Under a magnifying glass: On the use of experiments in strategy research

Giada Di Stefano
HEC Paris, France

Cédric Gutierrez
Bocconi University, Italy

Abstract
The rate at which experimental studies are published in the field of strategy has steadily increased over the past few years. Still, experimental papers account for only a small fraction of strategy papers. This may not come as a surprise given the skepticism surrounding the experimental method, which is often seen as uninterested in establishing external validity, and too “micro” for a field in which the level of analysis is primarily organizational and inter-organizational. Is this skepticism founded? To what extent can experiments be a useful tool for strategy research? To answer this question, we start by examining experimental strategy papers published between 1980 and 2016. Results from the analysis alleviate doubts about the suitability of experimental methods for the study of questions of strategic interest to firms. We next discuss the main advantages associated with the use of experiments and why they make strategy an exciting field in which to be an experimentalist today.

Keywords
apophenia, behavioral strategy, causal inference, decision-making, experiments, microfoundations

Introduction
It was 1982 when the Strategic Management Journal (SMJ) published a piece titled “Why sacrifice rigour for relevance? A proposal for combining laboratory and field research in strategic management.” In this article, Schwenk (1982) tried to counter common objections to the scant use of laboratory experiments in the field of strategy and made the case that experimental and observational data were complementary. Were we to judge by the rate at which strategy papers using randomized
laboratory or field experiments (simply “experiments” from now on) have been published since then, we would dismiss Schwenk’s point as outdated: Starting from the year 2000, data show a steady increase in the number of strategy publications using experiments (see Figure 1). Still, these manuscripts account for only a small fraction of the papers published in strategy; their use to date peaked at 5% of all papers published in SMJ in 2015.

This may not come as a surprise given widely diffused skepticism about the experimental method, which is often seen as uninterested in establishing external validity and too “micro” for a field in which the level of analysis is mostly organizational and inter-organizational (Croson et al., 2007). As experimentalists in strategy, we may be tempted to dismiss this criticism as unfounded; after all, experiments can also be used to answer “macro” strategy questions and their design can

Figure 1. Experimental strategy papers as a percentage of total papers published. Experimental strategy papers as a percentage of the total number of papers published in the seven leading management journals that publish empirical research (Academy of Management Journal, Administrative Science Quarterly, Journal of Management, Journal of Management Studies, Management Science, Organization Science, and Strategic Management Journal). The upper plot shows the total and the bottom plot is per journal (using a 5-year rolling window). Academy of Management Journal, Administrative Science Quarterly, Organization Science, and Strategic Management Journal appear to be leaders in publishing experimental research.
be tweaked to make findings more generalizable. To us, however, rebutting this criticism is not really the point: We need more experiments in strategy, we will argue, because they effectively address some fundamental needs of our field. Not only do experiments help us answer questions of strategic interest for firms but also they enable us to clearly assess causality and precisely measure the constructs of interest.

To develop our arguments, we start by examining how experiments have been used in the field of strategy using a co-citation analysis of papers published between 1980 and 2016. We specifically focus on the topic areas and research designs of extant experimental strategy papers in order to discuss the criticism described above. We next elaborate the advantages that experiments provide relative to other methods, and why these advantages make them so important for the development of our field.

What Experiments Are (Maybe) Not So Good At

Two common critiques to the use of experiments in strategy research relate to the fact that experiments cannot explain organizational-level outcomes (i.e. their level of analysis is too “micro”) and their results cannot be generalized beyond the specific context, task, and individuals under examination (i.e. they lack “external validity”). To what extent is this really the case? To answer to this question, we performed a co-citation analysis of all strategy papers that used a randomized (laboratory or field) experiment and were published in top management journals between 1980 and 2016 (see Supplementary Appendix). By examining the topic areas these papers cover and the experimental designs they adopt, we gain insights into the level of analysis and external validity of prior experimental work in our field.

