3

BEYOND THE LAB

Using Data from the Field to Increase Research Validity

Alicea Lieberman, Andrea C. Morales, and On Amir

When conducting or evaluating research, it is critical to understand the landscape of research methods and the value different approaches provide. Researchers have many methods to choose from, each with their own strengths, limitations, and variance in the degree of external and internal validity they provide. Which form of validity researchers tend to care about most changes with the goals of the research, but also over time, with certain decades favoring one aspect over the other. Currently, there is a push within consumer research to enhance external validity and include more field experiments in research papers (e.g., Gneezy, 2017). However, history tells us that this may change; the process is cyclical, and what is in today may be gone tomorrow, only to reappear the day after. For this reason, it is critical not to simply follow trends, but rather to consider your research goals and how to provide the strongest evidence to support your research question.

In this chapter, we will focus on methods to enhance the realism of research (i.e., external validity and generalizability)—namely, through field experiments, experiments in-the-field, and field data. However, before diving into the details of field experimentation, we believe it is imperative to consider the genesis and goals of gathering data from the field. As such, we will first discuss the underlying motivations to collect data from the field (i.e., external validity), then describe the different empirical methods that address these goals (along with their associated strengths and weaknesses), and conclude with practical suggestions for how to most effectively use these different research methods.

How Can Researchers Provide the Most Robust Research Findings?

A common goal of consumer research is to determine the impact of a given intervention, oftentimes to test a theory. To examine the effect of a treatment or intervention, researchers must compare those who received the treatment with those who did not. Such experimental research is often divided into randomized experiments (participants are randomly assigned to intervention treatments) and nonrandomized experiments (participants experience an intervention or exogenous event, but there is no random assignment). Randomized experiments are the gold standard and provide the strongest causal evidence; however, randomization is not always possible. Nonrandomized experiments, such as quasi-experiments and natural experiments, are also an excellent approach to answering research questions, but, because they lack randomization, they may lead to less valid conclusions than randomized experiments. Researchers may also assess the relationship between two variables using existing field data. In such designs, researchers will analyze existing data that involve no manipulation or variation from a given event or intervention. Field data can offer strong supplementary support and
provide insight into the relationship between two variables, but the conclusions that can be drawn from such data are limited to correlational evidence and cannot be used to make any causal claims. See Table 3.1 for an overview of these different approaches.

When trying to assess the merit of research, a lot hinges on the strength of the empirics, even though the Lewinian idiom of “there is nothing more practical than a good theory” is as true as ever (Lewin, 1943; but see also Lynch, 1999). Historically, researchers have grouped methods into buckets based on the degree of internal and external validity they provide. However, rather than separating methods this way, we view them along a continuum and suggest that nearly all methods can be designed to enhance the external validity and generalizability they offer (with some approaches being naturally stronger than others on these dimensions). Most importantly, we contend that the impact of the research critically depends on whether or not the results are perceived to be meaningful and real, which we suggest is a function of the following criteria:

- Whether the observed results actually are what the authors claim they are and accurately reflect the proposed causal chain (usually referred to as internal validity).
- Whether the observed results replicate and are not a fluke (usually referred to as statistically significant, robust, and/or replicable).
- Whether the observed results generalize:
  - Outside the artificial lab environment (often referred to as external validity).
  - Across situations and contexts (sometimes referred to as generalizability or conceptual replicability).

<table>
<thead>
<tr>
<th>Experimental Design</th>
<th>Examples</th>
<th>Strengths</th>
<th>Limitations</th>
</tr>
</thead>
</table>
| Randomized Experiment | Field experiments | - Gold standard  
- Direct causal inference  
- Strong external validity  
- More feasible and accessible than field experiments  
- Provide appropriate context | - Logistically challenging  
- Not always possible  
- Less experimental control  
- Less external validity than field experiments  
- Less control over parameters than lab  
- Can be logistically challenging and expensive  
- Less external validity |
| Experiments in-the-field | | | |
| Laboratory experiments | Quasi-experiments | - Strong internal validity  
- More feasible and accessible than randomized experiments  
- Moderate external validity  
- Can often use existing data  
- Simpler than randomized experiments | - Less causal inference than randomized experiments  
- Need to account for alternative explanations |
| Nonrandomized Experiment | Natural experiments | | |
| Field Data | Secondary and archival data; observational data | - Can use existing data  
- Strong external validity | - Correlational: does not provide causal evidence |

Table 3.1 Experimental Designs: Strengths and Limitations
Although we discuss methods individually in this chapter, it is important to recognize that the most robust and valid evidence (internal and external) is often gathered through a mixed-methods approach, which can provide convergent results, replicate effects, and counter the limitations inherent in any single method. Thus, the strongest evidence will be gathered using a mix of different methods (e.g., randomized experiments, quasi-experiments, and correlational data) across different locations (e.g., laboratory and field; for a more thorough review of different methodological approaches, we recommend McGrath, Martin, & Kulka’s *Judgment Calls in Research*, 1982).

