering organization theory. We highlight key advantages of experimental methods, including high internal and construct validity, vividness in communicating findings, the capacity to examine complex and understudied phenomena, and the identification of microfoundations and theoretical mechanisms. We alleviate some concerns about external validity and offer guidance for designing and conducting sound, reproducible experimental research. Ultimately, we contend that the current experimental turn holds the potential to reorient organization theory.

Keywords: Organization Theory, Experiment, Microfoundation, Mechanism

Funding: This work was supported by the Otto Moensted Foundation, which granted the Otto Moensted Visiting Professorship to S. S. Levine and a National Science Foundation CAREER Award from the Directorate for Social, Behavioral and Economic Sciences [Grant 1943688] to O. Schilke.

Acknowledgments: We are thankful for comments and suggestions from Gary Bolton, Jon Evans, Jana Gallus, Alfonso Gambardella, Patrick Haack, Hyunjin Kim, Ronald Klingebiel, Rembrand (Rem) Koning, Tamar Kugler, Yuxin Lin, Rosemarie Nagel, Henning Piezunka, Nathan Podsakoff, Phanish Puranam, Martin Reimann, Yueling Sima, Alexey Tyulyupo, Roberto A. Weber, Bart J. Wilson, Arthur Xue, Jihye Yeo, and Zixu Zhang. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Otto Moensted Foundation.

> "The first and most obvious distinction between Observation and Experiment is, that the latter is an immense extension of the former. It not only enables us to produce a much greater number of variations in the circumstances than nature spontaneously offers, but also, in thousands of cases, to produce the precise *sort* of variation which we are in want of for discovering the law of the phenomenon; a service which nature, being constructed on a quite different scheme from that of facilitating our studies, is seldom so friendly as to bestow upon us."

(John Stuart Mill, *A System of Logic*, 1874: p. 274, § 3)

How does life emerge? In mid-nineteenth century Europe, the prevailing theory was spontaneous generation, holding that organisms emerge from nonliving matter, like yeast spawning out of air. Louis Pasteur, a young scientist, had a different hypothesis: air merely carries organisms that are invisible to the eye. Seeking a design to test the idea, he examined the details of century-old experiments and came up with a new instrument: a glass container, filled with nutrient-rich broth, topped by a slender twisted tube. If air was the source of life, he reasoned, the added tube would not stop regeneration. Conversely, if life was carried by invisible air-borne spores, the tube would stop them from reaching the broth. He started with three identical flasks. In the first, the tube filtered the air reaching the liquid. In the second, he broke off the tube, letting the air meet the nutrients directly. In the third, he let the broth flow through the tube, similarly exposing it to air. The results—that yeast was absent from the first flask but developed in the others—contradicted the scientific consensus. Instead, the findings evidenced how tiny organisms, invisible to the naked eye, are carried in the air. Pasteur's discovery, easy to replicate, paved the way for germ theory, which revolutionized medicine and agriculture, bringing inventions like pasteurization and vaccines, saving millions of lives.

Louis Pasteur's work epitomizes a good experiment: It is simple, elegant — and decisive. Since they excel in proving causality, experiments have become the workhorses of the sciences. They are as central to physics as they are to biology and chemistry. In medicine, randomized control trials, a form of field experiments, are routinely deployed to study the efficacy of drugs, surgical techniques, medical devices, diagnostic procedures, and other treatments. In the social sciences, psychology was quick to adopt the experimental method. Since the late nineteenth century, the method has become central in most branches of psychology (and, subsequently, in micro-organizational behavior). In economics, starting in the mid-twentieth century, experiments began to assume a more central role, examining behaviors that had been theorized and studied in archival data, but rarely directly. The results shook well-established theories, as we exemplify throughout.

When organization theory was born, experiments were sitting bedside, in the delivery room. In a nine-page memorandum, Herbert Simon and Harold Guetzkow (1952) proposed to George L. Bach, their dean at Carnegie Mellon, a program of "research into behavior in Organizations."[1] The research proposal began with field studies "employing interviewing and observational techniques," which were to inspire ideas to be tested through "laboratory experimentation," culminating in "theory development." The grant proposal must have been fruitful, as their laboratory became a spring of experimental research, supplementing work that was purely theoretical or relied exclusively on archival data.

After reading hundreds of experimental papers, including those submitted to this special issue, we can offer some guidance on experimental research and its merits for organization theory. We outline why experiments are considered the gold standard of evidence and how they have propelled scientific disciplines, including psychology and economics, which feed organization theory. We then turn to the benefits of experimental methods for organization theory, particularly at the field's current stage of development. We also address some of the criticism leveled against experiments before providing guidance on what constitutes good experimental research. In the conclusion section, we elaborate on recent developments in the experimental research community. We call for new types of experiments—including the study of experimental organizations—to enrich our understanding. The current experimental turn, we believe, can fundamentally reorient the field of organization theory.

Experiments for Organization Theory

Although we may never achieve complete consensus on what precisely organization theory is and how it should be studied (Pfeffer 1993; Heugens 2008), the term in and of itself spells out what sets the field apart: (a) its central unit of analysis—the organization and (b) its relentless commitment to theory.

[1] We thank Mie Augier for unearthing the original request.

Both characteristics make experiments a uniquely suited method for organization theorists. *Organizations* can be defined as "systems of coordinated action among individuals" (March and Simon 1993, p. 2) who pursue collective goals (for a review, see Puranam 2018, pp. 4-9), and experiments are highly capable of zooming in on relevant aspects of these social systems, providing unparalleled analytic precision (Zelditch 1969). *Theory* "emphasizes the nature of causal relationships, identifying what comes first," and "delves into underlying processes" (Sutton and Staw 1995, p. 378). Experiments excel in identifying causal relationships and can provide strong evidence for the underlying processes, making them an appealing methodological choice for organization theorists. Finally, they offer an unparalleled ability to examine behavior in controlled circumstances and with robust measures, cutting through the complexity and ambiguity of organizations.

From their newly funded initiative in Pittsburgh, Herbert Simon, Harold Guetzkow, and their colleagues began advocating for experiments in organization theory. In an early *Management Science* article, Harold Guetzkow and Anne E. Bowes (1957, p. 381) were among the first to propose that it was "feasible to develop organizations in the laboratory." Their colleagues Richard Cyert, Edward Feigenbaum, and James March (1959, p. 94) insisted that "(m)any of the major propositions in organization theory depend on evidence generated by studies in the laboratory." In his comprehensive assessment of experimentation with organizations, Karl Weick (1965, p. 254) similarly concluded that experiments provide "an extremely powerful means to gain substantial knowledge about organizations," and Stanley I. Hyman (1970, p. 52) went as far as to argue that "[m]anagement, to become a useful science, must base its knowledge on scientific experiments."

We agree with these early observations and have no doubt that, as in biology (Siegfried 2022) or behavioral economics (Bardsley et al. 2010), experiments can elevate organization theory by promoting understanding of theoretical mechanisms. Like experimental markets, a methodological innovation that brought the Nobel prize to Vernon Smith (1962, 1982), experiments about organizations—and *experimental organizations*—will provide a unique window into complex, microfoundational processes that are otherwise invisible to the organization theorist. Our article discusses the various ways in which the experimental turn will facilitate—and be facilitated by—a reorientation of organization theory.

The Gold Standard of Evidence

Experiments provide two qualities crucial to science: variation and control (Falk and Heckman 2009). As Louis Pasteur did (and John Stuart Mills observed), experimenters can determine the causal role of one variable by varying it, holding everything else constant, and observing the outcome. Across disciplines and theories, the prime value of experiments lies in determining the causal chain of events. Experiments excel in unambiguously separating correlation from causation.

Aside from determining causality, experiments also let a researcher generate original data under controlled conditions with exact measurements. Such an environment allows describing behavior precisely and comparing it to what theory predicts. It allows for identifying boundary conditions that surround a theory or collecting data to "see what will happen" (Smith 1982). Some of the most influential studies in the social sciences used this approach, like those conducted by the psychologists Amos Tversky and Daniel Kahneman and the economist Vernon Smith, which reshaped fields of judgment and decision-making, as well as behavioral and experimental economics.

A look back. As evident, experimental methodology has long roots in science, going back to at least the sixteenth century, when Galileo Galilei demonstrated experimentally that the speed of an object's fall is independent of its mass (Caffarelli 2009). Isaac Newton (1999 [1713]) spoke of "experimental philosophy" (and warned against feigning hypotheses). Albert Einstein saw the discovery of the possibility to find out causal relationships by "systematic experiment" as one of the two great achievements undergirding the development of Western science (the other being the invention of formal logical system; Cohen 1994, p. 234).

Given their scientific gravitas, experiments were embraced by early social scientists, like Friedrich Christoph Hegelmaier (1852), who studied the effect of time delays on perceptual judgmwent accuracy. With Wilhelm Wundt's founding of the first experimental laboratory at the University of

Leipzig in 1879, and Norma Triplett's (1898) experiments on the effect of competition on task performance, the experimental method gained prominence in psychology. Soon after, sociologists began to embrace experiments (Webster and Sell 2014; Zelditch 2014), of which the so-called Hawthorne experiments (Mayo 1933; Roethlisberger and Dickson 1939)—exposing factory workers to a range of stimuli such as different levels of lighting and measuring these workers' productivity—have become particularly well-known (and their findings still debated to the present day; see Levitt and List 2011). Around the same time, experimental markets appeared in economics (Chamberlin 1948). In the emerging field of organization theory, the Carnegie school brought experiments to the fore (Guetzkow and Bowes 1957; Cyert et al. 1961), but the method also began to flourish in other streams of organization theory (see Weick 1965; Weick 1969 for early reviews).

This brief historical survey makes clear that experiments have long been the leading scientific method. But only recently have they witnessed an impressive surge in organization theory, as evidenced by the tremendous response to our call for papers, the two special issue conferences, and a series of professional development workshops over the past decade. We are not surprised, considering all this method has to offer to the organization theorist. In what follows, we discuss some of the major advantages of the experimental method, followed by a discussion of what experimental methods offer specifically for organization theory..

Internal Validity. The concept of internal validity refers to "the confidence with which we can draw cause and effect conclusions from our research results" (Aronson et al. 1998, p. 129), or, in short, the ability to demonstrate causality. High internal validity is the *sine qua non* for empirical findings to be useful (Campbell and Stanley 1963). Internal validity is a necessary condition for external validity (Schwenk 1982). If a study cannot provide convincing evidence for causality, considerations of external validity are rendered irrelevant (Schram 2005; Lude and Prügl 2021).

Internal validity is crucial to organization theory, a field that stresses causal relationships and strives to guide organizational decision makers. Unless causal effects can be established, we cannot be certain of our theoretical models, nor can we demonstrate to managers that changing a particular parameter will truly result in desirable organizational outcomes (Galbraith 2012; Steinmetz et al. 2012; Milkman et al. 2021).

There is consensus that experiments offer the strongest causal inference and thus the highest degree of internal validity (Colquitt 2008; Breitsohl 2021). This is the primary reason experiments are considered the gold standard of scientific research (Podsakoff and Podsakoff 2019), whereas "for many important strategy research questions, however, traditional empirical techniques are not sufficient to establish causal effects with high confidence." (Chatterji et al. 2016, p. 116). Generally, non-experimental findings—even those involving sophisticated statistical methods—are correlational. Typically, such methods identify associations between variables, so attempts to establish causality necessitate unverifiable assumptions about identifying variables and relationships.[2] Such approaches, including those that attempt to leverage the benefits of "as if random assignment," are susceptible to criticism of their identification assumptions (List et al. 2011).

The challenge of identifying causality is met by the experimental method, with Its qualities of variation and control. With properly designed samples and randomization, we expect that non-manipulated differences cancel each other, allowing us to assume probabilistic equivalence between potentially-confounding variables and thus isolate true effects (Aronson et al. 1998). Experiments curb not only known confounds but also those a researcher may be unaware of (Bitektine et al. 2022). As result, the only difference between observations in the distinct conditions is in the causal variable, which means that differences in the dependent variable can be clearly attributed to this cause, precisely as Louis Pasteur concluded when he examined the flasks. Moreover, as the enhanced control over independent and confounding variables reduces random errors, experimental designs improve the likelihood that relationships between the independent and dependent variables are detected, reducing the false negative

[2] As Rohen Shah (2023) raps: "If we want to know if preschool affects your shot at college / Our task is not so easy, you have to acknowledge / We can correlate the two, with results that are resounding / But that's likely an error, and no, not from rounding / Correlations aren't causal: that evidence is compounding / The problem's other variables that might be confounding."

(type 1) error that can lead to the mistaken rejection of a valid hypothesis (Podsakoff and Podsakoff 2019).

As such, experimental research can effectively circumvent endogeneity concerns that are top of mind for organization theorists, including the biases of self-selection and omitted variables (Shaver 1998; Bitektine et al. 2018a; Hill et al. 2021) —that is, the possibility that what has been modeled as the dependent variable could be causing the independent variable or that some third variable might be correlated with the independent and the dependent variable, producing a spurious relationship (Aronson et al. 1998).

Construct validity. Another important type of validity, construct validity, denotes the extent to which an investigation truly captures its concept of interest. Construct validity thus questions the fit between theoretical construct and empirical measure (Brutus et al. 2013). Convincing organization theory requires clear constructs (Kerlinger and Lee 2000). Constructs are the fundamental ingredient to any organization theory, tying ideas to the empirical world and giving theories meaning (Cohen 1989, p. Ch. 7). Experiments shine in construct validity, especially when it comes to cognitive or affective variables, which are central to many strands of organization theory. Experiments offer a high degree of precision because they allow for observing relevant variables more directly. For instance, how can one measure March's (1991) theoretical concept of "exploration"? One possible answer involves attempting to identify archival proxies that indicate how organizations solve problems. Another approach is to measure search and problem solving directly, in real time, by instructing decision makers to engage in a search task and choose between exploiting known solutions or exploring new ones (e.g., Laureiro-Martínez et al. 2015; Levine et al. 2021a; Evans and Schilke Forthcoming).

As Louis Pasteur did, experimentalists build on each other's work. They can develop a comprehensive repertoire of manipulations and dependent measures for central constructs in organization theory, making significant improvements in their measurement. Importantly, experiments allow for assessing rather than assuming construct validity through such means as manipulation checks and psychometric analyses (Grégoire et al. 2019). Moreover, many experimental articles report multiple studies, which helps demonstrate construct validity by changing the operationalization of key constructs from one study to the next (Lykken 1968). Researchers can go a step further and design an entire experimental study with the sole purpose of evaluating construct validity, rather than testing a theoretical proposition. For example , this can be done by manipulating a given construct and examining the effect of this manipulation on a set of items intended to measure this construct (Podsakoff et al. 2013). Such designs can effectively disentangle the causal direction between latent construct and observable items that creates the distinction between formative and reflective measures (Bollen 1989; Edwards and Bagozzi 2000),[3] a distinction that is difficult to identify using other methods. In sum, experiments have the potential to make empirical measurement precise and thus enhance confidence in the testing of organizational theories.

Vividness. Experiments excel in testing theories and producing new ones. But they also possess a particular advantage for organizational theorists: Produce evocative results, findings that are easy to describe and communicate, conclusions that engage readers. This feature, which goes beyond the oft-mentioned benefits of rigor, can be crucial for creating a positive change in the world, an admirable disciplinary goal (Howard-Grenville et al. 2019; Tsui 2021). A thoughtfully designed experiment can convey a theory, articulate a prediction, send forth the findings, a prerequisite for bringing benefits to people, organizations, or society. Ideally, the experimental design will evidence the theory, prediction, and findings. Then, the experiment serves not only as a scientific method, but also tells a compelling story about the research, captivates the scientific community, decision-makers, and the public.

Consider the Louis Pasteur illustration in our opening, a canonical experiment cited in science textbooks and replicated by high-school students. With just a paragraph of description, a reader immediately grasps how two theories competed—and which one won. In modern behavioral sciences, recall the elegant simplicity of the three-sentence vignette conjured by Amos Tversky and Daniel

[3] In reflective measurement models, items are assumed to be caused by (and thus a consequence of) the latent variable; in formative measurement models, items are assumed to cause (and thus be antecedents to) the latent variable (Diamantopoulos and Winklhofer 2001).

Kahneman (1982, p. 92, 1983) to demonstrate how people predictably fail to follow the rules of probability. They called it the conjunction fallacy, but the design was so evocative that most people know it simply as the Linda problem: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable? (1) Linda is a bank teller; (2) Linda is a bank teller and is active in the feminist movement." Once you read it, you intuitively grasp it. No matter how many times you repeat it, the result stubbornly contradicts what you ought to know about statistics. As Stephan J. Gould (1988), the evolutionary biologist and historian of science, commented, "I am particularly fond of [the Linda problem] because I know that the [conjoint] statement is least probable, yet a little homunculus in my head continues to jump up and down, shouting at me—'but she can't just be a bank teller; read the description.'"

For all of these reasons, we concur with Nobel Laurate Guido Imbens (2010, p. 407), a luminary of econometrics and statistics, when he declares: "experiments do occupy a special place in the hierarchy of evidence, namely at the very top."

An Ideal Method for Organization Theory

By now, it is clear why experiments have become scientific workhorses, whether in physics or biology, engineering or nutrition, psychology or economics. But experiments also bear some unique features that make them particularly appealing to students of organizations, at this point in the theory's arc of progress. As sociologist and organization theorist Richard Scott once commented to one of us (SSL), "there are hard sciences—and there are harder sciences." We believe the study of organizations belongs in the latter group. Researchers of organizations arguably face a more daunting task than do many natural scientists. Organizational features can rarely be isolated and manipulated as Louis Pasteur did with his identical flasks. Firms are complex. They are composed of people, who are organized in teams and hierarchies. The entire company is subject to industry dynamics and macroeconomic forces, is embedded in a geography, and exists at a specific point in time. But experiments can help. [4]

Dissect organizational complexity. Complexity is what drew many of us to organization theory but is also what encumbers our empirical work. In much of our research, we want to separate the effect of one organizational facet from that of all others. But in naturally occurring settings, many of these facets cling and cluster, making isolation difficult—often impossible. In contrast, experiments allow for varying these features separately and in an orthogonal fashion, helping us dissect complex organizational phenomena into their constituent parts.