With respect to the “micro” objection, we observe that strategy experiments have been used to answer questions about processes and outcomes not only at the individual level but also at the organizational level. Of course, the study of individual-level processes and outcomes is a privileged domain of application for experiments. In particular, in our sample, individual-level studies focus on how organization design elements are negotiated within the organization and ultimately impact workers’ motivation, productivity, and performance: Examples include Jordan (1986) examining the effects of extrinsic rewards on intrinsic motivation, Welsh et al. (1993) studying the effect of strategic human resource management on productivity, and Stajkovic and Luthans (2001) investigating the effects of incentives on performance. As these examples clearly show, despite being at the individual level, these studies have an important organizational component in that they examine individual-level variables in the context of organizations. Even when the outcomes under examination are at the individual level (e.g. productivity or motivation), they have clear effects on organizational-level outcomes (ultimately, performance). The same holds true for another topic that our analysis found was frequently investigated using experiments, namely, the effect of managerial and entrepreneurial decision-making on firm-level outcomes. This is the case, for instance, of Sutcliffe and Zaheer (1998) examining the effect of uncertainty on the decision to vertically integrate or Huang and Pearce (2015) investigating the role of “gut feel” in investment decisions. Once again, the emphasis is on individual-level decision-making, but the individuals examined are key decision makers whose choices have a clear impact on organizational-level outcomes. A final and perhaps less intuitive domain of application for experiments is the study of processes by which individuals, and the knowledge they possess, are integrated at the organizational level. Several papers in this domain focus on the integration process that follows mergers and acquisitions: Montmarquette et al. (2004), for instance, examined how, after a merger, teams are redesigned to bring together managers from companies with different compensation schemes, work habits, and organizational routines. Other papers examine the role of information
on organization design and division of labor, as in Reitzig and Maciejovsky (2015), who studied how steep corporate hierarchy and apprehension about evaluation affect the propensity to share knowledge upward in the hierarchy.

The second common criticism directed toward the use of experiments in strategy is related to their lack of *external validity* (Scandura and Williams, 2000), that is, the “extent to which the results generalize across time, settings, and individuals” (Brutus et al., 2013: 54). The question of external validity seems particularly salient for this method because of “the concern that evidence from a given experimental […] research design has little predictive value beyond the context of the original experiment” (Angrist and Pischke, 2010: 23). However, concerns about external validity are not restricted to experimental methods (Brutus et al., 2013), as “empirical evidence on any given causal effect is always local, derived from a particular time, place, and research design” (Angrist and Pischke, 2010: 23). Moreover, increasing external validity should not come at the price of jeopardizing the internal validity of a study, that is, ensuring the study design “successfully uncovers causal effects for the population being studied” (Angrist and Pischke, 2008). After all, “internal validity is a prerequisite for external validity and therefore, for generalizability” (Schwenk, 1982: 215). This is not to say that experimentalists should not aim for higher external validity, but that the extent to which they do may depend on the objective of the study. Some research questions may require the researcher to purposefully “abstract away many real-world complications” (Billinger et al., 2014: 96) in order to isolate the effect under examination and avoid potential confounds. Some others may allow researchers to look for a more general result, without compromising their ability to uncover the causal effect being studied. In line with this argument, results from our analysis show that strategy experiments have been conducted using a host of different designs, from laboratory experiments involving a very specific task for a specific population of participants, to complex field studies comparing the response of different managers or organizations to the same treatment. In terms of *context (laboratory vs field)*, most of the papers in our sample employ experiments conducted in a controlled laboratory setting. This allows researchers to purposefully exclude specific confounds and make sure the observed outcomes can be ascribed to the experimental treatments only (e.g. Billinger et al., 2014). In addition, our data show an increase in the popularity of field experiments, which grew from 19% of papers in 1980–2009 to 33% of papers in 2010–2016, with an additional 7% of papers using both field and lab experiments (e.g. Berg, 2016). This is in line with trends emerging in other macro fields (Bitektine et al., 2018; Carpenter et al., 2005; Duffy, 2014; Floyd and List, 2016): Think, for instance, about the work of Banerjee and Duflo (2011) with the Poverty Action Lab at MIT. In terms of *tasks adopted*, we observe anything from simple tasks performed in a laboratory setting to more complex experimental protocols aimed at replicating real-world situations. This may involve relying on business games simulating a simplified decision-making environment in which participants play the role of managers (e.g. Conlon and Parks, 1990; Kunc and Morecroft, 2010) or presenting real-world actors with stylized scenarios that represent naturally occurring settings (e.g. Bardolet et al., 2011; Loch et al., 2013). We also observe a tendency to complement laboratory data with field evidence (Kistruck et al., 2013; Moore et al., 2007) or observational data (Raveendran et al., 2016) in an attempt to generalize findings beyond the stylized tasks presented in a controlled laboratory setting. In terms of *participants recruited*, when the aim of the study is to investigate a generic pattern in human behavior, we often see researchers resort to undergraduate students: They are readily available and sensitive to participation incentives that are commonly offered (see, for instance, Bottom et al., 2006). On the other hand, when the aim of the study is to examine specific decisions, processes, or outcomes conducted by individuals or groups with specialized knowledge, or that affect only a particular population, we observe the recruitment of managers or entrepreneurs as participants (e.g. Hodgkinson et al., 1999; Mitchell et al., 2011).
We have shown that experiments are being successfully employed to answer “macro” questions in strategy, and that study designs can be optimized for external validity when needed. But does that mean that strategy as a field should embrace experimental methods more widely? We would argue yes. To understand why, a good place to start is by looking at the advantages of experiments over other empirical methods. In general terms, experiments enhance control (Shadish et al., 2002). They improve the ability of a researcher to isolate the relationship under examination and establish causal inference (Cook and Campbell, 1979). They also enable the researcher to better match the observation to the construct it is intended to measure, thus increasing construct validity (Cronbach and Meehl, 1955). Moreover, because of their micro nature, experiments can shed light on individual-level mechanisms behind the relationships we observe and on the microfoundations of strategy research more generally (Felin et al., 2012).