For the purposes of this chapter, we focus on how to select the best methods to increase external validity and generalizability. To start, it is important to be aware of the many factors that can reduce both external validity and generalizability. Some common threats stem from the participant sample (e.g., selection bias, homogeneity, fit with the research question of interest), testing effects (Hawthorne effect/demand characteristics/interaction with pretest that primes participants and influences their responses), and multiple treatments (if participants receive several treatments it can be hard to separate the effects of prior treatments). Although there are numerous ways to increase external validity, field experiments are one approach to improve external validity and generalizability. Particularly in the domain of consumer research, demonstration that proposed effects not only occur in a lab setting but are also realized in the marketplace is quite compelling. Indeed, when executed properly, field experiments can provide ecologically valid, robust findings, which is the primary reason they are often sought after and/or recommended.

Field experiments certainly have the potential to add value to research papers, but it is important for consumer scholars to recognize that they are not the only way to ensure that results are perceived as both meaningful and real and do not even necessarily guarantee external validity. Field experiments are definitely one option that can contribute to meeting the goal of enhancing meaningfulness and realism, but there are other experimental methods that are equally, and possibly at times even more, equipped to do so, depending on the circumstances. In this chapter, we attempt to position field experiments in the continuum of available methodologies and, therefore, to emphasize when their use may be advisable and/or necessary.

Viewed from a historical perspective, it is intriguing to note that many of the best-known and most paradigm-shifting behavioral experiments are not in fact field experiments. For example, experiments for Milgram’s seminal work on obedience in social psychology originally took place inside a laboratory on the Yale University campus, and participants were explicitly recruited to take part in an experiment on “learning” (Milgram, 1974). In addition, the Little Albert experiment, which provided empirical evidence for classical conditioning as well as critical insights about associative networks (Watson & Rayner, 1920), also took place in a laboratory setting and utilized an N of 1. Although the procedural details, ethics, and even research design of these historical studies have been highly disputed and likely would not gain IRB approval with the same experimental designs today, we reference them here because they are all non-field experiments whose results have provided important insights about human behavior. Further, they have had a lasting impact and are still currently being used to inform teaching, research, and policy. That is, they were powerful enough to render concerns of external validity or generalizability mute. In other words, they each delivered results that are considered meaningful and real and, consequently, are often discussed and referenced in psychology courses, academic papers, and the development of policy and interventions.

Looking specifically at the marketing literature, many well-known research studies, from both a theoretical and a practical perspective, are also not field experiments. For example, Kevin Keller’s work (1987) on advertising retrieval cues and memory utilized a single laboratory experiment in order to demonstrate that including information from an advertisement directly on a product package can increase brand evaluations. Similarly, in their research on information acceleration, Urban, Weinberg, and Hauser (1996) used a simulated showroom for an electric vehicle (EV)—equipped
with a prototype that participants could sit in—to have people go through a realistic EV customer experience and provide testimonials and ratings, which they then used to forecast EV adoption. In doing so, they were able to identify key attributes that impacted potential future adoption. Importantly, although neither of these papers utilized field experiments in their empirical work, the studies they ran in the lab and in the field were all high in experimental realism, which helped to increase the veracity and believability of the findings (Morales, Amir, & Lee, 2017).

Taken together, what these well-known papers in psychology and marketing demonstrate is that field experiments are not the only way to enhance the meaningfulness and impact of consumer research. Instead, what is needed are experiments that make it clear that the proposed/observed effects are not an artifact of the artificial environment but actually occur in real life. As we have alluded to, field studies are indeed one way to increase external validity and generalizability. However, there are multiple other options that can also increase the external validity and thus realism of research results. These methods lie along a continuum, with some natural overlap between them, and vary in the degree of external validity they provide. Throughout the chapter, we hope to emphasize that it is possible to increase external validity using almost any method and, although we group these methods by category for illustrative purposes in this chapter, we believe that a method should be chosen based foremost on the value that it provides in strengthening the research, rather than simply choosing a method because of its category. Below, we provide brief descriptions and examples of how to increase the external validity and realism using each method.

**Lab Studies with Realistic Independent and Dependent Variables**

Lab experiments are conducted in controlled settings; participants are aware that they are part of a research study and may participate in other concurrent studies before or after the focal one. Because laboratory experiments allow researchers to tightly control the manipulation and measurements, they are often employed to provide high internal validity. However, laboratory experiments can also be purposefully designed to increase external validity, namely, by using more realistic independent variables (IVs) and dependent variables (DVs).

The external validity of an experiment is greater when the research design uses IVs that are more realistic rather than hypothetical (e.g., actually interacting with another person in a sales encounter vs. imagining interacting with someone) and measures actual behavioral DVs (e.g., consequential choices, purchases with real money, amount of food eaten, gambles, effort exerted). For example, Duke and Amir (2018) examined whether temporally separating decision and consumption influences how guilty an individual feels post indulgence. During a laboratory experiment, the authors had participants make a decision to eat either a healthy food (i.e., edamame) or an unhealthy food (i.e., caramelized popcorn) and then varied the time gap between food choice and consumption. As a measure of how much guilt participants experienced, the authors measured how much of the unhealthy food participants ate in total. Thus, although this experiment took place in the lab, it utilized a realistic IV (i.e., choice of which food to eat and a delay between choice and consumption) and a realistic behavioral DV (i.e., amount of actual food consumed), thereby increasing its external validity.