This unique ability to divide complexity comes with three key benefits for organizational theorists. The most straightforward benefit relates to obtaining more precise estimates of causal effects, as alluded to in our earlier discussion of internal validity. Second, and relatedly, experiments offer the opportunity to "single out some issues that have gone unnoticed in the complexity of actual organizations" (Weick 1969, p. 294). Through the experimentalist's effort to dig deep enough to reach the root of a phenomenon, what was previously unnoticed is now discovered. Third, having separated distinct organizational causes, organizational scholars can reinfuse complexity into their investigation by examining how combinations of several independent variables produce particular effects. We always advocate starting with a simple design, as Louis Pasteur did, but experiments in organization theory can also accommodate complicated factorial designs (Gunst and Mason 2009). For instance, Lucas (2003a simultaneously manipulated a leader's gender (male vs. female), the reasoning for appointing the leader (appointed on ability vs. randomly assigned), and the institutionalization of female leadership in the organization (legitimated vs. experimental control). He hypothesized about the effects of theoretically meaningful combinations of these three factors on participants' acceptance of the leader's influence. Treating these three independent variables as orthogonal through systematic experimental manipulation

[4] The emergence of a community dedicated to Experimental Organization Science is heartening (https://www.experimentalorganization.science/)

enabled Lucas to tease out main and interaction effects, which naturally occurring covariation might have made impossible to detect.

Organizations are complex not just because of what they contain, but also because of the environment in which they operate. Here, again, experiments can help clarify. Typically, scholars take the perspective of the actor — the manager, the entrepreneur, the firm. But experiments can also be used to track the view of the audience, illuminating external social evaluations of organization, a growing stream of research in organization theory (Bitektine 2011; Pollock et al. 2019)

Discover mechanisms and processes. Experiments have long been tools of discovery: "a pretty experiment is in itself often more valuable than twenty formulae extracted from our minds," quipped Albert Einstein (quoted in Moszkowski 1970, p. 69). Herbert Simon (1989) and Allen Newell (1972) used experiments to uncover human cognitive functions, sometimes together with think-aloud protocol analysis. Vernon Smith (1982) created experimental markets (and later experimental worlds) to examine how institutions affect price, trade, and specialization, among others

Experiments can surface novel phenomena, but their ability to disentangle complex relationships also makes them powerful in unpacking mechanisms. Mechanisms are the causal reasons why an independent variable affects the dependent variable. Quasi-synonyms include underlying processes, intervening variables, and causal explanations. Many researchers in our field, such as Sutton and Staw (1995, p. 378), equate good theory with a solid grasp of underlying processes, putting mechanisms at the very heart of what separates theory from "dustbowl empiricism" (Mathieu et al. 2008, p. 203) and what distinguishes scholarship in organization theory (Davis and Marquis 2005).[5]

[5] Like the proverbial Russian Doll, mechanisms exist at multiple levels, often nested within each other. For instance, one might argue that relational capital is useful to explain risk taking, but that would hardly satisfy a social psychologist, who might want to measure trust. Still, such a measurement may fail to impress a neuroscientist or an evolutionary psychologist, who in turn would ask about the brain pathways or natural selection processes that give rise to relational capital or trust. In other words, one's "mechanism" can be another's "black box". Organizational scholars may benefit from biologist Ernst Mayr's (1961) influential distinction between causes that are proximate (driven by immediate factors) and ultimate (driven by evolution).

The experimental method is uniquely suited to advancing mechanism-focused organization theory for two reasons. First, experiments are conducive to a large variety of data collection techniques, such as interviews, behavioral data, neuropsychological measures, content analysis, and protocol analysis (Colquitt 2008; Laureiro-Martínez et al. 2010; Reypens and Levine 2017, 2018). This openness to distinct forms of data collection allows access to a variety of intraorganizational or psychological processes, which are at the core of contemporary organization theory — yet rarely captured by other methods. Second, experiments are highly compatible with analytic approaches to establishing a causal mechanism by providing evidence for the mediating role of some variables (MacKinnon et al. 2007). Experiments not only can accommodate the measurement-of-mediation logic that Baron and Kenny (1986) made famous, but also are able to augment such correlational approaches with designs that manipulate the mechanism. Because if you can manipulate the mechanism, not just measure it, you can determine causality (Spencer et al. 2005). Their unique capacity to model mechanisms also makes experiments prime candidates for testing and reconciling competing organizational theories, providing insights into *why* one theory predicts better than another (Croson et al. 2007).

Study microfoundations. Mechanisms may be most central in the flourishing study of microfoundations, a movement that has now permeated virtually all strands of modern organization theory (Felin et al. 2015; Cowen et al. 2022). Like no other method, experiments help identify all three types of multi-level mechanisms that together form the scaffolding of microfoundational inquiry in organization theory (Foss and Pedersen 2016; Haack et al. 2019; Bitektine et al. 2022). These include situational mechanisms (macro→micro), action-formation mechanisms (micro→micro), and transformational mechanisms (micro→macro) (Coleman 1990; Hedström and Swedberg 1998).

Much experimental research in sociology has demonstrated the situational effect of social structure on individuals (e.g., Cook et al. 1983), while experiments on judgment and decision making have shed light on important action-formation mechanisms (e.g., Tversky and Kahneman 1992), and jury experiments in political science have long investigated transformational mechanisms that shape collective outcomes (e.g., Guarnaschelli et al. 2000). In economics and strategy, researchers experimented to examine how organizational values affect the efficacy of managerial practices (Blader et al. 2019) and how managers' pay-for-performance affect workers (Bandiera et al. 2007). They have used the method to understand the value of information about a firm's competitors (Kim 2020) or why woman-owned businesses earn less than those owned by men (Delecourt and Ng 2021), among others.

Inspired by such disciplinary contributions, organizational scholars have started to devise experimental microfoundations studies, such as investigations of the effects of organizational identity on decision makers' attention, uncertainty, and mimetic behavior (Schilke 2018); founders' demographic composition on workers' effort (Kacperczyk et al. 2023); voting rules on aggregation of organizational decisions (Piezunka and Schilke Forthcoming); and individual cognition and behavior on market outcomes (Levine et al. 2014; Levine et al. 2017; Park et al. 2022; Levine and Zajac 2023).

Organization theory is not a pure macro-level discipline. As organizational theories begin to place greater emphasis on the micro level. this not only implies attention to cognitive and affective mechanisms, but also to the ways in which organizations aggregate individual behaviors (e.g., Puranam 2018; Klingebiel 2022; Piezunka and Schilke 2023). Either way, experiments will only continue to grow in their importance (Bolinger et al. 2022) no longer just appropriate; they are indispensable for organization theory to advance. They are the logical next step in the coevolution of organization theory and methods (Haack et al. 2021).

Improve external validity in organization research. As graduate students sitting in a research method class, some of us may have heard that experiments excel in identifying causality, but their "contrived and created settings" (McGrath 1981, p. 73) may not generalize (Jackson and Cox 2013). Yet, this dichotomy does not necessarily hold in organization research, where complex phenomena are joined by incomplete data to undermine external validity. Especially in organization science, experiments can play a crucial role in boosting external validity, the extent to which a causal relationship identified is generalizable.

First, the challenge: When studying organizations, data are often limited. Organization theorists rarely have access to data spanning the entire population of interest. Comprehensive samples are rare.

Databases can be narrow or fragmented. Information is typically available for a set of companies in a specific industry, at an idiosyncratic geographical location, in a particular timeframe. Moreover, databases can suffer from survivor bias, covering only companies that reached a certain lifespan or size (Yang and Aldrich 2012). As scholars readily acknowledge, the generalizability of findings from readily available data is questionable. An admirable effort to assess the generalizability of findings from archival data estimated that only about half held across timespans and places (Delios et al. 2022). Other scholars, using a variety of methods, produced similarly disheartening results (Goldfarb and King 2016; Bergh et al. 2017). (Encouragingly, Delios et al. also show that when the original findings are statistically reliable, they are more likely to reproduce).

The challenge of generalizability balloons when studying organizational practices. Available data may overrepresent what organizations have allowed or promoted. An examination of R&D projects likely overlooks those rejected by corporate decision-makers (Criscuolo et al. 2017; Klingebiel 2022). An analysis of patents granted shows only inventions filed and approved but misses those protected in a different, perhaps superior way (Levin et al. 1987; Criscuolo et al. 2019). Data typically show organizational practices that were adopted, but not those there were reassessed and rejected (Kraatz and Zajac 1996; Klingebiel 2018). Such sampling can produce biased results. Even if they hold in the settings examined, they may not generalize to other organizations (Denrell and Kovács 2015; Naumovska et al. 2021).

Limitations of data and sampling techniques constitute a severe challenge to organization research: "Most research on organizations had collected data on easy-to-observe organizations, generally the largest and most dominant. Such designs cause serious problems for studies that seek to understand the causal processes responsible for success and failure," observes Michael Hannan (2022, p. 3), a founding father of the field.

How can experiments improve the external validity of organizational scholarship? In experiments, a researcher collects original, primary data (Falk and Heckman 2009; List 2011). Thus, the scholar may overcome some of the sampling bias that may lurk in existing data. By carefully designing the sampling technique and considering the target population, the researcher can design an experiment whose findings generalize to theories that hold beyond the sample of participants (and non-participants; Banerjee and Duflo 2009). For instance, in studying startups, the scholar can recruit potential entrepreneurs, not just the successful ones (Camuffo et al. 2020). If examining behavioral differences between the sexes, an experimenter can construct a gender-balanced, nationally representative sample (Levine et al. 2021b). And if aiming to improve the workplace climate in large corporations, they can experiment in employees in such organizations (Alan et al. 2023).

Experiments can sidestep the problems besieging archival and observational studies. Their samples can minimize biases and their results provide decisive causal evidence. To achieve that, obviously, experiments must be well designed and executed. Our recommendations below provide some countermeasures to such ills that can infect even the best-intentioned researcher.

Furthermore, external validity is bolstered when we know not only *whether* but also *why* something works. Such knowledge of mechanisms can support sensible predictions of how a given manipulation will perform outside the initial experimental context (Baldassarri and Abascal 2017). As discussed, experiments have a unique capacity to identify relevant mechanisms and thus add substantial insight into the specific contexts in which a theorized effect is likely to occur.

Improve reproducibility of findings. Experiments can substantially enhance confidence in research findings because they are ideal for replication. If we observe the same effect under identical conditions across distinct studies, we can be more certain that the outcome is not haphazard but, rather, is caused by the hypothesized predictor (Lucas 2003b; Haslam and McGarty 2004). Put differently, the generalizability of a finding is not established in a single study, regardless of the methodology (Lynch 1982). Every empirical investigation is conducted in a specific setting, at a specific point in time, and among specific actors. Whether results from a given study can be extrapolated to other settings remains to be determined (Lonati et al. 2018). As such, generalizability can only be achieved through replication of results across different populations and contexts (Dipboye 1990; McDermott 2011).

In principle, all research can (and should) be replicated, including qualitative studies (Aguinis and Solarino 2019). But experiments lend themselves to more exact replications thanks to their minimal procedures and clearly structured protocols, which, as in Louis Pasteur's experiment, enable sharing and reconstruction of designs, facilitating the independent verification of findings (Croson et al. 2007). Because experiments are relatively easy to replicate, they lend themselves to probing robustness and thus external validity. Moreover, several meta-analyses suggest that experiments tend to produce results that are highly aligned with those of other methods (see Mitchell 2012; Vanhove and Harms 2015).

The simplicity of sharing materials and procedures is epitomized in the massive collaborations that have been testing the reproducibility of findings in the social sciences at large (Camerer et al. 2018), psychology (e.g., Open Science Collaboration 2015; Ebersole et al. 2016; Klein et al. 2018), economics (Camerer et al. 2016), and most recently — of landmark experiments published in *Management Science* (Davis et al. 2023). This replicability gave rise to meta-studies: massive experiments run by dozens of researchers, where many interventions are compared for the same outcomes in the same population and duration (Milkman et al. 2021).

In contrast, non-experimental methods often cannot afford a comparable degree of control over various industry- or organization-specific factors that affect research findings (Reypens and Levine 2017). Although replication has historically not been highly incentivized (Rubinstein 2001; Ryan and Tipu 2022), the replicability crisis in psychology (Maxwell et al. 2015; Shrout and Rodgers 2018) and its spillovers to organizational scholarship (Lewin et al. 2016; Pratt et al. 2020) have put replication at the top of the research agenda across the social sciences (Lewin et al. 2016; Freese and Peterson 2017; Camerer et al. 2019; Davis et al. 2023), giving a privileged position to experimental methodology and its ability to afford replication and collaboration. The collective response promises to enrich the scientific dialogue and bring us closer to the truth.

Even though replicability concerns originated in domains dominated by experimental methods, it was only thanks to experiments' high degree of verifiability and repeatability that these concerns could come to light in the first place. In their comprehensive replication of economic research, Camerer et al.

(2016, p. 1435) conclude "that laboratory experiments are at least as robust, and perhaps more robust, than other kinds of empirical economics." Given that science is necessarily a cumulative effort, we should favor methods that promote the substantiation of extant findings over those that make it prohibitively costly or impossible (Agarwal et al. 2010; Lewin et al. 2016; Christensen et al. 2019).

Uncover understudied topics. Another key benefit of experimental methods is that they "allow access to invisible domains of human experience" (Aviram 2012, p. 463). That is, they make possible the study of interesting topics on which other types of data are difficult to obtain (Schwenk 1982). This is because experiments can not only observe but also create the subject of inquiry. Going back to our opening quote from Mill's *A System of Logic*, an experiment can produce variations that are crucial to theory development. This capability is particularly important in organization theory because "any non-experimental science…can study only the systems that evolve," as Michael Hannan (2022, p. 23) readily acknowledges.

At least four types of relevant issues remain understudied in organization theory for reasons of data availability, where we believe experiments can help: (1) sensitive, (2) infrequent, (3) invariant, and (4) emergent topics. We begin with sensitive topics, like organizational misconduct, scandal, and discrimination, that are notoriously difficult to study with observational approaches. By definition, sensitive topics can pose a grave threat to those with deviant positions or behavior, and thus research participants may be unable or unwilling to report on them with accuracy (King et al. 2013). Even if one manages to recruit participants, social desirability bias may curtail honest responses. Experimental methods help address these problems—for instance, by using hypothetical scenarios or disguised measures (Weber 1992; Aronson et al. 1998).

Organizational events of great importance are usually infrequent. Companies do not often enter new markets, get acquired, or their political stance (Lampel et al. 2009; Beamish and Hasse 2022). Managers do not suddenly acquire strategic intelligence (Levine et al. 2017). Entrepreneurs do not often boost their social skills. These rare events, difficult to observe *in situ,* can mean the difference between organizational death and survival. Further complicating matters, such rare events tend to be idiosyncratic (no two political messages are identical) — and endogenous (social skills and strategic intelligence may be correlated with other desirable qualities). Therefore, they are difficult to compare. But experimentalists do not have to wait for these events to happen. They can create them: train entrepreneurs to expand their social networks (Dimitriadis and Koning 2022), focus business owners' attention on competitors (Kim 2020), alter a firm's political message (Burbano 2021). Experiments allow us to create events that are identical except for some crucial differences which are the subject of study.

Further, some phenomena simply lack variation in natural settings. Consider, for example, how isomorphic forces may have caused most organizations in a field to adopt the same solution (DiMaggio and Powell 1983). Alternative solutions may have been considered but suppressed or may have simply died out (or left out of the sample; as discussed above). With insufficient variation, it is impossible to detect covariation with other variables of interest. Without heterogeneity, typical statistical methods are rendered useless. Experiments may be able to extend the range of a variable to produce greater variation beyond what would be observable in natural environments. This was the precise purpose of an early organizational experiment (Guetzkow and Bowes 1957).

Finally, some phenomena are emergent; that is, they are not yet widely adopted in the field. An example is blockchains: Organization theorists show great interest in exploring the implications of these technologies for organizational design but lack archival data to study them. One solution is to go to the laboratory and isolate theoretically meaningful features of blockchain-based organizations, such as certain types of voting procedures or smart contracts (Wang et al. 2022), and examine their ramifications.

Overall, well-designed experiments have the critical advantage that they enable researchers to produce their own exogenous variation in theoretically relevant issues—no matter how sensitive, infrequent, invariant, or emergent they are in natural settings—and study the effects of such variation in unobtrusive ways. As a result, experiments allow for genuinely question-driven research and free investigators from the constraint of narrowing their interests to only those issues for which they can find existing data, as recent entrepreneurial experiments exemplify (Chatterji et al. 2016; Nai et al. 2022; Kotha et al. 2023).

Best Practices in Conducting Experiments

Given the importance of experiments for organization theory, two related questions remain: What are the best practices in conducting experiments? And what are the key criteria for evaluating experimental research? Seeking answers to these questions, we began our exploration by probing scholars' perceptions about practices in experimental research. In late 2022, we surveyed the participants of the Organization Science Special Issue Conference about a range of experimental practices. The survey was voluntary and anonymous (to ensure confidentiality, we avoided collecting demographic information, like gender and age). We received valid responses from 60 scholars (61 for some questions) present at the conference. Most respondents (56.67%) were not special issue authors. At the time, they had held a PhD for an average of 7.35 years (*SD* = 6.95; range: 0-28 years).

The survey questions probed how certain research practices may increase confidence in the findings presented in a study. For instance, the first question asked about the public availability of the data set used in the study. The responses were all provided on a five-point Likert scale capturing the degree to which a practice would increase confidence about the findings and ranging from 1 ("definitely not") to 5 ("definitely"). We probed survey respondents about a total of seventeen practices (Table 1).