Taken together, results from our analysis of strategy experimental papers help us respond to skeptics who question whether experiments can be productively used in strategy research. We have shown that, despite their micro nature, experiments can be used to study questions of strategic interest for firms. We have also shown that experimental evidence can be made more generalizable by acting upon the design of the experiment itself. Overall, these results suggest that the question we should ask ourselves is no longer whether experiments can be a useful tool for strategy research: They clearly are. What seems more interesting is understanding how to make the best use of them. And this is what our next section is about.

What Experiments Are (Surely) Good At

Let’s start with causality. Many studies in the field of strategy are based on empirical models that fail to account for the endogeneity of strategy choice (Bascle, 2008; Hamilton and Nickerson, 2003). As Shaver (1998) puts it,

firms choose strategies based on their attributes and industry conditions; therefore, strategy choice is endogenous and self-selected. Empirical models that do not account for this and regress performance measures on strategy choice variables are potentially misspecified and their conclusions incorrect. (p. 571)

Dealing with these issues has become an important endeavor, to the extent that some scholars talk about the field undergoing an “identification revolution.” As researchers expose scientific apophasia (Goldfarb and King, 2016), we are witnessing renewed interest in the use of empirical designs in which there is an exogenous treatment, as in the case of quasi- (Flammer, 2015; Gubler et al., 2016; Obloj and Zenger, 2017) and natural experiments (Lee and Puranam, 2017; Nagaraj, 2018; Starr et al., 2018; Zhang, 2018). By randomizing the assignment to such an exogenous treatment, randomized experiments enable researchers to make causal inferences that are simply not feasible in observational studies (Baldassarri and Abascal, 2017; Falk and Heckman, 2009; LaLonde, 1986). As an illustration, consider a researcher interested in understanding whether employees are willing to accept a lower wage in order to work for a socially responsible employer. One way to approach this question would be to compare salaries offered in firms that differ in terms of involvement in social responsibility. However, this would not take into consideration the fact that (a) firms choose the extent to which they want to be involved in socially responsible activities and (b) workers choose the firms they may consider working for. To overcome these selection issues, Burbano (2016) designed a field experiment in which workers were randomly provided with different levels of information about their employer’s social responsibility. By keeping the firm constant, while varying the degree to which substantially identical workers were aware of its involvement in socially responsible activities, the researcher managed to isolate the single effect of a firm’s involvement in corporate social responsibility on its employees’ demand for wages.
Experiments not only improve the ability of a researcher to make causal inferences, but they also offer a way to directly measure some constructs that may be difficult to measure in the field or for which researchers often resort to the use of proxies (Gutierrez et al., 2018). In other words, they can alleviate some issues related to construct validity, that is, the “fit between the measures employed and the constructs that they purport to represent” (Brutus et al., 2013: 54). This is important, since “the choice of an empirical measure designed to proxy for one theoretical object that others may legitimately interpret as a proxy for something else entirely” (Oxley et al., 2010: 383) has been identified as one of the common problems in our field. Indeed, in an analysis of the self-reported limitations of 1276 articles published in five management journals, Brutus et al. (2013) found that concerns about construct validity were negatively correlated with experiments. As an illustration, consider the study run by Agarwal et al. (2012) to assess the role of pre-acquisition alliances as a screening mechanism for potential acquisition targets. By designing their own experiment, the researchers were able to measure post-acquisition performance “without relying on noisy or possibly biased data like stock market returns” (p. 721), but simply based on whether “the total resource allocations by all divisions meets or exceeds the threshold amount required for synergistic profits from combined activities” (p. 723). Scholars have also resorted to experimental methods when the strategic decisions, processes, or outcomes are difficult, or impossible, to observe, which can be seen as an extreme case of low construct validity. Take, for instance, the private interactions among competing firms: In their research on the sanctioning of norm violations, Di Stefano et al. (2015) were interested in assessing how weak and strong reciprocity motivates lead industry players to respond to a norm violation with a sanction. Witnessing these behaviors is nearly impossible. The authors addressed this issue by examining intended, rather than real, action. After manipulating the likelihood that a norm violation would occur, instead of trying to observe the participants administer a sanction, they asked participants about their propensity to do so.