Other researchers have successfully increased external validity by recording participants’ behaviors in the lab that come with real monetary consequences. Both Winterich, Mittal, and Morales (2014) and Mazar, Amir, and Ariely (2008) had laboratory participants perform a task in another room (i.e., knowledge exam, coin flip, or unsolvable anagram) incentivized by a monetary reward. Afterwards, participants were asked to report their task outcome to the researcher and were able to tell the truth or to cheat and collect a greater payout. Thus, both the behavior (i.e., cheating) and the consequences (i.e., psychological effects of cheating and monetary reward) were real, increasing the external validity of their findings.
Experimental designs utilizing realistic IVs and behavioral DVs in the lab also help ensure that researchers indeed manipulate what they intend to manipulate. Many imaginary scenarios have the possibility to be misunderstood or yield some version of demand effects. For example, having participants imagine making a choice in a hypothetical situation in which they have never been, with money they have never had, makes it less likely that the results would be reflective of behavior outside the artificial scenario constructed. In contrast, offering participants the opportunity to purchase gum using real money, for example, carries substantially less risk of being misunderstood by participants or of producing biased results that reflect consumer lay theories of how they should behave versus how they actually behave (Lieberman, Duke, & Amir, in press). Experiments in the laboratory can move farther along the external validity continuum by surreptitiously measuring real behaviors that participants are unaware are part of an experiment. For example, Williams, Amir, and Lieberman (2019) manipulated instructions for students to sign up for a fun experiment in the laboratory, and the time for which students signed up was the DV. Thus, researchers can creatively design their laboratory experiments using realistic IVs and real behaviors as DVs to enhance the external validity of an experiment conducted in the laboratory.

**Secondary and Archival Data**

The use of existing data is an excellent nonexperimental method that provides researchers with more externally valid data from the field without the need to run a field experiment. Existing data are sometimes categorized into two types: secondary data, which are data usually collected by another agency, often for their own research purposes (e.g., purchase trends, TV ratings), and archival data, which are existing data often recording events or bureaucratic procedures (e.g., accounting, surveys, weather, applications). Because the use of secondary and archival data falls under nonexperimental design, it lacks the ability to offer causal inference and is limited to providing correlational evidence between variables. However, the use of secondary and archival data can provide excellent additional data that enhance the generalizability of a theory by demonstrating its presence in the field. As a result, many highly cited papers have taken advantage of existing data sets both to develop theory and to provide convergent evidence for laboratory experiments. For example, Moore (2012) uses Amazon.com reviews to examine the relationship between linguistic elements and star ratings for fiction books. She finds that the use of more language that includes explanations for why things happened (i.e., explaining language: “I bought this ice cream because I was hungry”) is correlated with less extreme book evaluations. By using archival data in this way, Moore is able to provide correlational evidence that supports her empirical findings that the use of explaining (vs. nonexplaining) language increases storytellers’ understanding of consumption experiences, influencing their evaluations and intentions toward different hedonic versus utilitarian experiences.

In another example, Nunes, Drèze, and Han (2011) use archival data, including product photos and prices, to examine the relationship between brand prominence and price of luxury goods. Specifically, the authors use archival data from several brands, including Louis Vuitton and Gucci, to demonstrate that luxury brands more prominently display their logos on lower-priced items, whereas higher-priced items are less conspicuous. In another example, Parmentier and Fischer (2015) analyze 2003–2012 Nielsen ratings for *America’s Next Top Model* to assess the relationship between destabilization of a brand identity and the decline of serial brand popularity. Finally, Hong and Sun (2012) combine both secondary and archival data to support their claim that romantic movies are more sought after when consumers are cold. The authors use historical temperature (archival data) and data on online movie rentals (secondary data) to show a negative relationship between temperature and the proportion of romantic movies watched.

Novel analysis methodologies further expand the ways in which data can be used to explore many old and new hypotheses. For instance, with the addition of packages allowing topic analysis
and semantic interpretation, qualitative data can be used in a manner equivalent to numerical data. For example, Tully, Cheema, Amir, and Proserpio (2018) used TripAdvisor and Expedia.com hotel reviews to test the effect of eliciting expectations in the review process on the valence of the review. Using Linguistic Inquiry and Word Count (LIWC) to code the emotionality of the review, they find a main effect of mentioning expectations on the overall number of stars given, as well as process evidence supporting their proposed mechanism.

**Observational Field Data**

As suggested by its name, this nonexperimental method involves the observation and collection of field data in their natural setting with people behaving as they would otherwise. Observational data offer externally valid data in a manner that is more accessible and, oftentimes, less time- and labor-intensive than a field experiment. However, like secondary and archival data, observational field data are correlational (not causal). Thus, observational data are best used as a complement to experimental methods, and many papers have done just that to successfully support their hypotheses. For example, Hoyer (1984) observed the decision-making process and timed how long real consumers took to select a laundry detergent in a grocery store to provide real-world data supporting his conjecture that repeated and unimportant purchases do not involve much in-store decision deliberation. In a recent paper on the relationship between status and nonconformity, Bellezza, Gino, and Keinan (2013) collected observational data at a professional academic conference. The authors surreptitiously coded the dress of 76 randomly selected conference attendees and compared their level of formality (from very casual to very formal) with their status in the field (i.e., number of academic publications). The authors found that more nonconforming (e.g., casual) dress was indeed significantly correlated with higher status (i.e., number of peer-reviewed academic publications; an effect that was especially strong for publications in the top marketing journals). These correlational data serve the same research-motivating role as a thought experiment (“consider a person shopping for laundry detergent . . .”) but help ground the paper more effectively in realistic contexts, as well as make the hypotheses seem more plausible. Indeed, the combination of correlational field data and causal laboratory tests is highly recommended to increase meaningfulness and realism.