Table 1: List of practices and reported mean scores (emphases in source)

Item	**Mean score**
If the authors include any of the following, are you more confident about the findings presented?	
The **data set** is publicly available	4.13
The **statistical code** is publicly available	4.31
The **instruments** used are publicly available (e.g., questionnaire)	4.46
Clear separation between (confirmatory) **hypothesis testing** and (exploratory) **post-hoc analysis**	4.43
Description of all **variables** (measures) for which data were collected, even if some were not used in analysis	4.25
Description of all **experimental conditions** (manipulations), even if some were not used in analysis	4.30
Disclosure of any **data exclusions**	4.33
Discussion of how the **sample size** was determined, including analysis of statistical power	4.23
Full **description** of variables (e.g., means, SD, CI)	4.45

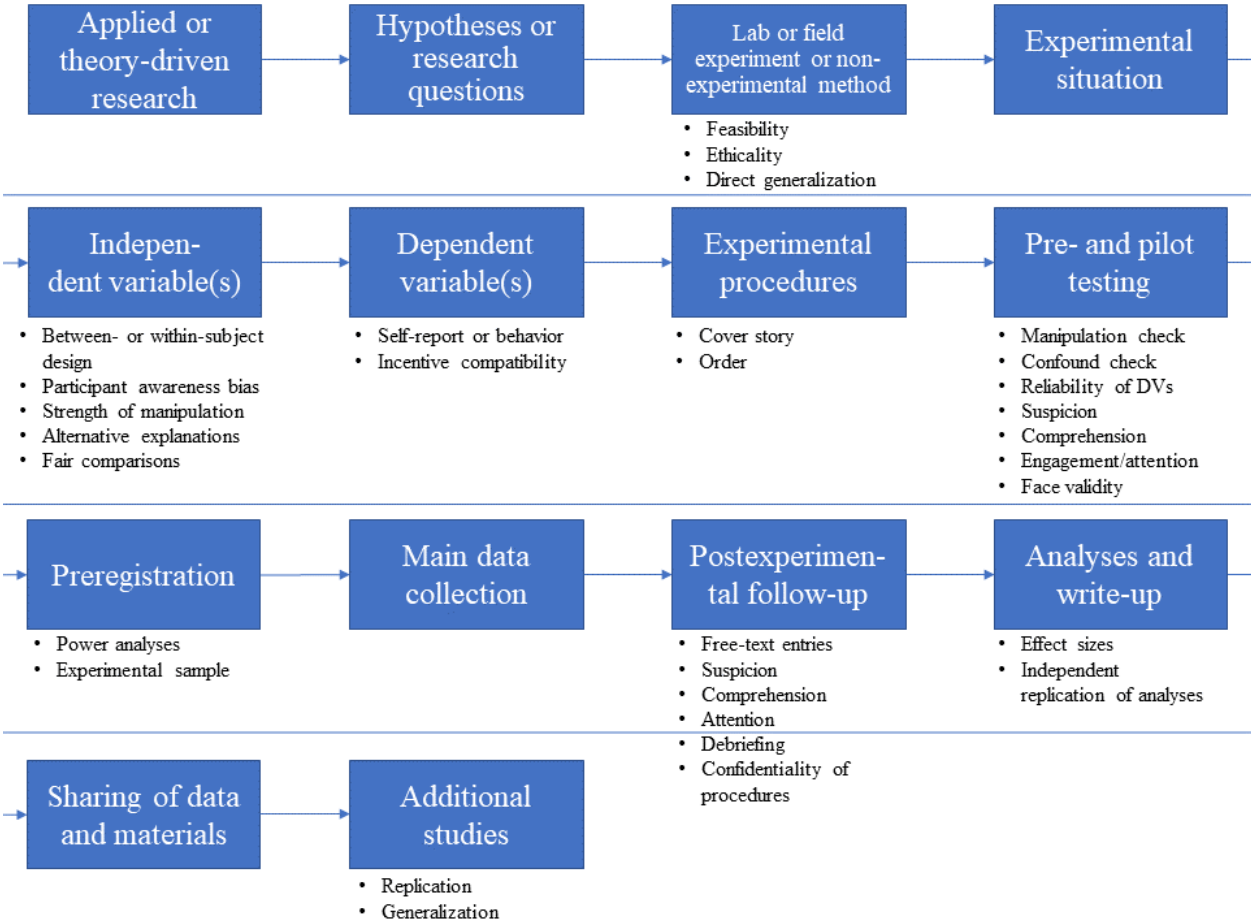
Visualizations of disaggregated raw data, e.g., scatterplot matrixes, binned scatterplots, residual plots, Cook's distance	3.80
The statistical test(s) used, and whether they are **one- or two-sided**	4.45
Exact p-values rather than stating that a finding is "significant"	4.25
Estimates of **effect sizes** (e.g., Cohen's d, Pearson's r)	4.32
Preregistered list of **hypotheses** or questions that motivate the study	4.00
Preregistered sampling plan, which specifies the sampling method, the sample size, and principles for removing data	4.05
Preregistered measurement plan, which details the measures used and how they will be coded	4.00
Preregistered analysis plan, which lists the measures (e.g., coefficients) that will be estimated and how they will be interpreted	3.87

Several conclusions emerged from the survey. First, all practices listed above received an average rating that was higher than the midpoint, 3 ("might or might not"). The median score was the maximum (5.00) for eight of the practices, suggesting the community considered these practices as very important in affecting one's confidence in empirical findings. In other words, there was little controversy about these practices overall.[6]

Below, we discuss in greater detail a range of best practices in experimental research. First, we synthesize the most important steps in an ideal-typical process diagram (see Figure 1). Second, we highlight some practices that we observed to be featured prominently in peer-review processes (see checklist in the appendix).

[6] In a correlational post-hoc analysis, we estimated OLS regressions to examine how the endorsement of experimental practices may have varied as a function of participant characteristics. Two interesting conclusions emerged. First, we found that respondents who were *not* special issue authors placed more value on a clear separation between (confirmatory) hypothesis testing and (exploratory) post-hoc analysis (effect: +0.22; $p<0.035$). Second, those who had received their PhD earlier (i.e., were more experienced) gained slightly more confidence from the specification of the statistical tests used (effect: +0.03; $p<0.039$) and the provision of the exact *p*-value (effect: +0.04; $p<0.027$). The materials, analyses, and data are w available https://osf.io/8duag/?view_only=6baf093381824550a2cb2b1e63589612

Figure 1: Simplified flow chart of experimental design



Note: Feedback loops omitted.

Simplify, simplify, simplify. What Albert Einstein advocated as the "supreme goal of all theory" applies to experiments, too: "make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience" (van Dongen 2010).

The simplest experiment is often the most compelling: it is easier to implement, easier to communicate, and easier to reproduce (remember the Linda problem, above). But, somewhat ironically, minimizing complexity in experimental research can be difficult: it requires careful, upfront planning and the ability to balance simplicity with the quest to unpack complex social phenomena. In addition, despite the clear benefits, some scholars may be reluctant to commit to a simplified design, anticipating a risk of losing scientific rigor or undermining ecological validity.

How can we navigate such trade-offs in experimental designs? A few rules apply. First, scholars may want to identify measures that not only map tightly onto the key variables of interest but also are easy to administer and interpret. When the stimuli are simple and precise, it is easier to rule out alternative explanations. Rather than creating original instruments, which may require extensive validation, it is often preferrable to use already-validated measures that have proven reliable and credible based on exiting research (Crano et al. 2014). To mitigate complexity, it is also useful to draw on prior research to guide an experimental design—including not only the choice of measures but also that of experimental procedures. For example, an audit or correspondence study, whereby fictitious resumes are sent out to real jobs, represents a common experimental design used in evaluating employment opportunities available to job seekers in the labor market (Pedulla 2016; Rivera and Tilcsik 2016; Kacperczyk and Younkin 2022a). While the development and application of audit studies represents a collaborative effort among researchers over many years, the methodology has been standardized and widely adopted, allowing for an easily replicable body of research.

Second, before launching the experiment, it is often instructive to conduct a pilot test of the research instrument and identify potential ambiguities, challenges, and unexpected complexities (Baker 1994 , p. 182-183). Such pilot tests can give scholars an advance warning of potential issues, indicating whether the proposed methods or instruments are appropriate or whether it is necessary to refine the design, the procedures, or the materials before conducting the main study (Wilson et al. 2010). But pilot testing can also serve another goal. Running a pilot of the entire experiment can help researchers evaluate the study's feasibility by providing an opportunity to do a trial run "in preparation for the major study" (Polit and Beck 2001, p. 467). As De Vaus (1993, p. 54) warns, "Do not take the risk. Pilot test first." Indeed, piloting a study may be especially valuable for field experiments, which rely on implementing specific interventions in real-world settings. By conducting pilot tests, researchers can test their interventions on a smaller scale to gauge their effectiveness, feasibility, and potential impact early on before scaling up. An opportunity to rectify shortcomings and bottlenecks early on via pilot studies can thus help minimize the risk and cost of deploying large-scale experiments in the real world (Perdue and

Summers 1986; Grégoire et al. 2019). And pilot data can be effectively used to conduct power analyses that are based on sensible assumptions.

Overall, simple experiments are more efficient, are easier to interpret, and can still effectively address complex research questions. While designing simplified experiments requires careful attention and time on the part of researchers, the long-term payoff in terms of the ease of interpretation, communicability, and execution can be considerable.

Replicate and support others in replicating your work. Replication is fundamental to reliable, credible experiments (Bettis et al. 2016; Lewin et al. 2016; Antonakis 2017; Bergh et al. 2017; Bamberger 2019; Davis et al. 2023). Researchers can take a range of steps to facilitate replication. Perhaps first and foremost, it is useful to maintain methodological transparency by clearly documenting the experimental protocol, including procedures, materials, and analyses (some specific recommendations appear in the appendix checklist).

Standardized methods, instruments, and procedures help minimize variability across replications and thus are often preferrable to new, untested procedures and instruments. Standardization can be especially valuable when studying organizations because organizations are, by default, idiosyncratic, unique, and difficult to compare. In other words, in studying organizations, experimental approaches would benefit from greater standardization of procedures and measures, including more tests of convergent and discriminant validity.

But even detailed descriptions of the experimental protocol in the study itself or prior to data collection may not suffice to guide researchers toward exact replication. Typically, data processing and analysis involve a "garden of forking paths": judgement calls or specifications in the code (Gelman and Loken 2013). These can make replications difficult, unless both the data and the code are publicly shared. Therefore, replicability is critically contingent on scholars' willingness to share the study's materials, including the provision of data and detailed annotations of the steps followed in the analyses.

Finally, replicability can be facilitated by agreeing to collaborations with scholars interested in expanding the paper's findings in novel and interesting ways. At the Special Issue Conference, our objective was to foster a collaborative environment, by encouraging authors to share their findings with others in the community. Through this collaborative exchange, we hoped to inspire future cooperative endeavors. In short, by following a few simple steps—such as using standardized procedures and instruments, preregistering the study's data collection plan, publicly sharing the data, and collaborating—researchers can enhance the replicability of their studies and contribute to a more transparent and robust community.

Use students, executives, or the general population. Experimental approaches often tap a wide range of participants, resulting in diverse (Brandon et al. 2013) and generalizable samples (Colquitt 2008), as we discussed above. Yet, for any given study, the critical question remains: How should we select participants? The choice of participants depends on the purpose of the study. Far too often, researchers consider managers (or entrepreneurs) as the default option. But selecting practitioners may undermine external validity if the study aims to investigate behaviors not specific to a managerial setting. In addition, recruiting practitioners as participants has the drawbacks of taking a long time and being expensive. As a result, studies drawing from managerial populations risk suffering from small samples and low statistical power.

As a viable alternative, researchers may want to consider students, especially considering the clear benefits of lower data collection costs and larger sample sizes. Although student samples are occasionally criticized on the basis of reduced generalizability, they can be appropriate for exploring a wide range of organizational questions, including questions that relate to cognition and psychological processes relevant to the general population (Stevenson and Josefy 2019). The use of student participants can also be justified based on conceptual grounds, considering that they will soon occupy organizational and entrepreneurial positions (and many already do, even throughout their studies).

Economists devoted much effort to explicitly comparing the choices of various participant populations in theoretically important settings (e.g., Cooper et al. 1999; Bolton et al. 2012). Recent reviews of this extensive literature find little systematic differences in the behaviors of managers as compared to students (Fréchette 2015, 2016). With this evidence in mind, a student sample may constitute a valid and high-powered sample while reducing the time and cost of research, ultimately enabling more studies.

Finally, participants can be sourced from the general population, including participants available through online platforms, such as Amazon Mechanical Turk and its ilk (Reypens and Levine 2017). Various studies indicate that participants recruited via such platforms are highly representative of the general population (Buhrmester et al. 2011; Weinberg et al. 2014; Palan and Schitter 2018). In summary, the decision of what type of experimental participants to use depends on the research question, objectives regarding generalization of results, resource constraints, time available to collect the data, ecological validity concerns, and ethical considerations.

Choose to measure behavior rather than self-reports. Whenever it is theoretically appropriate and empirically feasible, measuring behavior might be preferable to measuring self-reported intentions. Intentions reflect motivation to engage in certain behavior, but they do not always translate into actual behavior (Baumeister et al. 2007). For example, individuals may self-report the intention to foster innovation, but they could be unable to follow through with this intention because of external constraints such as the time or budget needed to develop a new product. In addition, intentions could be subject to social desirability bias (Edwards 1957) and demand characteristics (Orne 1962), such that participants may default on reporting intentions that are more socially acceptable or meant to satisfy (what participants believe to be) the researcher’s expectation.

Behavioral measures not only yield potentially more reliable and less biased data but also can be particularly useful in evaluating the effectiveness of interventions and treatments. For example, when assessing how to de-bias investors’ evaluations of women founders (e.g., Thébaud 2010; Guzman and Kacperczyk 2019) or learning from minority peers (Levine et al. 2021b), it is useful to measure choices directly (Kanze et al. 2018), as opposed to relying on self-reported intentions. But what should we do when behavioral measures remain out of reach? For example, in any employee recruitment setting, observing how employers evaluate job seekers can prove challenging. In some cases, such hurdles may be overcome by using less direct proxies for behavior or providing a strong justification for why intention can predict behavior in the given context. Indeed, scholars recognize that measuring intention is often useful in cases where behavior is difficult to detect (Di Stefano and Gutierrez 2019). In short, whereas scholars have often resorted to measuring self-reported intentions, experimental methods open the door to measuring actual behavior (Schilke et al. Forthcoming).

Ensure incentive compatibility. An experiment must reveal individuals' true preferences, behaviors, or reactions (Lonati et al. 2018), but incentives — whether withheld or provided, weak or strong — can unintentionally bias the responses of experimental participants. A good experimental design will assure that incentive do not bias responses, a feature economists call "incentive compatibility" (Azrieli et al. 2018). Similarly, psychologists use the term "experimental realism" to refer to conditions that serve to make a study meaningful and engaging to participants, eliciting responses that are spontaneous and natural (Aronson et al. 1990).

How can we achieve incentive compatibility in an experimental design? The answer depends on the purpose of the study and the characteristics of the experiment itself (Ariely and Norton 2007). Many use pecuniary rewards, rewarding participants with compensation according to their performance or the accuracy of their responses (Gneezy and List 2013). Such features of the experimental design help align incentives, but they arrive with a host of considerations, both experimental and ethical (Camerer and Hogarth 1999; Holt and Laury 2002, 2005; Levine and Zajac 2023).

Another approach to elicit unbiased behavior involves deception. For example, our understanding of employer bias during recruitment would not have progressed as far as it has without the extensive use of correspondence or resume-based audit studies. Such studies rely on submitting fictional resumes to real jobs in order to elicit employer callback for an interview, a measure of real behavior (Tilcsik 2010; Kacperczyk and Younkin 2022b). Of course, ethical issues and informed consent requirements must be carefully considered when using deception in experiments (Breitsohl 2021).

Finally, while there are common ethical principles and guidelines that govern research worldwide, the specific standards and processes of ethics committees can vary between different universities or institutions. For example, whether deception is considered harmful may depend on institutional policies, regulatory environment, or even the composition and expertise of individual committee members. In short, eliciting unbiased behavior may require paying close attention to incentive compatibility as well as ethical considerations in cases where deception is being used.

Preregister and separate your hypothesis tests from post-hoc analyses. In general, it is recommended, and sometimes required, to preregister. As detailed in our checklist (appendix), preregistration (or pre-specification) refers to specifying the techniques and analyses for confirmatory or deductive research studies before carrying out the actual research. During preregistration, researchers identify various aspects of their study, including the characteristics of the sample, the sample size, manipulations, measures, and other analytical choices. Specifically, a researcher may list (1) hypotheses or questions that motivate the study; (2) a sampling plan, which specifies the sampling method, the sample size, and principles for removing data; (3) a measurement plan, which details the measures used and how they will be coded; and/or (4) an analysis plan, which lists the measures (e.g., coefficients) that will be estimated and how they will be interpreted (van 't Veer and Giner-Sorolla 2016).

Preregistration has several benefits for researchers, their readers, and the field. By stating upfront what we expect to find, the practice guards us from overinterpreting unexpected pattern in the data and thus potentially being fooled by false positives (Moore 2016). Second, preregistration reduces the temptation of fishing for statistical significance (Forstmeier et al. 2017), what Bettis (2012) pejoratively calls "the search for asterisks" and Antonakis (2017) dubbed a "disease of *significosis*". As we know today, coefficients with an asterisk (or even three) are not necessarily true or interesting (Gelman and Stern 2006), but searching for them compromises our theories, which are never truly tested, just evoked once the findings are already known (Starbuck 2016). It is not a coincidence that leading scholars warn: "'statistically significant'—don't say it and don't use it" (Wasserstein et al. 2019, p. 2).

Preregistration has been embraced across the social sciences because it increases the transparency and reproducibility of findings (Munafò et al. 2017; Davis et al. 2023). Perhaps most importantly, the practice clarifies to the readers which elements of the study are deductive or hypothesis-driven — and which are exploratory, initiated after seeing the data (Nosek et al. 2018). When researchers preregister their studies, they establish what part of a study is pre-planned as opposed to a post-hoc finding. The preregistration record guards against accusations of *p*-hacking (Simonsohn et al. 2015) or selective publication of findings (Kasy 2021). A study of 15,992 randomized control trials published in 15 leading economics journals found that "preregistered studies that have a complete pre-analysis plans are significantly less *p*-hacked" (Brodeur et al. 2022, p. 3). (We suspect that experiments are generally less susceptible to p-hacking than other methods, simply because experiments are typically designed for a purpose, like testing specific hypotheses or exploring a process. In contrast, other methods rely on extensive data sets. The presence of numerous variables can tempt a researcher to try myriad tests. This could result, even if inadvertently, in p-hacking or HARKing (Hypothesizing After the Results are Known; Kerr 1998)).

Preregistration comes in handy when the results contradict expectations. Rather than struggle to explain contradictory or null results — peer comparison does not motivate but instead harms (Reiff et al. 2022); womens' ideas are not judged worse than those of men (Dahlander et al. Forthcoming) — the researchers can point to the preregistration, which testifies to the their sincere effort in conducting a sound theory test. When theory-anchored hypotheses are contradicted, the results may still be worth publishing, notwithstanding the known publication bias (Begg and Berlin 1988; Lewin et al. 2016). Some outlets now take the practice a step further by offering registered reports, an advance guarantee of publication regardless of eventual findings (Lewin et al. 2016; Academy of Management Discoveries 2023; Carsten et al. 2023).