A final point of strength of experiments is, perhaps paradoxically, the “micro” nature for which they are so often criticized. Recent years have witnessed the emergence of many different research streams whose interest is to “decompose macro-level constructs in terms of the actions and interactions of lower level organizational members” (Foss and Pedersen, 2016: E22). Think about behavioral strategy (Powell et al., 2011), research on the microfoundations of strategy (Felin et al., 2012), as well as strategic human capital (Coff and Kryscynski, 2011). Common to all of these is an interest in a more micro approach to strategy, where macro-level phenomena are reduced to their foundations. The ideal candidates for advancing insights in these areas are “simulation exercises [and] experiments” (Foss and Pedersen, 2016: E29). As an illustration, think about recent research on the foundations of managerial decision-making, where experimentalists have been examining the cognitive (Levine et al., 2017), affective (Hodgkinson and Healey, 2011), as well as neurobiological drivers of choices (Laureiro-Martinez et al., 2014).

Overall, we believe that experiments can greatly contribute to the development of the field of strategy in many different ways. They can help us establish causal effects between variables that are precisely measured, thus bringing clarity to those domains where empirical research has been merely correlational (Geyskens et al., 2006; Tosi et al., 1997). They can also help us identify the micro-level mechanisms behind observed effects, which is crucial for theory building and for contrasting competing theories that offer alternative explanations (Agarwal et al., 2012).

Concluding Remarks
Strategy is an exciting field in which to be an experimentalist today. The identification revolution has generated renewed interest in the great advantages experiments provide when it comes to establishing causality and increasing construct validity. Experiments also seem an obvious methodological choice
when it comes to hot topics such as behavioral strategy, strategic human capital, or the microfounda-
tions of strategy. While their use is still small in relative terms, our analysis shows that the contribu-
tions from experimental methods are significant, distinctive, and growing in number.

As the method grows in popularity, common practices used by experimentalists start to undergo
a more careful scrutiny. Think, for instance, about the debate on the use of subject pools drawn
entirely from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies. On the
one hand, there is evidence showing that WEIRD subjects can be particularly unusual compared
with the rest of the species (Henrich et al., 2010). On the other hand, in light of the observation that
“empirical evidence on any given causal effect is always local” (Angrist and Pischke, 2010: 23),
one could argue that there is always something to learn from observing how (actual or future)
managers or employees behave in an organizational context. Moreover, the increasing internation-
alization of the field, together with the emphasis recently put on the value of replication studies
(Ethiraj et al., 2016), surely helps counteract this tendency of all (and not just experimental) empir-
ical research. Overall, we are observing a tendency to constantly push forward the standards for
what constitutes a “good” experiment, with new practices being introduced such as the pre-regis-
tration of hypotheses (Chambers and Munafo, 2013), open access to data (see Levine et al., 2017,
for an exemplary implementation), and avoiding deception in treatments (Gutierrez et al., 2018).
Transparency about data collection and analysis is of pivotal importance to make sure the method
is implemented at its best. And so is providing adequate training for scholars and students to con-
duct experiments and, equally important, review experimental findings.