**Natural Experiments**

Natural experiments occur when a subset of individuals are “naturally” exposed to a set of circumstances that others are not, allowing for a comparison between groups. For example, in events such as natural disasters or policy changes, researchers may investigate the behavior of people who were and who were not exposed to a given situation. As such, natural experiments allow researchers to test the effects of treatments that they may otherwise not be able to randomly assign to participants. One well-known example of a natural experiment is the Oregon Health Insurance Experiment. In 2008, Oregon used a lottery to randomly select individuals from a pool of almost 90,000 low-income adults to receive the opportunity to sign up for Medicaid. This experiment served as a “naturally” occurring randomized control trial, allowing researchers to examine the causal effects of access to health insurance relative to not having access (Baicker et al., 2013; Finkelstein et al., 2012; Taubman, Allen, Wright, Baicker, & Finkelstein, 2014). Policy changes also offer opportunities for natural experiments. For example, Card and Krueger (1994) assessed the effects of a minimum wage increase by examining employment growth in fast food restaurants before and after the 1992 minimum wage increase in New Jersey (i.e., the treatment condition) compared with fast food restaurants just across the border in Pennsylvania, where there were no changes to the minimum wage (i.e., the control condition).
Natural disasters can also act as random assignment as they “randomly” affect only a portion of a population. Indeed, many researchers conducted natural experiments in the wake of Hurricane Katrina. For example, Kirk (2009, 2015) examined how both geographic dispersion and relocation of ex-prisoners influenced recidivism rates, and De Silva, McComb, Moh, Schiller, and Vargas (2010) analyzed the impact that the sudden influx of immigrants into the Houston area had on labor earnings.

A second form of “natural” experiments happens when a variable unrelated to the causal inquiry naturally varies across consumers. For example, county lines are often drawn as a result of political goals and may follow arbitrary geographic lines. As such, if people on either side of a geographic line are exposed to a different level, type, or content of advertising, one may use this separation to mimic random assignment (Shapiro, 2018). Further, a third type of natural quasi-experiment can be found in pre–post analyses, such as looking at thin slices of Internet behavior before and after a TV ad is aired. For example, Liaukonyte, Teixeira, and Wilbur (2015) find that action-oriented TV ad content increases brand website traffic and sales by examining behavioral data observed in two-minute increments immediately before and after brands’ TV ads are played on national television networks. This example serves as a good transition to an additional important class of design—quasi-experiments.

Quasi-Experiments

In quasi-experiments, researchers test a hypothesis by manipulating a variable and measuring its effect. Notably, however, quasi-experiments necessarily do not involve randomization. Rather, participants in quasi-experiments are assigned to a condition through self-selection (i.e., they choose their treatment or intervention group) or by researcher selection (i.e., the researcher chooses which participants or groups of participants to expose to which treatment or intervention). Because quasi-experimental designs lack random assignment, their conclusions may be less valid because of alternative explanations such as history (effects may be explained by something else having occurred during the measurement period); maturation (effects may be explained by normal changes that occur with time); selection differences (effects may be explained by inherent differences in the groups); testing (effects may be explained by a pretest having influenced behaviors); instrumentation and instability (effects may be explained by changes in measurement tools or forms of measurement); or regression (effects may be explained by extremes having regressed to the mean). However, with careful consideration, researchers can often use logic, experimental design, and a mixed-methods approach to address many (but never all) counterfactuals. For example, Godes and Mayzlin (2009) used a quasi-experiment to analyze which types of word-of-mouth (WOM) communication are most effective at driving sales. Specifically, they compared whether a WOM intervention was more effective when spread by a firm’s loyal customers or by people who were not customers of the firm. In this design, there may have been inherent differences between the customer and noncustomer populations that could not be accounted for because of the lack of randomization. In an attempt to address this limitation, the researchers used a mixed-methods approach and also conducted an online experiment to replicate their findings from the field and more directly test the underlying mechanism.

Quasi-experimental designs form an important (and very large!) class of methodology, each with their own set of strengths and limitations. We recommend interested readers refer to the original “bible” of quasi-experimental design by Campbell and Stanley (1963, 1966), and an updated edition by Shadish, Cook, and Campbell (2002). Although we certainly do not cover all quasi-experimental designs here, there are a few common designs worth noting. First is the one-group pretest–posttest design, in which observations are taken before and after an intervention, and the difference between these observations is attributed to the intervention. For example, Moorman (1996) conducted a
pre–post longitudinal quasi-experiment with more than 1,000 consumers in supermarkets across two states. Specifically, across 20 different product categories, Moorman measured consumer comprehension and acquisition of nutritional information 8 months before and 5 months after the implementation of the Nutrition Labeling and Education Act (NLEA). Moorman strengthened the validity of her results by testing for interactions between times (before and after NLEA implementation) and several consumer characteristics, such as individual differences in motivation and skepticism.