Preregistration aids not only individual researchers but also science. Collectively, preregistrations present a more complete picture of a field, the studies conducted in it, and the ones that resulted in publication. Such a comprehensive view allows for meta-analyses that mitigate publication bias, which when unchecked makes studies with "statistically significant" results more likely to be published, thereby exaggerating the likelihood of certain outcomes (Begg and Berlin 1988; Lewin et al. 2016). For this, some professional societies, such as the American Economic Association, encourage or even require preregistration of studies.

When can we deviate from pre-registration? It is recommended for most studies, experimental or otherwise. It is most beneficial when performing hypothesis testing, typically following a pilot study (De Vaus 1993). Explicating hypotheses in advance may be intimidating at first, but authors (and readers) should remember that preregistration is a plan, not a contract. Authors may be right in adding new hypotheses after seeing the data or in deciding not to test preregistered hypotheses, for example due to missing or invalid data. The key is a clear demarcation between preregistered and post-hoc results (i.e., those obtained after seeing the data). When scholars choose to deviate from a preregistration, they should disclose and justify their choices. As long as these are clearly explained and well-motivated, scholars may present preregistered analysis side-by-side with post-hoc results. And readers should practice moderation, remembering that overly restrictive preregistrations can hinder the dissemination of important but unexpected results (Banerjee et al. 2020).

Discussion

In this final section, we look ahead and address ensuing methodological developments, the question of mutually agreed-upon field norms, and the potential for experimental inquiry to fundamentally reorient organization theory.

Expanding the toolkit. As more and more scholars enter the field to examine organization science topics, the tools for experimental inquiry will surely continue to multiply. We highlight four opportunities. First, most experimental inquiry in organization theory has been decidedly deductive in nature, which is not unexpected given experiments' strength in testing causal theory. But experiments can do more. John Stuart Mill (1874) opined that "[an experiment] is immense extension of [an observation]." Nine decades later, Karl Weick (1965, p. 200) echoed him: "experiments are observation. Experiments permit induction as well as deduction. It is just as easy to 'look around and see what is going on' in the laboratory as it is in the field."

Vernon Smith's journey offers an apt example of how experimental exploration brings discovery. A classroom experiment that turned dramatically against theory and his "Harvard education," led him down a path anchored by experimental markets, a tool he developed. They enabled him to "see what will happen" and probe the boundaries of accepted theories (Smith 1962, 1982). Later in his career, Smith (1988) followed a similar strategy to reveal how price errors and subsequent bubbles emerge even highly efficient markets, contradicting established theory (for review, see Wilson 2017).

Experiments are thus not exclusively compatible with a positivist research agenda but can also inform social constructionism (Jost and Kruglanski 2002). In addition to testing theory, a common theme in this issue, experiments can be employed to explore and describe novel phenomena, develop new theory, and advise policymakers (Roth 1986), as some of our contributors do. Experimentalists may simultaneously pursue more than one of these objectives. Doing so may require modified procedures and criteria for evaluating experimental research, and it is important that scholars pay attention to such differences (Schram 2005; Lin et al. 2021).

For example—and this brings us to our second recommendation for expanding the experimental toolkit—endeavors to move beyond theory testing can benefit extraordinarily from multi-method investigations, where experiments are combined with other approaches. While such methodological triangulation (Denzin 1978; Jick 1979) might not be necessary when the focus is on rigorous testing alone, it is increasingly important as researchers seek more applied goals (Lin et al. 2021). Fortunately, the experimental method lends itself to creative combination with a variety of other methodologies. There has been a long history in social psychology of augmenting experiments with qualitative methods, and it is clear that integrating these two approaches can provide substantial leverage, especially on the theory development front (Fine and Elsbach 2000). The merger of qualitative and experimental methods should prove particularly powerful in abductive analysis (Tavory and Timmermans 2014; King et al. 2021), in which a qualitative study identifies unexpected themes that diverge from existing theory, which are then further pursued in an experimental study. Conversely, scholars may also use a design in which an experiment precedes a qualitative-interpretivist study that elaborates the theoretical mechanisms in greater depth (Kaplan 2015). Likewise, many topics in organization theory lend themselves to being studied quantitatively not only with experiments but also with complementary surveys (Ameri et al. 2020; Evans and Schilke Forthcoming) or archival data (e.g., Audia et al. 2000). The combination of two methods creates a contribution that is greater than the sum of its parts. Modeling is, by definition, not empirical. But whether they involve Greek letters or computer code, formal models can serve as springboards for experiments (Gibbons 1999; Fang 2012; Smith and Rand 2018; Knudsen et al. 2019).

Third, a distinction is often made between laboratory and field experiments, but this dichotomy misses nuanced differences and intermediate hybrid forms of experimentation—sometimes referred to as lab-in-the-field studies (Baldassarri and Grossman 2011)—that we believe can add substantial value. Harrison and List (2004) identify two such hybrids on the continuum between laboratory and natural field experiments: An artifactual field experiment resembles a laboratory experiment but relies on nonstandard participants, such as practitioners or a representative sample of a population. A framed field experiment additionally uses field elements and conducted in the participants' natural environment, but it is distinct from a "true" field experiment because participants are aware they are participating in an experiment (List 2023; Puranam 2023).

Fourth, as technology advances, it opens new pathways for experiments. The idea of going beyond text to include audio, video, and pictures in the experimental materials, in the hope of making the study more immersive and engaging for participants, is not new (Aiman-Smith et al. 2002). But the possibilities afforded by virtual and augmented reality could prove a game changer for framed field experiments, combining the exact controls of the laboratory and a stimulus-rich environment, resembling the participants' natural environment (Aguinis and Bradley 2014; Innocenti 2017).

Speaking of technology, we can only imagine how the method will benefit from artificial intelligence (AI), specifically machine learning (ML). For instance, we envision the technology used to identify potential targets for experimental investigation. AI/ML excels in pattern detection but not in determining causality. So, it can direct a researcher's attention to co-occurrences that are unexpected or potentially insightful, setting the stage for an experiment (Shrestha et al. 2020; Agrawal et al. 2022). The same technology can also play a role in the designing of experiments. For instance, it can create stimuli and items that systematically vary along a few dimensions. And while an experiment is running, the technology can automate power calculations, allowing response-adaptive allocation to study conditions, thus improving the efficiency of experiments (Kaibel and Biemann 2021). When analyzing responses, AI/ML can offer better yet systematic classification of participant qualitative responses (Gentzkow et al. 2019), permitting not just Likert-scale responses but writing and speaking. It may also excel is estimating typicality, systematically calculating how an object, whether a tweet or a painting, resembles a prototype. Ultimately, the combination of experimental and AI/ML techniques can aid in inferring the causes of heterogenous responses to a treatment and estimating their magnitude (Athey and Imbens 2016; Wager and Athey 2018; Lee et al. 2022).

And with the rapid improvements in large language models (LLM), like OpenAI's ChatGPT and its open-source peers, we foresee experiments in which not only humans but also autonomous computer algorithms participate, perhaps as organizational interlocutors, much in the spirit of Alan Turing's (1950) famous imitation game. As such tools become part of the organizational repertoire, researchers may even turn the experimental spotlight onto AI agents, making their behavior a subject of experimental studies. Our overall message is clear: We should embrace new methodological opportunities as they present themselves.

Agreeing on field norms. Thanks to its relative infancy, experimental organization theory is still low in institutionalization, making room for out-of-the-box thinking. At the same time, there is a need for some consensus on shared field norms, social and scientific. We hope that our suggestions on experimental design and transparency practices will contribute to scientific progress and rigor, but also to collaboration and camaraderie.

Our field is young, but the experimental method is not. With an approach so omnipresent, it is unsurprising that some distinct norms have emerged across disciplines, and sip into ours. Experimentalists in organization theory tend to draw on traditions that hail from psychology, sociology, or economics. Each comes with its own assumptions and values. We recognize that homogenization of scientific practices is not always desirable (March 2004) and worry that overly restrictive normative pressures may oppress methodological diversity and restrict intellectual exchange. However, the line between "best practice" and "oppressive diktat" is a thin one, so it may be helpful to enumerate and acknowledge differences in designing, conducting, and interpreting experimental results.

First, what counts as an experiment? Some insist that a study cannot be called "experimental" unless it features random assignment to conditions (Grant and Wall 2009), but others use the term more liberally, allowing more definitional breadth. Random assignment excels in ascertaining causality, everyone agrees. But experiments also shine in creating a setting where "adequate control can be maintained and accurate measurement of relevant variables guaranteed" (Wilde, p. 138). Indeed, control is often cited as the critical factor in designing and evaluating experimental research in the economics tradition (Smith 1976).

Experiments are often used to (1) test specific hypotheses and answer questions. But, as we discuss above, they can do more: Extend theory by (2) discovering the boundaries surrounding it — and give rise to theory by (3) exploring behaviors. Turing Award laureate Allen Newell and Nobel laurates Herbert Simon and Vernon Smith relied on experiments for exploration, what Simon (1989, p. 393) and Newall (1972) casually called "just looking" and Smith (1982, pp. 940-3) labeled "see what will happen" or "heuristic" experiments. Experiments can also be used to identify where one theory ends, where its predictions fail (Smith's "boundary" experiments, 1982). Elinor Ostrom, another Nobel laurate, famously used laboratory and field experiments to show how people can self-govern shared resources, contradicting gloomy predictions about the tragedy of the commons (Ostrom 1990; Ostrom et al. 1994). A grand set of framed field experiments across diverse cultures asked how universal is *homo economicus* (Henrich et al. 2001), and showed that people willingly bear the cost to punish unequal behavior and promote cooperation, contradicting assumptions of narrow self-interest (Henrich et al. 2006). Vernon Smith and colleagues (Crockett et al. 2009) created experimental worlds to study how people move from self-sufficiency to discover exchange and specialization.

Disciplines conceptualize experiments differently, and they also evolved distinct norms on how to design them. Take for instance, the debate on the right degree of realism in experiments. One tradition, long rooted in organizational theory, sociology, and economics, insists that experiments can be stylistic, distilling a theory to its essence (Zelditch 1969; Stolte et al. 2001; Lucas 2003b). "The greatest benefits of experiments reside in the fact that they are artificial," proclaim sociologists Webster, Jr. and Sell (2014). As Simon and Smith urged, and like qualitative research, experiments uncover "what can happen" more than "what does happen" in everyday life, so life-like tasks are not a necessity (Mook 1983; Aronson et al. 1998).

This approach does not object to generalizability, of course, but it does not seek it in the experimental design: "One *never* generalizes the empirical findings from laboratory experiments to a larger population of interest," opines Kanazawa (italics in source; 1999, p. 452). Rather than attempting to mirror naturally occurring situations, the experiment should be designed to enable testing of relevant theoretical principles and informing the underlying theory, which serves as the bridge between the experimental study and other settings (Bitektine et al. 2018b).[7]

But in much social psychology – and in economic field experiments – another tradition is prevalent. Many social psychologists "are very careful in experiments to represent those contextual aspects that seem most crucial to the real-world occurrence of the specific phenomenon under investigation" (Ariely and Norton 2007, p. 337). For that they use cover stories, confederates — and deception (Moffitt et al. 2011). In laboratory experiments, economists eschew deception entirely (even if they disagree what it means in practice ; Charness et al. 2021). For their objection, economists often cite concerns about polluting the subject pool with suspicion (and hence some deception is permissible in field experimetns, such as correspondence and audit studies) Yet others add ethical concerns (Ortmann and

[7] Consider, for instance, the classic institutional theory experiments by Zucker (1977. Judging the movement of a pinpoint of light is not an activity common to many organizations outside the lab, yet the task is an ideal fit with the theoretical requirements for testing Zucker's hypotheses on the effects of institutionalization, the results of which have had a fundamental influence on a long line of institutional inquiry in organization theory (Heath and Sitkin 2001).

Hertwig 2002). Sociologists (and the authors of this article) tend to take a middle-ground stance on the acceptability of deception (Cook and Yamagishi 2008; Webster and Sell 2014), such that deception—especially in its passive form[8]—is deemed admissible if theory testing would be impossible or highly impractical without it and if participants are debriefed (Schilke et al. 2023). However, whenever an experimental design can be implemented that avoids deception without sacrificing the validity of the research results, this design should be preferred.

Economists can be equally concerned about context, because "the environment of the experiment can also influence behavior. The environment can provide context to suggest strategies and heuristics" (Harrison and List 2004, p. 1013). In economics, concerns about the context have fueled the rise of field experiments (including RCTs), which require realistic tasks, familiar settings, and suitable participants (e.g., List and Lucking-Reiley 2002; Bertrand and Duflo 2017).

It is not clear to us that one approach is superior to the other. The organizational experimentalist should weigh the question and the theory (Bitektine, Lucas, and Schilke 2018) because "what is most important about any particular experiment is that it be relevant to its purpose. If its purpose is to test a theory, then it is legitimate to ask whether the elements of alleged 'unrealism' in the experiment are parameters in the theory" (Smith 1982, p. 937).

As experiments have become more widely embraced, disparate norms may emerge, as has been the case already in psychology and economics (Ariely, 2007 #1814) and between laboratory and field experimenters (Falk and Heckman 2009; Maniadis et al. 2014). But incompatible definitions and traditions of what is a "proper" experiment, combined with restrictive normative pressures, risk oppressing methodological diversity and productive intellectual exchange. We concur that "it has been more important for experimentalists to present a rich variety of examples of their work than abstract explanations of why one might perform experiments" (Smith 1982, p. 923). We prefer to maintain rather than quell this diversity of approaches. Our hope is that organization theorists—including authors,

[8] Passive deception involves withholding information from participants or not correcting their misconceptions about certain aspects of the study, whereas deception occurs when researchers give false information to participants.

reviewers, editors, and ultimately readers and users—will remain open-minded. It would be unfortunate if promising papers are rejected or useful data discarded simply due to the gatekeepers' discipline-based preferences. Of course, this is not to say that "everything goes," yet we need to realize that even when it comes to seemingly clear-cut best practices, such as preregistration, there are well-reasoned differences in opinions on what constitutes the "right approach" (Krishna 2021).

Unequivocally, we advocate for more transparency in research. But we also wish to establish a climate of mutual support and encouragement rather than suspicion and reprimand. "Scientists are humans" (Clark et al. 2022, p. 46), and the costs, monetary and mental, of dealing with unsubstantiated allegations of errors and guilt-by-association tactics are very real. The increased transparency afforded by the experimental method should not lower the bar for publicly condemning a study, colleague, or research stream. Just like the results of an empirical investigation, allegations of methodological weaknesses can suffer from false positives, and we must weigh the considerable benefits of fraud prevention against the downsides of false accusations (Goldenring 2010; Bouter and Hendrix 2017). Correspondingly, we should calibrate the tone of the debate and agree on appropriate channels through which academic controversy happens, such that the emphasis is on objective scrutiny and constructive critique, not "denunciation mobs" (Stevens et al. 2020) sharing opinions on social media (Hillygus 2018).

We applaud those who seek to better the transparency and reproducibility of our research. We recognize that such qualities are essential to the standing of our profession and its societal contribution (Starbuck 2016; Tsui 2021). Yet we hold such truth-seekers to the same standards as the authors they invigilate: If you detect hints of negligence (or worse), share all data, materials, statistical codes, analyses, and results. Moreover, seek response from the original authors and share it, too. And remember: Many studies are coauthored. A researcher may be sloppy (or fraudulent), but their associates should not be automatically declared guilty by association. Sometimes, the whistleblowers are the junior authors, as in the case of Marc Hauser, once a prolific researcher and popular teacher, who was turned in by his graduate students (Carpenter 2012).

Whenever poor judgment (or malicious intent) is suspected, begin by identifying the culprit. In recent cases, publicly available data files allowed peer investigators to hone-in on one of the authors. Without data, all coauthors may have been guilty by association, with their reputations tarnished. For this reason, it is in everyone's interest – researchers, reviewers, and readers – to boost transparency. As Brian Nosek (2023) of The Open Science Framework (OSF) puts it, "a commitment to transparency by researchers, journals, funders, and institutions will not prevent fraud, but it will make fraud more inconvenient and easier to detect."

Indeed, truth-seekers and whistle-blowers can only do so much. In the long run, the battle for research quality will be won by institutional efforts (Aguinis et al. 2022), such as journals demanding transparent reporting (Lewin et al. 2016; Antonakis 2017; Schwab and Starbuck 2017) and providing journal pages for replication studies and registered reports (Bettis et al. 2016; Ryan and Tipu 2022; Academy of Management Discoveries 2023; Carsten et al. 2023; Davis et al. 2023). Such efforts have led to some striking revisions of once-established findings (e.g., Goldfarb et al. 2018; Ertug and Maoret 2020; Goldfarb and Yan 2021). Also, reform is afoot thanks to social movements such as the Community for Responsible Research in Business and Management, a collaboration between distinguished scholars, business school deans, leading journals, and industry associations like AACSB (www.rrbm.network; RRBM 2020).

Advancing from experiments to experimental organizations. Management scholars are well acquainted with Nobel laureate Daniel Kahneman, but they pay too little attention to the contribution of the man who shared the prize with him: Vernon Smith, who was praised "for having established laboratory experiments as a tool in empirical economic analysis" (Nobel Prize 2002). If Kahneman used experiments to rewrite the theory on decision-making, Smith was the pioneer who drove the method to the heart of economics. The introduction of experiments as a legitimate empirical method was not entirely planned. After struggling to convey microeconomic theory to students, Smith (1962, 1982) designed a classroom demonstration modeled after the stock and commodities exchanges. To his surprise, the tiny experimental market revealed uncommon insights. He altered some conditions and continued experimenting.

The reception was lukewarm: his first paper, documenting years of market experiments, was published after two revisions, four negative referee reports, and an initial rejection. But to Smith (1991, p. 154), this was but a step in "a continuing struggle of escape from the prison of conventional patterns of economic thought." He may have been encouraged to find that, around the same time, Reinhard Selten had been pioneering economics experiments in Germany (Nagel et al. 2016). He, too, won the Nobel Prize.