Being an experimentalist in strategy today is also exciting because, as a community, we are giving
ourselves the liberty to experiment with the method (no pun intended) and push our creativity further
in a variety of ways. Some of us are on a quest to better use the laboratory to conduct research at the
organizational and inter-organizational level. Think about Weber and Camerer (2003), who opted for
studying the role of culture in acquisition success in the laboratory by allowing participants in ficti-
tious firms to develop a culture and then merge firms. As explained by the authors,
while existing studies provide support for the hypothesis that cultural integration plays an important role
in mergers, the causal effects are not clearly determined. This is not because of a flaw in the way the
research was conducted, but due to problems that frequently arise when working with real-world data. […]
While experiments often give up realism and generalizability, the added control and the ability to precisely
measure variables of interest mean that they can be a useful complement to field studies. (pp. 403–404)

At the opposite end of the spectrum, some scholars have embarked on the creation of ambitious
field studies. Two notable examples are the Harvard Crowd Innovation/NASA Tournament Lab
(Boudreau and Lakhan, 2015) and the MIT Bitcoin Study (Catalini and Tucker, 2017). These are
extreme cases that require important investments and may not be easy to replicate, but they vividly
portray the increasing interest in conducting strategy experiments “in the wild.”

In this article, we have tried to shed light on the value of the experimental method, together with
its limitations and the design challenges it poses. Far from saying that experiments are the perfect
method and suitable for all seasons, our aim has been to highlight the many ways that experiments
are contributing to successful research in strategy. In so doing, we hope to finally allow ourselves
to embrace the beauty of this method and use it to generate new, exciting research.

Acknowledgements
The authors would like to thank Andy King, Guilhem Bascle, Joao Albino-Pimentel, Olivier Chatain, Pier
Vittorio Mannucci, Tomasz Obloj, as well as participants in the 2017 SnO Research Day, for their insightful
comments. They are indebted to Rahul Anand for excellent research assistance, and to Erin Lee Martin for
being, once again, an amazing copy editor. A heartfelt thank also goes to co-editors Ann Langley, Glen
Dowell, and Robert David, as well as former co-editor Gianmario Verona, for supporting our essay and providing guidance in its development.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References


Chambers C and Munafò M (2013) Trust in science would be improved by study pre-registration. The Guardian. Available at: https://www.theguardian.com/science/blog/2013/jun/05/trust-in-science-study-pre-registration


Author biographies

Giada Di Stefano is an Associate Professor in the Department of Strategy at HEC Paris. Her general research interests are in the area of innovation, knowledge, and organizational learning. More specifically, she studies...
the creation and transfer of knowledge, with particular emphasis on the underlying behavioral mechanisms. Empirically, she has been adopting different methods, with a preference for field experiments conducted in real organizational settings. Her work has been published in the *Academy of Management Journal, Academy of Management Perspectives, Advances in Strategic Management, Industrial and Corporate Change, Research Policy*, and *Strategic Management Journal*. She currently serves in the editorial board of *Strategic Organization* and the *Strategic Management Journal*. She has also been elected as a Representative-at-Large for the SMS Behavioral Strategy Interest Group, a position she previously held for the AoM STR and TIM Divisions. Giada received her PhD from Bocconi University. Address: Department of Strategy and Business Policy, HEC Paris, 1 rue de la Libération, Jouy-en-Josas 78351, France. [email: di-stefano@hec.fr]

**Cédric Gutierrez** is an Assistant Professor in the Department of Management and Technology at Bocconi University. His research interests are situated at the intersection of behavioral decision-making, entrepreneurship and strategy. In particular, he examines how behavioral mechanisms (e.g. ambiguity aversion, perception of time) and cognitive biases (e.g. overconfidence) influence entrepreneurs and managers’ willingness to take risk. Empirically, he uses methodologies and models from experimental economics and decision sciences. His experimental work has been recently published in the *Journal of Risk and Uncertainty*. Cédric received his PhD from HEC Paris. Address: Department of Management and Technology, Bocconi University, Via Roentgen 1, 20136 Milan, Italy. [email: cedric.gutierrez@unibocconi.it]