Another common quasi-experimental design is a nonequivalent group design. Nonequivalent group designs evaluate the effects of different treatments or interventions to which participants were not randomly assigned, and outcomes are compared across intervention group. For example, Ariely, Ockenfels, and Roth (2005) compare the behavior of customers of online auctions on Ebay and on Amazon, exploiting the different ending rules they employ to learn about strategic behavior and the psychology of consumers. Although one could argue that very different people used each site, the large amount of data, coupled with lab experiments, drove the point home quite compellingly.

**Experiments In-the-Field**

Experimentation in-the-field (Morales et al., 2017) is an experimental design in which participants, usually consumers in real consumption situations, are aware that they are taking part in some form of research. Experiments in-the-field are often implemented directly by the research team and, as a result, allow researchers to maintain more control over implementation and data collection relative to field experiments. Because of this, experiments in-the-field can be more accessible than field experiments, as they do not always require partnerships or reliance on others to manage the procedure and data collection, as field experiments often do.

For example, McFerran and Argo (2014) used an experiment in-the-field to demonstrate that the presence of others can influence a VIP’s feelings of status. To examine this effect, the authors rented a luxury suite at a professional football game and invited fans seated in the regular seating to watch one quarter of the game from the suite, where they received VIP treatment (e.g., catering, etc.). They found that fans who brought an entourage with them to the suite reported feeling higher status than those who went to the suite alone.

Another benefit of experiments in-the-field is that they can be run almost anywhere, ranging from a sports game to nearly any location where people are engaging in some form of consumption experience, even on a university campus. For example, Argo, Dahl, and Morales (2006, 2008) asked people on a university campus to visit the university bookstore and informed them that they would be assigned a task and asked to share their thoughts about the experience. Although all participants carried out the same task (i.e., trying on a t-shirt), the experience itself differed across participants (e.g., volume, proximity of other shoppers, etc.) as a result of being run in a real store during regular business hours, providing a robust test of the hypotheses. Further, although participants knew that they were taking part in research, they were not aware which elements of the experience were being manipulated or which behaviors were being monitored. In another context, Kristofferson, McFerran, Morales, and Dahl (2017) sent participants to interact with a rigged vending machine in the basement of the building and monitored the level of aggression exhibited when the product got stuck and did not come out. Again, although participants knew they were engaged in some research, they did not know they were taking part in this particular research study and were unaware that their interaction with the machine was being observed.

As demonstrated in each of these examples, by exhibiting a bit of creativity in selection of stimuli and location, researchers can gain a great deal in terms of the degree to which their results seem believable and meaningful. In addition, by changing the location and running a study in an actual consumption context, researchers are better able to definitively classify the work as consumer research (Deighton, 2007). As a result, we propose that experiments in-the-field may currently be
the most underutilized marketing-centric methodology in consumer research and have the ability to greatly increase the external validity and generalizability of consumer behavior research.

**Field Experiments**

Researchers often define field experiments based on three criteria: participants are engaged in a real consumption experience; there is a surreptitious manipulation of which participants are unaware; and, participants do not realize that they are part of a research study (Charness, Gneezy, & Kuhn, 2013). The premise underlying this definition is the notion that mere knowledge that one is being observed may affect and even change behavior. For consumer behavior research in particular, we suggest that there may be some flexibility in these criteria. Namely, we believe that an experiment can still be considered a *field experiment* even if consumers are made aware that they are part of an experiment, as long as consumers do not know that they are part of a field experiment *during* the manipulation and consumption experience. If the goal of the field experiment is to enhance the external validity of the research in order to gain insight into how consumers behave in real consumption experiences, we contend that the most critical piece is that consumers engage in a “clean” consumption experience where they are unaware that there is anything different from those in which they typically engage.

The main advantage of this more flexible definition of field experiments from a methodological perspective is that it allows researchers to choose to inform participants about the experiment at the end and ask them to complete a survey (to collect additional data). As an example of this type of experiment, Morvinski, Amir, and Muller (2017) hired sales people to sell products to individuals walking down the main drag of a university campus and, only after they made their consumption decision, did they ask them to complete a survey. Because participants were unaware that they were part of an experiment during the manipulation and when making their purchase decision, this experiment fits our definition of a field experiment. Further, it also allowed researchers to obtain additional information from participants to provide even more insight into the research question.

As we have well noted, field experiments have become a common method, and researchers are using increasingly creative ways to test hypotheses as part of a real consumption experience. For example, Wu, Samper, Morales, and Fitzsimons (2017) provide support for their hypothesis that consumers will use less of an aesthetically appealing product by testing the amount of toilet paper used when the paper either was more or less aesthetically pleasing. The authors partnered with a fitness studio and, over the course of two weeks, stocked the bathrooms with either plain white toilet paper (lower aesthetic appeal) or white toilet paper printed with festive holiday motifs (higher aesthetic appeal). In support of their hypothesis, they found that customers used significantly less of the aesthetically appealing toilet paper than the plain white toilet paper. This example underscores the truly unique contribution of field experiments: it would be extremely difficult to devise a method to test this particular hypothesis with actual consumption in the lab. Nevertheless, properly testing the given hypothesis still required finding an industry partner who would cooperate on all aspects of the study design and measures.