Smith published his first experiments in 1962. A few years earlier, Selten published what may be the first strategy experiment, assigning three participants to run a simulated bank (Sauermann and Selten 1959). Later, Selten travelled to Pittsburgh, fascinated to meet Herbert A. Simon and his associates, who were busy conducting experimental research in organization theory (Nobel Prize 1994; Mitzkewitz and Nagel 2020). As we discussed above, they pointed out that experiments were crucial to gaining knowledge about organizations.

We envision a bright future for the experimental organization—an organization created for experimental purposes (Simon and Guetzkow 1952; Guetzkow and Bowes 1957; Weick 1965). As a good design requires, such an organization is distinguished by its theoretically-accurate features, like systems of coordinated action and collective goals, structure and incentives, and the involvement of more than one members (Camerer and Weber 2013). Naturally, the experimental organization may be simpler than many. It exists in a controlled environment. It may appear slightly artificial.

But these are features, not bugs: Simplicity, control, and exact measurements are crucial to discovery and testing (Zelditch 1969; Wilde 1981; Kriss and Weber 2013; Webster and Sell 2014). The lilliputian size of the experimental organization may raise eyebrows, but recall that size is not among the definitional criteria for an organization (Puranam et al. 2015; Puranam 2018). The world of entrepreneurship is filled with two-person startups, and many economies are dominated by small firms.

Few will argue that these are excluded from organization theory.[9] A stream of research studying experimental organizations could bring to our field the vast gains that economics enjoyed by studying experimental markets.

Some forms of experimental organizations have already been put to work by researchers studying topics such as incentives in teams (e.g., Brandts and Cooper 2006; Hamman et al. 2007) and leadership (Weber et al. 2001; Brandts et al. 2015; Cooper et al. 2020; Antonakis et al. 2022). Others studied experimental organizations to understand collective learning (Blume et al. 2009), identity (Chen and Chen 2011; Chen et al. 2021), culture (Weber and Camerer 2003), and office politics (Carpenter et al. 2010). Much of these studies are informed by economic theory and employ associated designs. Organizational scholars could contribute much by examining experimental organizations through the diversity of thought and perspectives that characterizes our field.

With the tight fit between recent theoretical trends and the strengths of the experimental method, we envision experiments contributing to a profound reorientation of empirical research toward the definitional features that set organization theory apart: an uncompromising dedication to clearly identify—rather than hint at—causal effects and theoretical mechanisms.

Whether you are an organizational scholar pondering experiments—or already using them—this is a promising time. There is considerable momentum in this burgeoning space, yet many important contributions are still to be made. In the coming years, organization theory may find its Louis Pasteur moment.

[9] To be clear: Size does not define an organization, but it can be a variable of interest for research. Further, because organizations can be very small, one can wonder about the differences between teams and organizations. These may be related to hierarchy (teams are typically a subunit of a larger organization) and communication patterns (perhaps more complete in teams) as well as other characteristics.

Epilogue: How The Contributions Here Benefit from the Experimental Method

We are delighted to introduce this *Organization Science* Special Issue, inspired by the growing trend in the organization theory of scholars utilizing experimental methodology. We aimed to create a forum for this emerging community to unite and push conceptual and methodological frontiers. We also intended to showcase the key merits that experiments offer to the organization theorist, jumpstart discussions, and facilitate the exchange of best practices and emerging norms in this burgeoning community.

In soliciting submissions, we explicitly encouraged disciplinary and methodological diversity, encompassing studies informed by sociology, psychology, or economics and employing laboratory, artefactual, framed, or field experiments, often combined with non-experimental studies. We were pleased by the overwhelming response, with numerous papers received and reviewed.

The results of this comprehensive review process are presented in this Special Issue. Each article in this collection individually makes essential contributions that enhance our understanding of organization theory. Collectively, they serve as a pivotal benchmark for experimental research. The work contained in the Special Issue sheds new light on the structures and operations of organizations and the embeddedness of organizational actors in their social environment, leveraging the strengths of experimental methodology in various ways. Next, we discuss just that (an editor's assessment of each article appears in the online companion).

Many of the articles in the issue excel in leveraging experiments to capture **theoretical mechanisms** that may be difficult to identify with alternative, non-experimental methodologies, such as protocol analysis, which relies on verbal reports as data (Laureiro-Martinez et al. Forthcoming). This capacity also allows for teasing apart competing mechanisms, such as in van den Broek et al.'s (forthcoming) investigation of legitimacy. The article shows that peer endorsement stimulates evaluators to express their judgments, particularly among those who expend limited cognitive effort, and that evaluators in different evaluative modes act differently when their propriety evaluations are based on instrumental, moral, or relational considerations.

Uncovering **microfoundations** of organization theory is a central theme across multiple articles in the Special Issue. Many of the investigations are breaking new ground in understanding cross-level mechanisms. For example, Shen et al. (forthcoming) add considerable richness by disaggregating field norms and studying the effect of consensus among technology experts on investments in new market sectors. While previous studies have shown that establishing evaluation criteria benefits well-performing firms, the authors argue that consensus among experts will lead to increased investments in all firms within the new sector. Bergenholtz et al. (forthcoming), who make microfoundations front and center in their article's title, spotlight the importance of cognitive abilities and styles in organizations' adaptive search. Methodologically, this is an excellent example of a design that relies not on manipulation but on careful control of conditions in a way that is rarely possible outside the laboratory. The authors have combined qualitative interviewing with laboratory and labor market studies (Amazon Mechanical Turk). They have generously shared all their data, scripts, and instruments for use by others. Moreover, they assured readers that they had provided a complete account of all studies and variables conducted and had not resorted to selective reporting of results (footnote 1).

Experiments allow us to access **difficult-to-observe phenomena**. Despite its importance for the country's economy, little research has been conducted on the Cuban cigar industry, so Verhaal et al. (forthcoming) designed a series of vignette experiments to dig into the organizational consequences of authenticity claims in this setting. The article shows that provenance-based claims of authenticity limit a firm's ability to expand production, while association with an iconic individual allows for more flexibility without undermining authenticity claims. Further, Gorbatai et al. (forthcoming demonstrate the advantage of experiments in shedding light on theoretical processes that can be difficult to observe, such as discrimination. This paper examines how high-salience racial events affect biases and acts of discrimination. The authors propose that discrimination reflects individual and environmental differences, and they find that racially salient events depress the quality evaluations and success odds of African-American entrepreneurs, leading to a significant disadvantage for minority groups.

Further, some articles stand out in their efforts to conduct **elaborate field experiments**. Calls for experiments conducted in the field abound. Still, a key challenge for researchers following these calls is collecting a sufficiently powered sample while maintaining control over the experimental manipulation. Several of the authors in this issue were able to tackle this challenge. For instance, Boudreau and Kaushik (forthcoming) conducted a large-scale field experiment under highly controlled conditions, assessing undergraduate students' willingness to participate in a tech training program. The study relied on a clever design: as part of the launching campaign, an email was sent out to numerous potential adopters, involving competition or collaboration treatments to manipulate perceptions of interactions with other participants. A notable strength of the study was minimizing contamination threats by creating two separate platforms (one for each condition) and ensuring participants were only engaged on their platform. For example, the authors had some control over the sharing of messages among participants, by asking them not to forward any messages and by detecting and excluding any minimal instances of sharing. This type of micro-level intervention on the part of researchers is rare in field experiments, making this design an exemplar. The article by Boss et al. (forthcoming) is another apt example of a well-designed field experiment. The researchers conducted a field experiment with hundreds of students enrolled in a lean startup entrepreneurship course over 11 weeks. The aim was to understand the separate and combined effects of granting autonomy in choosing ideas and team members compared to a baseline treatment where views and team members were preassigned.

Several articles set out to combine the benefits of laboratory experiments with relevant aspects from the field by conducting **lab-in-the-field studies**, which could be either artefactual or framed field experiments (Harrison and List 2004). Hergueux et al. (forthcoming) assess the relationship between reciprocal preferences and group performance using lab-in-the-field experiments in open-source software development. In one leg, they contacted numerous developers of open-source software and enrolled them in an online public goods experiment to gauge their cooperative tendencies. In the other leg, they collected archival data about actual contributions of software code by their participants, but also by non-participants. Combined, these data allowed them to draw original conclusions about reciprocity and organizations' success (or failure).

Similarly, Molina et al. (forthcoming employ lab-in-the-field studies to examine why leaders of organizations cooperate with others they may never transact with again. The article finds that belief in the reliability of robust social norm enforcement leads to a higher probability of cooperation with strangers, highlighting the importance of social norms in facilitating cooperation in a market economy.

Another set of articles let **experimental organizations** spring to life in the lab. The study by Christensen et al. (forthcoming is a prime example of how experimental organizations, minimalistic entities created with theoretically relevant characteristics, can help produce intriguing insights. Using two-member organizations, the authors manipulated decision-making structures in terms of aggregation rule (consensus), incentives, and assignment of positions (fixed versus random). To study how organizations make decisions regarding team selection and delegation of authority in uncertain environments, Hamman and Martínez-Carrasco (forthcoming) modeled a decision environment in which a manager both determines the skill heterogeneity of the workers and whether to retain or delegate the ability to allocate tasks. The authors designed a multiple-round game, with participants randomly assigned to manager and worker roles in three-person organizations. The study emphasizes the potential costs associated with managers seeking decision simplification, but also highlights how deliberative thinking and risk tolerance can enhance performance in a complementary manner.

Sometimes, experiments have received critiques for favoring a static perspective, but some articles in this issue capture important **dynamics** in their experimental studies. Alves et al. (forthcoming, for example, analyze the role of type 1 (intuitive) and type 2 (reflective) processing in the adaptation of routines during discontinuous change. The authors find that type 1 processing facilitates organizational adaptation more than type 2 processing, particularly in highly ambiguous environments where information is limited and difficult to verify. Likewise, Radoynovska and Ruttan (forthcoming extend research on hybridity (as a state) to study hybridization (as a process), analyzing how audience judgments of organizational authenticity and the ability to attract potential employees are affected by organizations transitioning from single to multiple categories. They find that hybridization triggers audience cynicism and leads to decreased judgments of authenticity. Still, these penalties are only observed when organizations move away from field-level profit-status norms. Finally, the experiments by Rahmandad and Gary (forthcoming make time for their focal concern by investigating the influence of delays and outcome uncertainty on strategic choices, learning, and performance outcomes. Their findings reveal that longer delays lead to a preference for alternatives with rapid returns, even if they are suboptimal, and induce learners to persist with their initial strategies. Delays and uncertainty can undermine performance, reduce heterogeneity in strategies, and lead to convergence in more dynamic and uncertain environments.

Experiments can effectively **complement a variety of non-experimental methods**, and several papers in this special issue showcase this complementarity, such that the individual studies genuinely come together and the whole is greater than the sum of its parts. Di Stefano and Micheli (forthcoming, is an excellent example of a mixed-methods approach. Phases 1 and 2 of their research focused on building theory through desk research, interviews, and observations at CERN. These phases refined the empirical strategy and narrowed the focus. Phase 3 involved a lab-in-the-field study administered to scientists at CERN. In Phases 4 and 5, feedback was collected, and additional interviews were conducted to make sense of the empirical evidence. Phase 6 involved designing and implementing two laboratory experiments conducted among Prolific panelists to test the theoretical model. The article by Pamphile and Ruttan (forthcoming) is another example of a mixed-study approach. The authors start with archival data collected from Glassdoor to detect correlations between the key variables of interest. Two experimental studies were then used to investigate the causal relationships between variables, eliminate alternative explanations, and unpack the underlying mechanisms. The authors find that perceptions of authenticity are influenced by the congruence between an organization's stated values (values it claims to have) and its lived values (values members perceive the organization to embody). To investigate the mechanism behind their findings, Raveendran, Srikanth, and Ungureanu (2023) add an agent-based simulation to experimental findings obtained from humans in universities and online. The computational model —

essentially a synthetic experiment (Levine & Prietula 2012) — simulates how adaptive learning affects the exploration-exploitation choice under different performance goals.

In terms of **construct validity**, several studies have reported pertinent analyses, and some have conducted additional studies to demonstrate the robustness of the results to different operationalizations. In doing so, Wang et al. (forthcoming have made good use of experiments' ability to measure actual behavior (rather than behavioral intentions). This paper examines how activists gain allies in organizations. It shows that potential allies' support for feminist policies depends on their level of feminist and organizational identification and the labeling of the policy. In one of the experiments, participants had the opportunity to support activists' efforts by completing an email task, which allowed for measuring performance on the task as a behavioral measure of policy support.

The articles in the Special Issue excel in **internal validity**, leveraging experiments' ability to identify causality and rule out alternative explanations, such as self-selection. But, one paper goes further and explores differences in observed effects depending on whether participants were randomly assigned or could self-select into study conditions. Gibbons et al. (forthcoming investigate how parties build relational contracts to adapt to unforeseen changes efficiently. They reveal that principle-based agreements, endogenously emerging, achieve significantly higher levels of performance after an exogenous change in the environment. Still, when participants were nudged into such agreements, performance did not improve. Another article that stands out on the internal validity front is that by Bernstein et al. (forthcoming), who pioneered an experimental design in a field dominated by non-experimental findings. To complement extant evidence, the authors turned to experiments to study how communication network structure affects an organization's adaptability to a shifting environment. They utilized an original "murder mystery" task designed to lead participants into settling on an early but wrong conclusion, requiring them to shift their understanding as more information became available. The authors shared all their data and code for use by others and generously made available the instrument used in the study: a software platform to conduct experiments involving real-time collaboration through an exogenously determined communication network structure.

Finally, all articles in this Special Issue follow state-of-the-art recommendations for **transparency and reproducibility**, disclosing data and code to increase the findings' reproducibility. For example, Ghosh and Wu (forthcoming do an exemplary job separating confirmatory hypotheses testing from exploratory post-hoc analyses. They preregistered the hypotheses underlying the laboratory experiment, thus assuring the reader that the hypotheses tested were indeed conjectured before the study (and not after seeing the results). They report statistical power analyses to probe the robustness of the results and generously provide their recruiting materials, instruments, data, and statistical code. Another example of high analytical transparency and reproducibility is the article by Pittnauer et al. (forthcoming. To benefit future research, they labored to provide their materials, even creating a video documenting the experimental interface as the participants saw. To mitigate the risk that participants were influenced by others who participated earlier, Pittnauer and colleagues took the admirable step of conducting the experiment across five universities. Throughout the article, they report exact *p*-values, as recommended (Wasserstein et al. 2019), and they do not just describe their results but report them in visually striking diagrams (Greve 2018; Levine 2018).

Overall, the Special Issue on experimental approaches to organization theory showcases a diverse portfolio of high-quality experimental studies, which we hope will contribute to setting a standard in the field. Together, the studies underscore the value of the experimental approach to organization theory. They exude abundant fertility and hold auspicious potential to yield substantial progress in the times ahead.

Appendix: Manuscript Checklist

This checklist recaps some of the best practices, as reported by organization researchers and based on the recommendations of scientific associations and leading journals. It can be used when designing and conducting research, whether experimental or not. It can also be used when reviewing manuscripts or evaluating the gravitas of published research.

Data, Analytic Methods (Code), and Research Materials Transparency

Present	*n/a*	*Comments*	
☐	☐	☐	Links to publicly available data set (example) and/or statement on restriction on data availability
☐	☐	☐	Links to statistical or mathematical code (e.g., log files)
☐	☐	☐	Links to a description (and/or demonstration) of the instrument used, sufficiently detailed to allow a reader to recreate it (example)

The data, code, and instruments should be detailed enough so that any reader can understand and assess the conclusions of the manuscript (§2, 3, 4 in Nosek et al. 2015; Davis et al. 2023, pp. 13-4).

Design and Analysis Transparency

(Sources: Nature Research, §5 in Nosek et al. 2015; Davis et al. 2023, pp. 13-4). For a statistical expansion, see https://www.nature.com/collections/qghhqm/pointsofsignificance

☐	☐	☐	Clear separation between (confirmatory) hypothesis testing and (exploratory) post-hoc analysis, ideally by preregistration, prespecification, or registered trial (see next ;Kerr 1998; Murphy and Aguinis 2019)
☐	☐	☐	Description of <u>all</u> variables (measures) for which data were collected, even if some were not used in analysis (to avoid the temptation of p-hacking;
☐	☐	☐	Description of <u>all</u> experimental conditions (manipulations), even if some were not used in analysis
☐	☐	☐	Disclosure of any data exclusions (e.g., based on attention checks or other criteria)
☐	☐	☐	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☐	☐	Discussion of how the sample size was determined, including analysis of statistical power (ideally, a priori) to show that the sample size justifies the conclusions drawn from it (Button et al. 2013)
☐	☐	☐	Description of any assumptions or corrections, if applicable, such as tests of normality and adjustment for multiple comparisons
☐	☐	☐	Full description of variables including
☐	☐	☐	central tendency (e.g., means) or other basic estimates (e.g., regression coefficient)
☐	☐	☐	variation (e.g., standard deviation) or associated estimates of uncertainty (e.g., confidence intervals)
☐	☐	☐	visualizations of disaggregated raw data, e.g., scatterplot matrixes, binned scatterplots, residual plots, and Cook's distance (Greve 2018; Levine 2018; Starr and Goldfarb 2020)
☐	☐	☐	The statistical test(s) used, and whether they are one- or two-sided
☐	☐	☐	For null hypothesis testing, the test statistic (e.g., F, t, r) with confidence intervals, effect sizes, degrees of freedom

Present	*n/a*	*Comments*	
☐	☐	☐	For hierarchical and complex designs, identification of the appropriate level of analysis for tests and full reporting of outcomes
☐	☐	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☐	☐	Exact p-values rather than stating that a finding is "significant" (Wasserstein et al. 2019)
☐	☐	☐	Estimates of effect sizes (e.g., Cohen's d, Pearson's r)

Pre-Specification \ Pre-registration
(§6 in Nosek et al. 2015)

☐	☐	☐	Statement of the questions or hypotheses that motivate the study
☐	☐	☐	A preregistered sampling plan, which specifies the sampling method, the sample size, and principles for removing data (e.g., outliers or erroneous entries)
☐	☐	☐	A preregistered measurement plan, which details the measures used and how they will be coded
☐	☐	☐	A preregistered analysis plan, which lists the measures (e.g., coefficients) that will be estimated and how they will be interpreted
☐	☐	☐	If a preregistration was created, the analysis should be reported as prespecified and linked to the pre-specification plan. Prespecification does not preclude additional analyses, exploratory or post-hoc.

Citation Standards
(§1 in Nosek et al. 2015)

☐	☐	☐	If any existing data sets, research materials, or code were used as part of this manuscript, include a citation to each in the main reference section, so that they are recognized as original contributions

IRB/Ethical Review

☐	☐	☐	A statement confirming compliance with the relevant ethical regulations

Academy of Management Discoveries. (2023). Registered Reports from https://aom.org/events/event-detail/2024/01/20/higher-logic-calendar/registered-reports
Agarwal R, Croson R, Mahoney JT (2010) The Role of Incentives and Communication in Strategic Alliances: An Experimental Investigation. *Strat Manag J* 31(4):413-37.
Agrawal A, Gans J, Goldfarb A (2022). *Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence* (Harvard Business Review Press, Boston).
Aguinis H, Archibold EE, Rice DB (2022) Let's Fix Our Own Problem: Quelling the Irresponsible Research Perfect Storm. *J Manag Stud* 59(6):1628-42.
Aguinis H, Bradley KJ (2014) Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods* 17(4):351-71.
Aguinis H, Solarino AM (2019) Transparency and Replicability in Qualitative Research: The Case of Interviews with Elite Informants. *Strategic Management Journal* 40(8):1291-315.
Aiman-Smith L, Scullen SE, Barr SH (2002) Conducting Studies of Decision Making in Organizational Contexts: A Tutorial for Policy-Capturing and Other Regression-Based Techniques. *Organizational Research Methods* 5(4):388-414.
Alan S, Corekcioglu G, Sutter M (2023) Improving Workplace Climate in Large Corporations: A Clustered Randomized Intervention. *Q J Econ* 138(1):151-203.
Alves MFR, Vastola V, Galina SVR, Zollo M (forthcoming) When Reflection Hurts: The Effect of Cognitive Processing Types on Organizational Adaptation to Discontinuous Change. *Organization Science*.
Ameri M, Rogers SE, Schur L, Kruse D (2020) No Room at the Inn? Disability Access in the New Sharing Economy. *Academy of Management Discoveries* 6(2):176-205.
Antonakis J (2017) On Doing Better Science: From Thrill of Discovery to Policy Implications. *Leader Q* 28(1):5-21.
Antonakis J, d'Adda G, Weber RA, Zehnder C (2022) "Just Words? Just Speeches?" on the Economic Value of Charismatic Leadership. *Manag Sci* 68(9):6355-81.
Ariely D, Norton MI (2007) Psychology and Experimental Economics - a Gap in Abstraction. *Curr Directions Psychol Sci* 16(6):336-9.
Aronson E, Ellsworth P, Carlsmith J, Gonzales M (1990). *Methods of Research in Social Psychology*, 2nd ed (McGraw-Hill).
Aronson E, Wilson TD, Brewer MB (1998) Experimentation in Social Psychology. Gilbert DT, Fiske ST, Lindzey G, eds. *The Handbook of Social Psychology (Vol. 1, Pp.)* (McGraw-Hill, Boston, MA), 99–142.
Athey S, Imbens G (2016) Recursive Partitioning for Heterogeneous Causal Effects. *Proc Natl Acad Sci USA* 113(27):7353-60.
Audia PG, Locke EA, Smith KG (2000) The Paradox of Success: An Archival and a Laboratory Study of Strategic Persistence Following Radical Environmental Change. *Academy of Management Journal* 43(5):837-53.
Aviram H (2012) What Would You Do? Conducting Web-Based Factorial Vignette Surveys. Gideon L, ed. *Handbook of Survey Methodology for the Social Sciences* (Springer, New York, NY), 463-73.
Azrieli Y, Chambers CP, Healy PJ (2018) Incentives in Experiments: A Theoretical Analysis. *J Polit Econ* 126(4):1472-503.
Baker TL (1994). *Doing Social Research*, 2 ed (McGraw-Hill, New York, NY).
Baldassarri D, Abascal M (2017) Field Experiments across the Social Sciences. *Annual Review of Sociology* 43(1):41-73.
Baldassarri D, Grossman G (2011) Centralized Sanctioning and Legitimate Authority Promote Cooperation in Humans. *Proceedings of the National Academy of Sciences* 108(27):11023-7.
Bamberger PA (2019) On the Replicability of Abductive Research in Management and Organizations: Internal Replication and Its Alternatives. *Academy of Management Discoveries* 5(2):103-8.
Bandiera O, Barankay I, Rasul I (2007) Incentives for Managers and Inequality among Workers: Evidence from a Firm-Level Experiment*. *Q J Econ* 122(2):729-73.

Banerjee A, Duflo E, Finkelstein A, Katz LF, Olken BA, Sautmann A (2020) In Praise of Moderation: Suggestions for the Scope and Use of Pre-Analysis Plans for Rcts in Economics. *Working Paper*.
Banerjee AV, Duflo E (2009) The Experimental Approach to Development Economics. *Annual Review of Economics* 1(1):151-78.
Bardsley N, Cubitt R, Loomes G, Moffatt P, Starmer C, Sugden R (2010). *Experimental Economics: Rethinking the Rules* (Princeton University Press, Princeton).
Baron RM, Kenny DA (1986) The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *J Pers Soc Psychol* 51(6):1173-82.
Baumeister RF, Vohs KD, Funder DC (2007) Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspect Psychol Sci* 2(4):396-403.
Beamish PW, Hasse VC (2022) The Importance of Rare Events and Other Outliers in Global Strategy Research. *Global Strategy Journal* 12(4):697-713.
Begg CB, Berlin JA (1988) Publication Bias: A Problem in Interpreting Medical Data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 151(3):419-63.
Bergenholtz C, Vuculescu O, Amidi A (forthcoming) Microfoundations of Adaptive Search in Complex Tasks: The Role of Cognitive Abilities and Styles. *Organization Science*.
Bergh DD, Sharp BM, Aguinis H, Li M (2017) Is There a Credibility Crisis in Strategic Management Research? Evidence on the Reproducibility of Study Findings. *Strat Organ* 15(3):423-36.
Bernstein ES, Shore JC, Jang AJ (forthcoming) Network Centralization and Collective Adaptability to a Shifting Environment. *Organization Science*.
Bertrand M, Duflo E (2017) Field Experiments on Discrimination. Banerjee AV, Duflo E, eds. *Handbook of Economic Field Experiments* (North-Holland), 309-93.
Bettis RA (2012) The Search for Asterisks: Compromised Statistical Tests and Flawed Theories. *Strat Manag J* 33(1):108-13.
Bettis RA, Helfat CE, Shaver JM (2016) The Necessity, Logic, and Forms of Replication. *Strat Manag J* 37(11):2193-203.
Bitektine A (2011) Toward a Theory of Social Judgments of Organizations: The Case of Legitimacy, Reputation, and Status. *Academy of Management Review* 36(1):151-79.
Bitektine A, Lucas J, Schilke O (2018a) Institutions under a Microscope: Experimental Methods in Institutional Theory. Bryman A, Buchanan DA, eds. *Unconventional Methodology in Organization and Management Research* (Oxford University Press, Oxford), 147-67.
Bitektine A, Lucas J, Schilke O, Aeon B (2022) Experiments in Organization and Management Research *Oxford Research Encyclopedia of Business and Management*.
Bitektine A, Lucas JW, Schilke O (2018b) Institutions under a Microscope: Experimental Methods in Institutional Theory. Bryman A, Buchanan DA, eds. *Unconventional Methodology in Organization and Management Research* (Oxford University Press, Oxford), 147-67.
Blader S, Gartenberg C, Prat A (2019) The Contingent Effect of Management Practices. *Rev Econ Stud* 87(2):721-49.
Blume A, Duffy J, Franco AM (2009) Decentralized Organizational Learning: An Experimental Investigation. *Am Econ Rev* 99(4):1178-205.
Bolinger MT, Josefy MA, Stevenson R, Hitt MA (2022) Experiments in Strategy Research: A Critical Review and Future Research Opportunities. *Journal of Management* 48(1):77-113.
Bollen KA (1989). *Structural Equations with Latent Variables* (Wiley, Toronto).
Bolton GE, Ockenfels A, Thonemann UW (2012) Managers and Students as Newsvendors. *Manag Sci* 58(12):2225-33.
Boss V, Dahlander L, Ihl C, Jayaraman R (forthcoming) Organizing Entrepreneurial Teams: A Field Experiment on Autonomy over Choosing Teams and Ideas. *Organization Science*.
Boudreau KJ, Kaushik N (forthcoming) Gender Differences in Responses to Competitive Organization? A Field Experiment on Differences between Stem and Non-Stem Fields from an Internet-of-Things Platform. *Organ Sci* 0(0):null.

Bouter LM, Hendrix S (2017) Both Whistleblowers and the Scientists They Accuse Are Vulnerable and Deserve Protection. *Accountability in Research* 24(6):359-66.
Brandon DM, Long JH, Loraas TM, Mueller-Phillips J, Vansant B (2013) Online Instrument Delivery and Participant Recruitment Services: Emerging Opportunities for Behavioral Accounting Research. *Behavioral Research in Accounting* 26(1):1-23.
Brandts J, Cooper DJ (2006) A Change Would Do You Good .... An Experimental Study on How to Overcome Coordination Failure in Organizations. *Am Econ Rev* 96(3):669-93.
Brandts J, Cooper DJ, Weber RA (2015) Legitimacy, Communication, and Leadership in the Turnaround Game. *Manag Sci* 61(11):2627-45.
Breitsohl H (2021) Experimental Designs in Business Research *Oxford Research Encyclopedia of Business and Management* (Oxford University Press).
Brodeur A, Cook N, Hartley JS, Heyes A (2022) Do Pre-Registration and Pre-Analysis Plans Reduce P-Hacking and Publication Bias? *Working Paper*.
Brutus S, Aguinis H, Wassmer U (2013) Self-Reported Limitations and Future Directions in Scholarly Reports: Analysis and Recommendations. *Journal of Management* 39(1):48-75.
Buhrmester M, Kwang T, Gosling SD (2011) Amazon's Mechanical Turk: A New Source of Inexpensive, yet High-Quality, Data? *Perspectives on Psychological Science* 6(1):3-5.
Burbano VC (2021) The Demotivating Effects of Communicating a Social-Political Stance: Field Experimental Evidence from an Online Labor Market Platform. *Manag Sci* 67(2).
Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafo MR (2013) Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience. *Nature Reviews Neuroscience* 14(5):365-76.
Caffarelli RV (2009). *Galileo Galilei and Motion: A Reconstruction of 50 Years of Experiments and Discoveries* (Springer, New York, NY).
Camerer C, Weber R (2013) Experimental Organizational Economics. Gibbons R, Roberts J, eds. *Handbook of Organizational Economics* (Princeton University Press, Princeton, NJ), 213-62.
Camerer CF, Dreber A, Forsell E, Ho T-H, Huber J, Johannesson M, Kirchler M et al. (2016) Evaluating Replicability of Laboratory Experiments in Economics. *Science* 351(6280):1433-6.
Camerer CF, Dreber A, Holzmeister F, Ho T-H, Huber J, Johannesson M, Kirchler M et al. (2018) Evaluating the Replicability of Social Science Experiments in Nature and Science between 2010 and 2015. *Nat Hum Behav* 2(9):637-44.
Camerer CF, Dreber A, Johannesson M (2019) Replication and Other Practices for Improving Scientific Quality in Experimental Economics. Schram A, Ule A, eds. *Handbook of Research Methods and Applications in Experimental Economics* (Edward Elgar Publishing), 83-102.
Camerer CF, Hogarth RM (1999) The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework. *J Risk Uncertainty* 19(1):7-42.
Campbell DT, Stanley JC (1963) Experimental and Quasi-Experimental Designs for Research. Gage NL, ed. *Handbook of Research on Teaching* (Rand McNally, Chicago, IL), 171-246.
Camuffo A, Cordova A, Gambardella A, Spina C (2020) A Scientific Approach to Entrepreneurial Decision-Making: Evidence from a Randomized Control Trial. *Manag Sci* 66(2):564-86.
Carpenter J, Matthews PH, Schirm J (2010) Tournaments and Office Politics: Evidence from a Real Effort Experiment. *Am Econ Rev* 100(1):504-17.
Carpenter S (2012) Harvard Psychology Researcher Committed Fraud, U.S. Investigation Concludes *Science*.
Carsten M, Clapp-Smith R, Alexander Haslam S, Bastardoz N, Gooty J, Connelly S, Spain S (2023) Doing Better Leadership Science Via Replications and Registered Reports. *Leadership Quarterly*:101712.
Chamberlin EH (1948) An Experimental Imperfect Market. *Journal of Political Economy* 56(2):95-108.
Charness G, Samek A, van de Ven J (2021) What Is Considered Deception in Experimental Economics? *Exp Econ*.
Chatterji AK, Findley M, Jensen NM, Meier S, Nielson D (2016) Field Experiments in Strategy Research. *Strat Manag J* 37(1):116-32.

Chen R, Chen Y (2011) The Potential of Social Identity for Equilibrium Selection. *Am Econ Rev* 101(6):2562-89.
Chen R, Chen Y, Riyanto YE (2021) Best Practices in Replication: A Case Study of Common Information in Coordination Games. *Exp Econ* 24(1):2-30.
Christensen G, Freese J, Miguel E (2019). *Transparent and Reproducible Social Science Research: How to Do Open Science* (University of California Press).
Christensen M, Dahl CM, Knudsen T, Warglien M (forthcoming) Context and Aggregation: An Experimental Study of Bias and Discrimination in Organizational Decisions. *Organization Science*.
Clark CJ, Honeycutt N, Jussim L (2022) Replicability and the Psychology of Science. O'Donohue W, Masuda A, Lilienfeld S, eds. *Avoiding Questionable Research Practices in Applied Psychology* (Springer International Publishing, Cham), 45-71.
Cohen BP (1989). *Developing Sociological Knowledge: Theory and Method*, 2nd ed (Nelson-Hall, Chicago).
Cohen HF (1994). *The Scientific Revolution: A Historiographical Inquiry* (University of Chicago Press).
Coleman JS (1990). *Foundations of Social Theory* (Harvard University Press, Cambridge, MA).
Colquitt JA (2008) Publishing Laboratory Research in Amj: A Question of When, Not If. *Academy of Management Journal* 51(4):616-20.
Cook KS, Emerson RM, Gillmore MR, Yamagishi T (1983) The Distribution of Power in Exchange Networks: Theory and Experimental Results. *American Journal of Sociology* 89(2):275-305.
Cook KS, Yamagishi T (2008) A Defense of Deception on Scientific Grounds. *Social Psychology Quarterly* 71(3):215-21.
Cooper DJ, Hamman JR, Weber RA (2020) Fool Me Once: An Experiment on Credibility and Leadership. *Econ J* 130(631):2105-33.
Cooper DJ, Kagel JH, Lo W, Gu QL (1999) Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers. *Am Econ Rev* 89(4):781-804.
Cowen AP, Rink F, Cuypers IRP, Grégoire DA, Weller I (2022) Applying Coleman's Boat in Management Research: Opportunities and Challenges in Bridging Macro and Micro Theory. *Academy of Management Journal* 65(1):1-10.
Crano WD, Brewer MB, Lac A (2014). *Principles and Methods of Social Research*, 3 ed (Routledge, New York, NY).
Criscuolo P, Alexy O, Sharapov D, Salter A (2019) Lifting the Veil: Using a Quasi-Replication Approach to Assess Sample Selection Bias in Patent-Based Studies. *Strat Manag J* 40(2):230-52.
Criscuolo P, Dahlander L, Grohsjean T, Salter A (2017) Evaluating Novelty: The Role of Panels in the Selection of R&D Projects. *Acad Manage J* 60(2):433-60.
Crockett S, Smith VL, Wilson BJ (2009) Exchange and Specialisation as a Discovery Process. *Econ J* 119(539):1162-88.
Croson R, Anand J, Agarwal R (2007) Using Experiments in Corporate Strategy Research. *Eur Manag Rev* 4(3):173-81.
Cyert RM, Feigenbaum EA, March JG (1959) Models in a Behavioral Theory of the Firm. *Behavioral Science* 4(2):81-95.
Cyert RM, March JG, Starbuck WH (1961) Two Experiments on Bias and Conflict in Organizational Estimation. *Manag Sci* 7(3):254-64.
Dahlander L, Thomas A, Wallin MW, Ångström RC (Forthcoming) Blinded by the Person? Experimental Evidence from Idea Evaluation. *Strat Manag J* n/a(n/a).
Davis AM, Flicker B, Hyndman K, Katok E, Keppler S, Leider S, Long X et al. (2023) A Replication Study of Operations Management Experiments in Management Science. *Manag Sci*.
Davis GF, Marquis C (2005) Prospects for Organization Theory in the Early Twenty-First Century: Institutional Fields and Mechanisms. *Organization Science* 16(4):332-43.
De Vaus DA (1993). *Surveys in Social Research*, 3rd ed (UCL Press, London).
Delecourt S, Ng O (2021) Does Gender Matter for Small Business Performance? Experimental Evidence from India. *Working Paper*.