Although field experiments are often considered “the gold standard” in increasing external validity, the explicit categorization of a method as a field experiment or an experiment in-the-field is less essential than the value that these different methods achieve. When choosing or evaluating a research method, rather than focusing on the designation or category of the method, it is most important to focus on whether a method maximizes the external validity and realism of the experiment. In this way, both field experiments and experiments in-the-field (along with many other carefully designed methods mentioned in this chapter) can be used to enhance the meaningfulness and impact of research.
Utilizing More Appropriate Participants (e.g., the Elderly, Low SES Consumers)

A final fundamental, overarching way to increase external validity and generalizability is to use participants who represent the population that you hope to study. For example, to study how vulnerable populations respond to repeated financial incentives to undergo cancer screening, Lieberman, Gneezy et al. (2019) conducted a longitudinal field experiment in which they offered differing levels of financial incentives ($5, $10, or no incentive) to a low-income, uninsured population, annually for three years. They found that, among this population, offering financial incentives did not significantly impact screening completion in the short or long run. In another example, Gneezy, Leonard, and List (2009) sought to test whether cultural contexts influenced gender differences in competitive behaviors. To provide support for their theory, the authors conducted experiments with members of both a patriarchal society in Tanzania and a matrilineal society in India. Participants were asked to toss tennis balls into a bucket and were rewarded for each successful shot (i.e., the ball entering and staying in the bucket). Prior to beginning, participants were asked to select between two payment schemes: The first scheme was a less competitive option that offered piecemeal compensation for each successful shot regardless of the performance of other players; the second scheme was a more competitive option and offered a higher incentive for successful shots, but required participants to outperform other players. Gneezy and colleagues found that, in the patriarchal society, men were more likely than women to choose the competitive scheme, whereas, in the matrilineal society, this pattern was reversed, and women were more likely than men to select the tournament option. Studying the “standard” participant population in this case would have had no hope of answering the real research question.

Which Method Is Best?

Given the multitude of approaches that may increase external validity, you may find yourself asking which method is the very best. Well, our fundamental advice is to use the method that most effectively allows you to answer the consumer behavior question you are asking. Before even selecting a method, it is important to think about the intricacies of your research question, solidify who it is you will be studying (i.e., your population of interest), which behavior it is you are interested in measuring (i.e., your DV), what you think is most likely to drive that behavior (i.e., your IV), and in which contexts you think these behaviors will most often take place (i.e., potential moderators and necessary controls). Additionally, one should scrutinize these choices based on the intended target audience. If one hopes to convince marketers, one should select different research design elements than if one hopes to convince policymakers or even scholars in a different field who may have different evaluative standards. Once you nail down these elements, you can choose a research design that best fits your research question and criteria, as well as available resources.

As we alluded to previously, many people immediately think that the best way to increase external validity is to run a field experiment. However, we agree with both Gneezy (2017) and Lynch (1999), who claim that external validity is a reflection of theory rather than method. Instead of immediately planning to run a field experiment because you think that it will make a stronger paper, first decide on the specific research question that you want to answer and then decide on the most compelling way to answer it. Jumping the gun and choosing a method based on what you think will “look best” rather than what best fits your research question not only does not guarantee a stronger paper, but also runs the risk of weakening the paper’s overall empirical package. In fact, there is often a mismatch in papers between the proposed theory/conceptual framework of the paper and the empirical evidence provided. Remember, the best papers are those in which the empirics directly test the proposed theory and where the experimental methods/designs provide the most compelling support for the predictions! Of course, it also helps to have meaningful predictions.
A Broader Approach to Increasing impact

Taking a broader approach of first considering your research question and then asking which method will be best to test your theory will allow you to increase the impact of consumer research in the most effective and efficient manner possible. Although a large-scale field experiment may be needed in some cases, other times it might be best to collect compelling, meaningful data using an alternative method. Once you have identified your research question and chosen a method that best answers that question, it is essential to use that method properly. Below, we provide some “tips and tricks” on how to best collect data with each of the abovementioned methods.

Lab Studies with Realistic IVs and DVs

What is the difference between behavioral research in psychology and that in marketing? Although the discipline identity may be beside the point, research on consumer behavior over the years stands out as attempting to address more actionable market-related issues and phenomena. Though at times mere semantics, the choice of appropriate stimuli or the context of the research question may make all the difference in the applicability and relevance of research (Amir et al., 2005). As such, we encourage consumer researchers to use realistic stimuli and employ behavioral measures in research designs (Morales et al., 2017). This can sometimes be as simple as using products or advertisements for experimental manipulations or giving participants a choice of real products as a DV. For example, in order to investigate the impact of scarcity ads on aggressive behavior, Kristofferson et al. (2017) gave participants a scarcity ad to evaluate (similar to those that consumers encounter around Black Friday) and then measured their level of aggression in various behavioral tasks. Notably, the researchers did not have participants think about a time where they experienced scarcity and write about it or search for scarcity-related words in a word search—both of which are manipulations that are commonly used but are lower on realism and may miss key physiological mechanisms. Instead, because they were interested specifically in the impact that scarcity ads had on aggressive behavior, they had participants actually look at a scarcity ad and then engage in tasks where their degree of aggression could be measured (e.g., playing a shooting or boxing video game, interacting with a vending machine). By enhancing the realism of both the IVs and DVs, the researchers were better equipped to answer the specific research question they were asking.