Delios A, Clemente EG, Wu T, Tan H, Wang Y, Gordon M, Viganola D et al. (2022) Examining the Generalizability of Research Findings from Archival Data. *Proc Natl Acad Sci USA* 119(30):e2120377119.
Denrell J, Kovács B (2015) The Effect of Selection Bias in Studies of Fads and Fashions. *PLOS ONE* 10(4):e0123471.
Denzin NK (1978). *The Research Act* (McGraw-Hill, New York, NY).
Di Stefano G, Gutierrez C (2019) Under a Magnifying Glass: On the Use of Experiments in Strategy Research. *Strategic Organization* 17(4):497-507.
Di Stefano G, Micheli MR (forthcoming) To Stem the Tide: Organizational Climate and the Locus of Knowledge Transfer. *Organization Science*.
Diamantopoulos A, Winklhofer HM (2001) Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research* 38(2):269-77.
DiMaggio P, Powell WW (1983) The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review* 48(2):147-60.
Dimitriadis S, Koning R (2022) Social Skills Improve Business Performance: Evidence from a Randomized Control Trial with Entrepreneurs in Togo. *Manag Sci* 68(12):8635-57.
Dipboye RL (1990) Laboratory Vs. Field Research in Industrial and Organizational Psychology. *International Review of Industrial and Organizational Psychology* 5:1-34.
Ebersole CR, Atherton OE, Belanger AL, Skulborstad HM, Allen JM, Banks JB, Baranski E et al. (2016) Many Labs 3: Evaluating Participant Pool Quality across the Academic Semester Via Replication. *J Exp Soc Psychol* 67:68-82.
Edwards AL (1957). *The Social Desirability Variable in Personality Assessment and Research* (Dryden Press, New York).
Edwards JR, Bagozzi RP (2000) On the Nature and Direction of Relationships between Constructs and Measures. *Psychological Methods* 5(2):155-74.
Ertug G, Maoret M (2020) Do Coaches in the National Basketball Association Actually Display Racial Bias? A Replication and Extension. *Academy of Management Discoveries* 6(2):206-34.
Evans JB, Schilke O (Forthcoming) The Power to Reward Vs. The Power to Punish: The Influence of Power Framing on Individual-Level Exploration. *Organ Sci*.
Falk A, Heckman JJ (2009) Lab Experiments Are a Major Source of Knowledge in the Social Sciences. *Science* 326(5952):535-8.
Fang C (2012) Organizational Learning as Credit Assignment: A Model and Two Experiments. *Organ Sci* 23(6):1717-32.
Felin T, Foss NJ, Ployhart RE (2015) The Microfoundations Movement in Strategy and Organization Theory. *Acad Manag Ann* 9(1):575-632.
Fine GA, Elsbach KD (2000) Ethnography and Experiment in Social Psychological Theory Building: Tactics for Integrating Qualitative Field Data with Quantitative Lab Data. *Journal of Experimental Social Psychology* 36(1):51-76.
Forstmeier W, Wagenmakers EJ, Parker TH (2017) Detecting and Avoiding Likely False-Positive Findings: A Practical Guide. *Biological Reviews* 92(4):1941-68.
Foss NJ, Pedersen T (2016) Microfoundations in Strategy Research. *Strategic Management Journal* 37(13):E22-E34.
Fréchette GR (2015) Laboratory Experiments: Professionals Versus Students. Fréchette GR, Schotter A, eds. *Handbook of Experimental Economic Methodology* (Oxford University Press, Oxford, UK).
Fréchette GR (2016) Experimental Economics across Subject Populations. Kagel JH, Roth AE, eds. *Handbook of Experimental Economics, Vol. 2* (Princeton University Press), 435-80.
Freese J, Peterson D (2017) Replication in Social Science. *Annu Rev Sociol* 43:147-65.
Galbraith JR (2012) The Future of Organization Design. *Journal of Organization Design* 1(1):3-6.
Gelman A, Loken E (2013). *The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No "Fishing Expedition" or "P-Hacking" and the Research Hypothesis Was Posited Ahead of Time* (Columbia University, New York).

Gelman A, Stern H (2006) The Difference between “Significant” and “Not Significant” Is Not Itself Statistically Significant. *Amer Statist* 60(4):328-31.
Gentzkow M, Kelly B, Taddy M (2019) Text as Data. *J Econ Lit* 57(3):535-74.
Ghosh M, Wu A (forthcoming) Iterative Coordination and Innovation: Prioritizing Value over Novelty. *Organization Science*.
Gibbons R (1999) Taking Coase Seriously. *Admin Sci Quart* 44(1):145-57.
Gibbons R, Grieder M, Herz H, Zehnder C (forthcoming) Building an Equilibrium: Rules Vs. Principles in Relational Contracts. *Organization Science*.
Gneezy U, List JA (2013). *The Why Axis: Hidden Motives and the Undiscovered Economics of Everyday Life* (Public Affairs, New York, NY,).
Goldenring JR (2010) Innocence and Due Diligence: Managing Unfounded Allegations of Scientific Misconduct. *Academic Medicine* 85(3).
Goldfarb B, King A (2016) Scientific Apophenia in Strategic Management Research: Significance Tests & Mistaken Inference. *Strat Manag J* 37(1):167-76.
Goldfarb B, Yan L (2021) Revisiting Zuckerman's (1999) Categorical Imperative: An Application of Epistemic Maps for Replication. *Strat Manag J* 42(11):1963-92.
Goldfarb B, Zavyalova A, Pillai S (2018) Did Victories in Certification Contests Affect the Survival of Organizations in the American Automobile Industry During 1895-1912? A Replication Study. *Strat Manag J* 39(8):2335-61.
Gorbatai A, Younkin P, Burtch G (forthcoming) Collateral Damage: The Relationship between High-Salience Events and Variation in Racial Discrimination. *Organization Science*.
Gould SJ (1988) The Streak of Streaks *New York Review*.
Grant AM, Wall TD (2009) The Neglected Science and Art of Quasi-Experimentation: Why-to, When-to, and How-to Advice for Organizational Researchers. *Organ Res Meth* 12(4):653-86.
Grégoire DA, Binder JK, Rauch A (2019) Navigating the Validity Tradeoffs of Entrepreneurship Research Experiments: A Systematic Review and Best-Practice Suggestions. *J Bus Venturing* 34(2):284-310.
Greve HR (2018) Show Me the Data! Improving Evidence Presentation for Publication. *Manag Organ Rev* 14(2):423-32.
Guarnaschelli S, McKelvey RD, Palfrey TR (2000) An Experimental Study of Jury Decision Rules. *American Political Science Review* 94(02):407-23.
Guetzkow H, Bowes AE (1957) The Development of Organizations in a Laboratory. *Manag Sci* 3(4):380-402.
Gunst RF, Mason RL (2009) Fractional Factorial Design. *Wiley Interdisciplinary Reviews Computational Statistics* 1(2):234-44.
Guzman J, Kacperczyk AO (2019) Gender Gap in Entrepreneurship. *Research Policy* 48(7):1666-80.
Haack P, Schilke O, Zucker LG (2021) Legitimacy Revisited: Disentangling Propriety, Validity, and Consensus. *Journal of Management Studies*. 58(3):749-81.
Haack P, Sieweke J, Wessel L (2019) Microfoundations and Multi-Level Research on Institutions. *Research in the Sociology of Organizations* 65A:11-40.
Hamman J, Rick S, Weber RA (2007) Solving Coordination Failure with “All-or-None” Group-Level Incentives. *Exp Econ* 10(3):285-303.
Hamman JR, Martínez-Carrasco MA (forthcoming) Managing Uncertainty: An Experiment on Delegation and Team Selection. *Organization Science*.
Hannan MT. (2022, Sep. 13). Rethinking Organizational Ecology in Light of Developments in Cognitive Science and Natural-Language Processing. Retrieved Sep. 11, 2022, from https://doi.org/10.31235/osf.io/vmuan
Harrison GW, List JA (2004) Field Experiments. *J Econ Lit* 42(4):1009-55.
Haslam SA, McGarty C (2004) Experimental Design and Causality in Social Psychological Research. Sansone C, Morf CC, Panter AT, eds. *The Sage Handbook of Methods in Social Psychology* (Sage, Thousand Oaks, CA), 237-64.

Heath C, Sitkin SB (2001) Big-B Versus Big-O: What Is Organizational About Organizational Behavior? *Journal of Organizational Behavior* 22(1):43-58.
Hedström P, Swedberg R (1998). *Social Mechanisms: An Analytical Approach to Social Theory* (Cambridge University Press, New York, NY).
Hegelmaier FC (1852) Über Das Gedächtnis Für Linearanschauungen (on Memory for Visually Perceived Lines). *Archiv für physiologische Heilkunde* 11:844-53.
Henrich J, Boyd RT, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R (2001) In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *Am Econ Rev* 91(2):73.
Henrich J, McElreath R, Barr A, Ensminger J, Barrett C, Bolyanatz A, Cardenas JC et al. (2006) Costly Punishment across Human Societies. *Science* 312(5781):1767-70.
Hergueux J, Henry E, Benkler Y, Algan Y (forthcoming) Social Exchange and the Reciprocity Roller Coaster: Evidence from the Life and Death of Virtual Teams. *Organization Science*.
Heugens PPMAR (2008) Organization Theory: Bright Prospects for a Permanently Failing Field *ERIM Inaugural Address Series*. Erasmus Research Institute of Management, Rotterdam.
Hill AD, Johnson SG, Greco LM, O'Boyle EH, Walter SL (2021) Endogeneity: A Review and Agenda for the Methodology-Practice Divide Affecting Micro and Macro Research. *J Manage* 47(1):105-43.
Hillygus DS (2018) Navigating Scholarly Exchange in Today's Media Environment. *Journal of Politics* 80(3):1064-8.
Holt CA, Laury SK (2002) Risk Aversion and Incentive Effects. *Am Econ Rev* 92(5):1644-55.
Holt CA, Laury SK (2005) Risk Aversion and Incentive Effects: New Data without Order Effects. *Am Econ Rev* 95(3):902-12.
Howard-Grenville J, Davis GF, Dyllick T, Miller CC, Thau S, Tsui AS (2019) Sustainable Development for a Better World: Contributions of Leadership, Management, and Organizations. *Academy of Management Discoveries* 5(4):355-66.
Hyman SI (1970) Management: An Experimental Science? *Management International Review* 10(6):51-9.
Imbens GW (2010) Better Late Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). *J Econ Lit* 48(2):399-423.
Innocenti A (2017) Virtual Reality Experiments in Economics. *Journal of Behavioral and Experimental Economics* 69:71-7.
Jackson M, Cox DR (2013) The Principles of Experimental Design and Their Application in Sociology. *Annual Review of Sociology* 39(1):27-49.
Jick TD (1979) Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly* 24(4):602-11.
Jost JT, Kruglanski AW (2002) The Estrangement of Social Constructionism and Experimental Social Psychology: History of the Rift and Prospects for Reconciliation. *Personality and Social Psychology Review* 6(3):168-87.
Kacperczyk AO, Younkin P (2022a) A Founding Penalty: Evidence from an Audit Study on Gender, Entrepreneurship, and Future Employment. *Organization Science* 33(2):716-45.
Kacperczyk AO, Younkin P, Rocha V (2023) Do Employees Work Less for Female Leaders? A Multi-Method Study of Entrepreneurial Firms. *Organization Science* 34(3):1111-33.
Kacperczyk O, Younkin P (2022b) A Founding Penalty: Evidence from an Audit Study on Gender, Entrepreneurship, and Future Employment. *Organization Science* 33(2):716-45.
Kaibel C, Biemann T (2021) Rethinking the Gold Standard with Multi-Armed Bandits: Machine Learning Allocation Algorithms for Experiments. *Organizational Research Methods* 24(1):78-103.
Kanazawa S (1999) Using Laboratory Experiments to Test Theories of Corporate Behavior. *Ration Soc* 11(4):443-61.
Kanze D, Huang L, Conley MA, Higgins ET (2018) We Ask Men to Win and Women Not to Lose: Closing the Gender Gap in Startup Funding. *Academy of Management Journal* 61(2):586-614.
Kaplan S (2015) Mixing Quantitative and Qualitative Research. Elsbach KD, Kramer RM, eds. *Handbook of Qualitative Organizational Research* (Routledge, New York, NY), 423-33.

Kasy M (2021) Of Forking Paths and Tied Hands: Selective Publication of Findings, and What Economists Should Do About It. *J Econ Perspect*.
Kerlinger FN, Lee HB (2000). *Foundations of Behavioral Research*, 4 ed (Harcourt, Fort Worth, TX).
Kerr NL (1998) Harking: Hypothesizing after the Results Are Known. *Pers Soc Psychol Rev* 2(3):196-217.
Kim H (2020) The Value of Competitor Information: Evidence from a Field Experiment. *Working Paper*.
King A, Goldfarb B, Simcoe T (2021) Learning from Testimony on Quantitative Research in Management. *Acad Manage Rev* 46(3):465-88.
King EB, Hebl MR, Botsford Morgan W, Ahmad AS (2013) Field Experiments on Sensitive Organizational Topics. *Organizational Research Methods* 16(4):501-21.
Klein R, Vianello M, Hasselman F, Adams B, Reginald B. Adams J, Alper S, Aveyard M et al. (2018) Many Labs 2: Investigating Variation in Replicability across Sample and Setting. *Working Paper*.
Klingebiel R (2018) Risk-Type Preference Shifts in Response to Performance Feedback. *Strat Organ* 16(2):141-66.
Klingebiel R (2022) Motivating Innovation: Tunnels Vs. Funnels. *Strategy Science* 7(4):300-16.
Knudsen T, Levinthal DA, Puranam P (2019) A Model Is a Model. *Strategy Science* 4(1):1-3.
Kotha R, Vissa B, Lin Y, Corboz A-V (2023) Do Ambitious Entrepreneurs Benefit More from Training? *Strat Manag J* 44(2):549-75.
Kraatz MS, Zajac EJ (1996) Exploring the Limits of the New Institutionalism: The Causes and Consequences of Illegitimate Organizational Change. *Amer Sociol Rev* 61(5):812-36.
Krishna A (2021) The Need for Synergy in Academic Policies: An Introduction to the Dialogue on Pre-Registration. *Journal of Consumer Psychology* 31(1):146-50.
Kriss PH, Weber RA (2013) Organizational Formation and Change: Lessons from Economic Laboratory Experiments. Grandori A, ed. *Handbook of Economic Organization* (Edward Elgar, Northampton, MA), 245–69.
Lampel J, Shamsie J, Shapira Z (2009) Experiencing the Improbable: Rare Events and Organizational Learning. *Organization Science* 20(5):835-45.
Laureiro-Martinez D, Arrieta J, Brusoni S (Forthcoming) Microfoundations of Problem Solving: The Role of Attentional Engagement in Predicting Problem-Solving Strategies. *Organ Sci*.
Laureiro-Martínez D, Brusoni S, Canessa N, Zollo M (2015) Understanding the Exploration–Exploitation Dilemma: An fMRI Study of Attention Control and Decision-Making Performance. *Strat Manag J* 36(3):319-38.
Laureiro-Martínez D, Brusoni S, Zollo M (2010) The Neuroscientific Foundations of the Exploration−Exploitation Dilemma. *Journal of Neuroscience, Psychology, and Economics* 3(2):95.
Lee A, Inceoglu I, Hauser O, Greene M (2022) Determining Causal Relationships in Leadership Research Using Machine Learning: The Powerful Synergy of Experiments and Data Science. *Leader Q* 33(5):101426.
Levin RC, Klevorick AK, Nelson RR, Winter SG, Gilbert R, Griliches Z (1987) Appropriating the Returns from Industrial Research and Development; Comments and Discussion. *Brookings Pap Econ Activ* 3:783-831.
Levine SS (2018) Show Us Your Data: Connect the Dots, Improve Science. *Manag Organ Rev* 14(2):433-7.
Levine SS, Apfelbaum EP, Bernard M, Bartelt VL, Zajac EJ, Stark D (2014) Ethnic Diversity Deflates Price Bubbles. *Proc Natl Acad Sci USA* 111(52):18524–9.
Levine SS, Bernard M, Nagel R (2017) Strategic Intelligence: The Cognitive Capability to Anticipate Competitor Behavior. *Strat Manag J* 38(12):2390-423.
Levine SS, Reypens C, Riedl C, Mauersberger F (2021a) What Directs Search? Behavioral Evidence on the Role of Performance Stability. *Acad Manag Proc* 2021.
Levine SS, Reypens C, Stark D (2021b) Racial Attention Deficit. *Sci Adv* 7(38):abg9508.
Levine SS, Zajac EJ (2023) The Other Invisible Hand: How Markets — as Institutions — Propagate Conformity and Valuation Errors. *Strategy Science* 8(3):323-404, C3.

Levitt SD, List JA (2011) Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *American Economic Journal: Applied Economics* 3(1):224-38.
Lewin AY, Chiu C-Y, Fey CF, Levine SS, McDermott G, Murmann JP, Tsang E (2016) The Critique of Empirical Social Science. *Manag Organ Rev* 12(4):649-58.
Lin H, Werner KM, Inzlicht M (2021) Promises and Perils of Experimentation: The Mutual-Internal-Validity Problem. *Perspectives on Psychological Science* 16(4):854-63.
List J (2023) A Summary of Framed Field Experiments on Fieldexperiments.Com: The Who's, What's, Where's, and When's. *Working Paper*.
List J, Sadoff S, Wagner M (2011) So You Want to Run an Experiment, Now What? Some Simple Rules of Thumb for Optimal Experimental Design. *Exp Econ*:1-19.
List JA (2011) Why Economists Should Conduct Field Experiments and 14 Tips for Pulling One Off. *J Econ Perspect* 25(3):3-16.
List JA, Lucking-Reiley D (2002) The Effects of Seed Money and Refunds on Charitable Giving: Experimental Evidence from a University Capital Campaign. *J Polit Econ* 110(1):215-33.
Lonati S, Quiroga BF, Zehnder C, Antonakis J (2018) On Doing Relevant and Rigorous Experiments: Review and Recommendations. *Journal of Operations Management* 64:19-40.
Lucas JW (2003a) Status Processes and the Institutionalization of Women as Leaders. *Amer Sociol Rev* 68(3):464-80.
Lucas JW (2003b) Theory-Testing, Generalization, and the Problem of External Validity. *Sociological Theory* 21(3):236–53.
Lude M, Prügl R (2021) Experimental Studies in Family Business Research. *Journal of Family Business Strategy* 12(1):100361.
Lykken DT (1968) Statistical Significance in Psychological Research. *Psychological Bulletin* 70(3):151-9.
Lynch JG (1982) On the External Validity of Experiments in Consumer Research. *Journal of Consumer Research* 9(3):225-39.
MacKinnon DP, Fairchild AJ, Fritz MS (2007) Mediation Analysis. *Annual Review of Psychology* 58:593-614.
Maniadis Z, Tufano F, List JA (2014) One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *Am Econ Rev* 104(1):277-90.
March JG (1991) Exploration and Exploitation in Organizational Learning. *Organ Sci* 2(1):71-87.
March JG (2004) Parochialism in the Evolution of a Research Community: The Case of Organization Studies. *Manag Organ Rev* 1(1):5-22.
March JG, Simon HA (1993) Introduction to the Second Edition. March JG, Simon HA, eds. *Organizations*, 2 ed (Blackwell, Cambridge, MA), 1-19.
Mathieu JE, DeShon RP, Bergh DD (2008) Mediational Inferences in Organizational Research: Then, Now, and Beyond. *Organ Res Meth* 11(2):203-23.
Maxwell SE, Lau MY, Howard GS (2015) Is Psychology Suffering from a Replication Crisis? What Does "Failure to Replicate" Really Mean? *American Psychologist* 70(6):487.
Mayo E (1933). *The Human Problems of an Industrial Civilization* (Macmillan, New York, NY).
Mayr E (1961) Cause and Effect in Biology. *Science* 134(3489):1501-6.
McDermott R (2011) Internal and External Validity. Lupia A, Greene DP, Kuklinski JH, Druckman JN, eds. *Cambridge Handbook of Experimental Political Science* (Cambridge University Press, Cambridge), 27-40.
McGrath JE (1981) Dilemmatics: The Study of Research Choices and Dilemmas. *Amer Behav Sci* 25(2):154-79.
Milkman KL, Gromet D, Ho H, Kay JS, Lee TW, Pandiloski P, Park Y et al. (2021) Megastudies Improve the Impact of Applied Behavioural Science. *Nature* 600(7889):478-83.
Mill JS (1874). *A System of Logic (Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation)*, 8th ed (Harper & Brothers, New York).