The enhancement of realism does not have to be physical in nature, however. The current state of technology allows more opportunities than ever to assimilate realistic stimuli into laboratory experiments, even if done electronically or virtually. Shopping environments and consumption contexts can now be simulated online or in virtual reality. Even social interactions between consumers and/or employees in the marketplace can be replicated in the lab. At a minimum, pictures, sound, and video can be added to any electronic survey to increase the level of realism, even in a computer-based survey. All this is to say that not including realistic stimuli or behavioral measures is most often the result of a lack of awareness or sheer laziness. Moreover, it often results in less meaningful and impactful research.

Secondary and Archival Data

Although not all existing data sets are publicly available, there are many that are free of charge and easily accessible. For example, state and government agencies, such as the Centers for Disease Control (CDC), the National Center for Health Statistics (NCHS), and the Census Bureau, provide free downloadable data sets. As noted in our earlier examples, there are also opportunities to access pricing and sales data, historical weather data, or even outcomes of election results. We recommend first investigating which data sets are publicly available, as you might be surprised how much data
you can get your hands on. If there are no secondary or archival data sets available to the public that answer your question, you may consider approaching a company and simply asking them if they would be willing to share their data—which is especially likely if your questions of interest can be addressed without individuating or other sensitive parameters. Another approach is to consider who in your professional and personal circles may have access to data sets that could be of interest to your research. Oftentimes, simply talking to the right person in a company will greatly increase your chances of accessing data. Sometimes you may even be working at a company that has data that can answer your question (e.g., a university; for example, Netzer, Lattin, & Srinivasan, 2008).

Finally, it is perhaps pertinent to note that the majority of consumer behavior research would most likely require individual-level data to properly address research questions. This kind of data is less freely available than the aggregate sets (though some of these have been used to address a few research questions successfully, as noted above), as disaggregate data may pose privacy and security concerns. It is advisable to discuss with your local IRB the kinds of care and restriction required when dealing with sensitive data of these types.

**Observational Field Data**

As there is essentially no limit on the variety of observation opportunities and techniques, our advice here is to simply be creative! From the recording of digital clickstreams to good old-fashioned human recording of instances of behavior with a clipboard, or from using location heat maps of customer whereabouts to drone surveillance, there is no real limit on observational data (except ethical and legal ones, of course). As noted earlier, even academic conferences can serve as fruitful opportunities for data collection from the field (Bellezza et al., 2013). If there is no natural experiment or variation, and no manipulation, the analysis may be limited to correlational conclusions, but, as discussed above, correlational results can also have far-reaching implications when interpreted and used correctly. And, in some situations, it can be very hard to argue reverse causality (e.g., increased consumption correlating with ambient air temperature; Hong & Sun, 2012). As indicated by the examples outlined above, there is a great deal of research that has relied on observational data to provide meaningful insights with substantial impact.

Observational data from the field can also serve as compelling motivation and support (in the front end of a paper or elsewhere) to underscore why a research question is worth examining. For example, before running their lab studies, Peck and Childers (2003a, 2003b) started by simply observing the number of people who picked up and touched produce in a grocery store before purchasing it. Although these observational data did not make it through the review process into the formal publication, it still served a critical role by demonstrating to both the researchers and the reviewers that physical touch is an important part of the buying process and is worthy of formal examination. Similarly, to motivate their investigation of busyness as a status signal, Bellezza, Paharia, and Keinan (2016) examined 1,100 tweets from celebrities and famous people and coded them to show that 12 percent made references to being busy, working hard, and having a lack of free time. Using observational data from the field to demonstrate that the topic of interest exists in the marketplace is an effective way to immediately bolster the perceived meaningfulness and potential impact of a research paper.

**Natural Experiments**

By virtue of their name, natural experiments are not created or controlled by the researcher. However, they can be extremely powerful for precisely this reason. To increase the likelihood your research may benefit from natural experimentation, here are a few helpful pointers:
With a well-defined theoretical prediction, one is much better equipped to search for and identify relevant natural experiments. In other words, do your homework first.

Keep tabs on policy changes and/or policy differences between potentially matched areas that may provide good opportunities (e.g., differing policies on plastic bag use while grocery shopping; Homonoff, 2015).

Keep an open mind and be cognizant that many natural experiments present the opportunity to use the same exogenous event to answer a multitude of research questions (as demonstrated by the wide range of research that occurred after Hurricane Katrina).

Quasi-Experiments

As mentioned earlier in the chapter, quasi-experiments cover a vast array of different designs and are often much more feasible than randomized field experiments, making them a great method with which to be familiar and a great way to increase the generalizability and realism of research. However, when considering quasi-experiments, it is imperative to remember that, because they do not involve randomization, they are subject to reduced validity. It is important to be familiar with the alternative explanations that quasi-experiments cannot rule out on their own and account for them, as best you can, using other methods. Given the popularity of quasi-experiments, there are many excellent resources that provide in-depth guidance on what can (and cannot) be concluded from quasi-experiments and how to design experiments (and triangulate methods) in order to minimize alternative explanations.