Mitchell G (2012) Revisiting Truth or Triviality: The External Validity of Research in the Psychological Laboratory. *Perspectives on Psychological Science* 7(2):109-17.
Mitzkewitz M, Nagel R (2020) Reinhard Selten. Frantz R, ed. *The Beginnings of Behavioral Economics* (Academic Press), 57-9.
Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington H, Houts R et al. (2011) A Gradient of Childhood Self-Control Predicts Health, Wealth, and Public Safety. *Proc Natl Acad Sci USA*.
Molina M, Nee V, Holm H (forthcoming) Cooperation with Strangers: Spillover of Community Norms. *Organization Science*.
Mook DG (1983) In Defense of External Invalidity. *American Psychologist* 38:379-87.
Moore DA (2016) Preregister If You Want To. *Amer Psychol* 71(3):238-9.
Moszkowski A (1970). *Conversations with Einstein* (Horizon Press, New York).
Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, Simonsohn U et al. (2017) A Manifesto for Reproducible Science. *Nat Hum Behav* 1:0021.
Murphy KR, Aguinis H (2019) Harking: How Badly Can Cherry-Picking and Question Trolling Produce Bias in Published Results. *Journal of Business and Psychology* 34.
Nagel R, Bayona A, Kheirandish R, Mousavi S (2016) Reinhard Selten, the Dualist *Routledge Handbook of Behavioral Economics* (Routledge), 66-87.
Nai J, Lin Y, Kotha R, Vissa B (2022) A Foot in the Door: Field Experiments on Entrepreneurs' Network Activation Strategies for Investor Referrals. *Strat Manag J* 43(2):323-39.
Naumovska I, Gaba V, Greve HR (2021) The Diffusion of Differences: A Review and Reorientation of 20 Years of Diffusion Research. *Acad Manag Ann* 15(2):377-405.
Newell A, Simon HA (1972). *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, N.J.).
Newton I (1999 [1713]). *Philosophiae Naturalis Principia Mathematica, General Scholium*, Second ed (University of California Press).
Nobel Prize. (1994). Reinhard Selten – Biographical. from https://www.nobelprize.org/prizes/economic-sciences/1994/selten/lecture/
Nobel Prize. (2002). Vernon L. Smith - Facts. from https://www.nobelprize.org/prizes/economic-sciences/2002/smith/facts/
Nosek B (2023) A Commitment to Transparency by Researchers, Journals, Funders, and Institutions Will Not Prevent Fraud. But, It Will Make Fraud More Inconvenient and Easier to Detect., 14:06 ed. Twitter.
Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, Buck S et al. (2015) Promoting an Open Research Culture. *Science* 348(6242):1422-5.
Nosek BA, Ebersole CR, DeHaven AC, Mellor DT (2018) The Preregistration Revolution. *Proc Natl Acad Sci USA* 115(11):2600-6.
Open Science Collaboration (2015) Estimating the Reproducibility of Psychological Science. *Science* 349(6251):aac4716.
Orne MT (1962) On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications. *Amer Psychol* 17(11):776–83.
Ortmann A, Hertwig R (2002) The Costs of Deception: Evidence from Psychology. *Exp Econ* 5(2):111-31.
Ostrom E (1990). *Governing the Commons* (Cambridge University Press, Cambridge, UK).
Ostrom E, Gardner R, Walker J (1994). *Rules, Games, and Common-Pool Resources* (University of Michigan Press, Ann Arbor, Michigan).
Palan S, Schitter C (2018) Prolific.Ac—a Subject Pool for Online Experiments. *Journal of Behavioral and Experimental Finance* 17:22-7.
Pamphile VD, Ruttan RL (forthcoming) The (Bounded) Role of Stated-Lived Value Congruence and Authenticity in Employee Evaluations of Organizations. *Organization Science*.
Park S, Gonzalez C, Puranam P (2022) Decision Centralization and Learning from Experience in Groups: Separating Context from Aggregation Effects. *Manag Sci* Forthcoming.
Pedulla DS (2016) Penalized or Protected? Gender and the Consequences of Nonstandard and Mismatched Employment Histories. *American Sociological Review* 81(2):262-89.

Perdue BC, Summers JO (1986) Checking the Success of Manipulations in Marketing Experiments. *Journal of marketing research* 23(4):317-26.
Pfeffer J (1993) Barriers to the Advance of Organizational Science: Paradigm Development as a Dependent Variable. *Academy of Management Review* 18(4):599-620.
Piezunka H, Schilke O (2023) The Dual Function of Organizational Structure: Aggregating and Shaping Individuals' Votes. *Organ Sci*.
Piezunka H, Schilke O (Forthcoming) The Dual Function of Organizational Structure: Aggregating and Shaping Individuals' Votes. *Organization Science*.
Pittnauer S, Hohnisch M, Ostermaier A, Pfingsten A (forthcoming) Effects of Social Information on Risk Taking and Performance: Understanding Others' Decisions Vs. Comparing Oneself with Others in Short-Term Performance. *Organization Science*.
Podsakoff NP, Podsakoff PM, MacKenzie SB, Klinger RL (2013) Are We Really Measuring What We Say We're Measuring? Using Video Techniques to Supplement Traditional Construct Validation Procedures. *Journal of Applied Psychology* 98:99-113.
Podsakoff PM, Podsakoff NP (2019) Experimental Designs in Management and Leadership Research: Strengths, Limitations, and Recommendations for Improving Publishability. *Leadership Quarterly* 30(1):11-33.
Polit DF, Beck CT (2001). *Essentials of Nursing Research: Methods, Appraisal, and Utilization* (Lippincott Williams & Wilkins, Philadelphia, PA).
Pollock TG, Lashley K, Rindova VP, Han J-H (2019) Which of These Things Are Not Like the Others? Comparing the Rational, Emotional, and Moral Aspects of Reputation, Status, Celebrity, and Stigma. *Academy of Management Annals* 13(2):444-78.
Pratt MG, Sarah K, Richard W (2020) The Tumult over Transparency: Decoupling Transparency from Replication in Establishing Trustworthy Qualitative Research. *Admin Sci Quart* 65:1-19.
Puranam P (2018). *The Microstructure of Organizations* (Oxford University Press, Oxford).
Puranam P. (2023). How to Rapidly Test New Organisation Designs. Retrieved May 30, 2023, from https://knowledge.insead.edu/leadership-organisations/how-rapidly-test-new-organisation-designs
Puranam P, Stieglitz N, Osman M, Pillutla MM (2015) Modelling Bounded Rationality in Organizations: Progress and Prospects. *Acad Manag Ann* 9(1):337-92.
Radoynovska N, Ruttan R (forthcoming) A Matter of Transition: Authenticity Judgments and Attracting Employees to Hybridized Organizations. *Organization Science*.
Rahmandad H, Gary MS (forthcoming) Delays Impair Learning and Can Drive Convergence to Inefficient Strategies. *Organization Science*.
Reiff JS, Zhang JC, Gallus J, Dai H, Pedley NM, Vangala S, Leuchter RK et al. (2022) When Peer Comparison Information Harms Physician Well-Being. *Proc Natl Acad Sci USA* 119(29):e2121730119.
Reypens C, Levine SS (2017) To Grasp Cognition in Action, Combine Behavioral Experiments with Protocol Analysis. Sund KJ, Hodgkinson GP, Galavan RJ, eds. *Methodological Challenges and Advances in Managerial and Organizational Cognition* (Emerald, Bingley, UK), 123-46.
Reypens C, Levine SS (2018) Behavior in Behavioral Strategy: Capturing, Measuring, Analyzing. Augier M, Fang C, Rindova V, eds. *Behavioral Strategy in Perspective* (Emerald), 221-46.
Rivera LA, Tilcsik A (2016) Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market. *American Sociological Review* 81(6):1097-131.
Roethlisberger FJ, Dickson WJ (1939). *Management and the Worker* (Harvard University Press, Cambridge, MA).
Roth AE (1986) Laboratory Experimentation in Economics. *Economics & Philosophy* 2(2):245-73.
RRBM C-fo (2020) A Vision for Responsible Research in Business and Management: Striving for Useful and Credible Knowledge.
Rubinstein A (2001) A Theorist's View of Experiments. *European Economic Review* 45(4):615-28.
Ryan JC, Tipu SA (2022) Business and Management Research: Low Instances of Replication Studies and a Lack of Author Independence in Replications. *Research Policy* 51(1):104408.

Sauermann H, Selten R (1959) Ein Oligopolexperiment. *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*(H. 3):427-71.
Schilke O (2018) A Micro-Institutional Inquiry into Resistance to Environmental Pressures. *Academy of Management Journal* 61(4):1431-66.
Schilke O, Powell A, Schweitzer M (Forthcoming) A Review of Experimental Research on Organizational Trust. *Journal of Trust Research*.
Schilke O, Powell A, Schweitzer ME (2023) A Review of Experimental Research on Organizational Trust. *Journal of Trust Research* 13(2):102-39.
Schram A (2005) Artificiality: The Tension between Internal and External Validity in Economic Experiments. *Journal of Economic Methodology* 12(2):225-37.
Schwab A, Starbuck W (2017) A Call for Openness in Research Reporting: How to Turn Covert Practices into Helpful Tools. *Acad Manag Learn Educ* 16(1):125-41.
Schwenk CR (1982) Why Sacrifice Rigour for Relevance? A Proposal for Combining Laboratory and Field Research in Strategic Management. *Strategic Management Journal* 3(3):213-25.
Shah R. (2023). Field Experiments Rap. Retrieved June 1, 2023, from https://youtu.be/Z7feq-zl9SI
Shaver JM (1998) Accounting for Endogeneity When Assessing Strategy Performance: Does Entry Mode Choice Affect Fdi Survival? *Manag Sci* 44(4):571-85.
Shrestha YR, He VF, Puranam P, Krogh Gv (2020) Algorithm Supported Induction for Building Theory: How Can We Use Prediction Models to Theorize? *Organ Sci* 32(3):527-908.
Shrout PE, Rodgers JL (2018) Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual review of Psychology* 69:487-510.
Siegfried T (2022) Louis Pasteur's Devotion to Truth Transformed What We Know About Health and Disease. *Science News* 202(9):22-6.
Simon HA (1989) The Scientist as Problem Solver. Klahr D, Kotovsky K, eds. *Complex Information Processing: The Impact of Herbert A. Simon* (Lawrence Erlbaum Associates, Hillsdale, N.J.), 375-98.
Simon HA, Guetzkow H (1952) Research into Behavior in Organizations - Proposed Program. Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Penn.
Simonsohn U, Simmons JP, Nelson LD (2015) Better P-Curves: Making P-Curve Analysis More Robust to Errors, Fraud, and Ambitious P-Hacking, a Reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General* 144(6):1146-52.
Smith EB, Rand W (2018) Simulating Macro-Level Effects from Micro-Level Observations. *Manag Sci* 64(11):5405-21.
Smith VL (1962) An Experimental Study of Competitive Market Behavior. *J Polit Econ* 70(2):111-37.
Smith VL (1976) Experimental Economics: Induced Value Theory. *Am Econ Rev* 66(2):274-9.
Smith VL (1982) Microeconomic Systems as an Experimental Science. *Am Econ Rev* 72(5):923-55.
Smith VL (1991) Experimental Economics at Purdue. Smith VL, ed. *Papers in Experimental Economics* (Cambridge University Press, Cambridge), 154-8.
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):845-51.
Starbuck WH (2016) 60th Anniversary Essay: How Journals Could Improve Research Practices in Social Science. *Admin Sci Quart* 61(2):165-83.
Starr E, Goldfarb B (2020) Binned Scatterplots: A Simple Tool to Make Research Easier and Better. *Strat Manag J* 41(12):2261-74.
Steinmetz J, Bennett C, Håkonsson DD (2012) A Practitioner's View of the Future of Organization Design: Future Trends and Implications for Royal Dutch Shell. *Journal of Organization Design* 1(1):7-11.
Stevens ST, Jussim L, Honeycutt N (2020) Scholarship Suppression: Theoretical Perspectives and Emerging Trends. *Societies* 10(4):82.
Stevenson RM, Josefy M (2019) Knocking at the Gate: The Path to Publication for Entrepreneurship Experiments through the Lens of Gatekeeping Theory. *Journal of Business Venturing* 34(2):242-60.

Stolte JF, Fine GA, Cook KS (2001) Sociological Miniaturism: Seeing the Big through the Small in Social Psychology. *Annu Rev Sociol* 27(1):387-413.
Sutton RI, Staw BM (1995) What Theory Is Not. *Admin Sci Quart* 40(3):371-84.
Tavory I, Timmermans S (2014). *Abductive Analysis: Theorizing Qualitative Research* (University of Chicago Press, Chicago, IL).
Thébaud S (2010) Gender and Entrepreneurship as a Career Choice: Do Self-Assessments of Ability Matter? *Social Psychology Quarterly* 73(3):288-304.
Tilcsik A (2010) From Ritual to Reality: Demography, Ideology, and Decoupling in a Post-Communist Government Agency. *Academy of Management Journal* 53(6):1474-98.
Triplett N (1898) The Dynamogenic Factors in Pacemaking and Competition. *American Journal of Psychology* 9(4):507-33.
Tsui AS (2021) Responsible Research and Responsible Leadership Studies. *Academy of Management Discoveries* 7(2):166-70.
Turing AM (1950) Computing Machinery and Intelligence. *Mind* LIX(236):433-60.
Tversky A, Kahneman D (1982) Judgments of and by Representativeness. Tversky A, Kahneman D, Slovic P, eds. *Judgment under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge), 84-98.
Tversky A, Kahneman D (1983) Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychol Rev* 90:293-315.
Tversky A, Kahneman D (1992) Advances in Prospect Theory: Cumulative Representation of Uncertainty. *J Risk Uncertainty* 5(4):297-323.
van 't Veer AE, Giner-Sorolla R (2016) Pre-Registration in Social Psychology—a Discussion and Suggested Template. *Journal of Experimental Social Psychology* 67:2-12.
van den Broek T, Langley DJ, Ehrenhard ML, Groen A (forthcoming) When Do Evaluators Publicly Express Their Legitimacy Judgments? An Inquiry into the Role of Peer Endorsement and Evaluative Mode. *Organization Science*.
van Dongen J (2010) On the Method of Theoretical Physics *Einstein's Unification* (Cambridge University Press, Cambridge), 36-57.
Vanhove AJ, Harms PD (2015) Reconciling the Two Disciplines of Organisational Science: A Comparison of Findings from Lab and Field Research. *Applied Psychology* 64(4):637-73.
Wager S, Athey S (2018) Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. *J Amer Statist Assn* 113(523):1228-42.
Wang CL, Whitson JA, King BG, Ramirez RL (forthcoming) Social Movements, Collective Identity, and Workplace Allies: The Labeling of Gender Equity Policy Changes. *Organization Science*.
Wang W, Lumineau F, Schilke O (2022). *Blockchains: Strategic Implications for Contracting, Trust, and Organizational Design* (Cambridge University Press, Cambridge).
Wasserstein RL, Schirm AL, Lazar NA (2019) Moving to a World Beyond "P < 0.05". *Amer Statist* 73(sup1):1-19.
Weber J (1992) Scenarios in Business Ethics Research: Review, Critical Assessment, and Recommendations. *Business Ethics Quarterly* 2(2):137-60.
Weber R, Camerer C, Rottenstreich Y, Knez M (2001) The Illusion of Leadership: Misattribution of Cause in Coordination Games. *Organ Sci* 12(5):582-98.
Weber RA, Camerer CF (2003) Cultural Conflict and Merger Failure: An Experimental Approach. *Manag Sci* 49(4):400-15.
Webster M, Jr., Sell J (2014) Why Do Experiments? Webster M, Jr., Sell J, eds. *Laboratory Experiments in the Social Sciences*, 2nd ed (Academic Press).
Weick KE (1965) Laboratory Experimentation with Organizations. March JG, ed. *Handbook of Organizations* (Rand McNally), 194-260.
Weick KE (1969) Laboratory Organizations and Unnoticed Causes. *Admin Sci Quart* 14(2):294-303.

Weinberg JD, Freese J, McElhattan D (2014) Comparing Data Characteristics and Results of an Online Factorial Survey between a Population-Based and a Crowdsource-Recruited Sample. *Sociological Science* 1(19):292-310.
Wilde LL. 1981. On the Use of Laboratory Experiments in Economics *Philosophy in Economics: Papers Deriving from and Related to a Workshop on Testability and Explanation in Economics held at Virginia Polytechnic Institute and State University, 1979*. Springer, 137-48.
Wilson BJ (2017) How We Think About Economics. *South Econ J* 83(3):655-7.
Wilson TD, Aronson E, Carlsmith K (2010) The Art of Laboratory Experimentation. Fiske ST, Gilbert DT, Lindzey G, eds. *Handbook of Social Psychology* (McGraw-Hill, Boston, MA), 51-81.
Yang T, Aldrich HE (2012) Out of Sight but Not out of Mind: Why Failure to Account for Left Truncation Biases Research on Failure Rates. *J Bus Venturing* 27(4):477-92.
Zelditch M, Jr. (1969) Can You Really Study an Army in a Laboratory? Etzioni A, Lehman EW, eds. *A Sociological Reader on Complex Organizations* (Holt, Rinehart, and Winston, New York, NY), 528-39.
Zelditch M, Jr. (2014) Laboratory Experiments in Sociology. Webster M, Jr., Sell J, eds. *Laboratory Experiments in the Social Sciences*, 2nd ed (Academic Press).
Zucker LG (1977) The Role of Institutionalization in Cultural Persistence. *Amer Sociol Rev* 42(5):726-43.