Experiments in-the-Field

Somewhere between the lab and the field is the method of running controlled experiments in and around the natural consumption setting. From interviewing customers in the mall about their shopping trip (Vohs et al., 2008) to having customers at car dealerships view ads and answer questions about their perceptions and attitudes toward a car (Brough, Wilkie, Ma, Isaac, & Gal, 2016), participants know they are taking part in a study, but are representative of the ecologically valid participant pool. Importantly, one of the defining characteristics of experiments in-the-field is that the DVs are recorded or measured in the relevant context of investigation. Thus, because realism is enhanced in the IVs, the meaningfulness and validity of the DVs is increased as well (Morales et al., 2017).

In trying to decide whether an experiment in-the-field is the appropriate methodology for investigating your research topic, it is important to consider the answer to two questions: (1) Is realism in the IVs helpful in being able to study and document the specific behavior of interest? And (2) is it easier to create/bring the realism into the lab or to bring the participants into the real consumption context of interest? Although we contend that realism is typically helpful in studying all consumer research topics, there are clearly cases where it is more helpful than others. In particular, when consumers are unaware, have an erroneous theory, or are unwilling to admit that a factor may be influencing their behavior, hypothetical scenario-based studies are particularly uninformative. For example, although people generally respond quite negatively to clothing they believe has been touched by other people (Argo et al., 2006), they may be unwilling to acknowledge the impact that this factor may have on their self-reported product evaluations of imaginary clothing. However, when they are forced to try on real clothing (that they believe has been touched by others) in an actual dressing room inside a real store, it is easier to assess their real reactions. Thus, although it may be the case that consumer lay theories or predictions about their behavior sometimes do match their actual behavior, when the research question is focused on better understanding consumer behavior, it can only be helpful to enhance realism in the IVs, and one way to do so is by running an experiment in-the-field.
The answer to the second question is harder to determine and will likely differ across research projects as well as individual studies. In some cases, creating realism inside the laboratory is not only feasible, but quite simple. For example, in Castro, Morales, and Nowlis (2013), the researchers were interested in how consumers responded to various cues in product displays of consumer packaged goods. As such, they were able to purchase industrial shelving (from a grocery store supplier) and stock it with products to resemble an actual grocery store display without much difficulty. However, in contrast, in Wang, Krishna and McFerran (2017), the researchers were interested in how consumers responded to various eco-friendly cues inside a hotel room. In this case, rather than constructing a hotel room from scratch inside the laboratory, the researchers instead had participants go to an existing hotel to complete the study in the context of interest. Depending on the specific research question, taking participants into the field may not only be an opportune way to enhance realism, but it may in fact be simpler than trying to bring the same degree of realism into the lab.

The options for experiments in-the-field are bound only by researcher creativity, and the risks tend to be much lower than a pure field experiment. Some shortcuts that have proven effective involve taking advantage of local areas/campus environments and facilities (e.g., restaurants, coffee shops, bookstores, movie theatres). University-affiliated businesses tend to be more amenable to collaborating with university researchers. From a logistical standpoint, it also can streamline the recruitment of participants if the field location is close to campus, and ideally even close enough to send participants directly from the lab. Indeed, some schools have sent participants from their lab subject pools (as part of the lab session) to participate in an experiment in-the-field, and then brought them back to the lab to complete more studies.

One aspect that is particularly appealing about experiments in-the-field is the use of confederates. Research assistants, posing as store employees or other shoppers, can be extremely useful for increasing the level of realism in this type of study. For researchers studying social influence, in particular, being able to construct and observe social interactions between individuals in a real-world setting is highly beneficial. The downside is that, with so many real shoppers and real employees in the field setting, it can sometimes be difficult to identify research participants. Some researchers have found it helpful to have participants wear a lanyard around their neck, so that research assistants/confederates can easily identify the next participant. Use of two-way radios and/or text messages to provide members of the research team with relevant information in real time is also extremely useful.

Finally, for studies where subject pool participants cannot be utilized, we also have some recommendations to help recruit participants for experiments in-the-field. First, it is important to minimize the hassle for potential participants, as people hate transaction costs, and any such barriers can prevent participation. Setting up a table outside the experiment location so that all participants have to do is walk inside to participate is highly recommended. Our experience is that people respond well to small token incentives when approached for participation, and gift cards, candy bars, and other types of food are often more appealing than the equivalent amount of cash. Another strategy that some researchers have found particularly effective for recruiting participants is to use a “bingo cage” to make the choice of incentive seem more like a prize they win, and therefore more appealing. Using the exact same set of incentives but allocated by spinning the bingo cage increased interest substantially for studies run by researchers at Arizona State University, for example.

Field Experiments

When it is the case that a field experiment is, in fact, the desired methodology to most effectively meet your goals, there are several key considerations that may help. Importantly, most companies do not actively wander around searching for academics with whom to partner and conduct experiments. Therefore, it is often very helpful to utilize existing relationships to reach the right person in the company to discuss the possibility of running experiments.
If you have to start a new collaboration from scratch, try to establish a relationship with the company by building trust. Running several field experiments with them where they are able to see the benefits over time has worked well for some researchers. Building on this latter point, when approaching a new company, try to do so with a research idea that will be beneficial to their bottom line as well. When incentives are aligned between the researcher and the business partner, it is far easier to get momentum for running a field experiment. Although not always feasible, smaller companies and privately owned businesses are typically easier to work with for field experiments, as they do not have the red tape often associated with larger organizations. If you can locate and work with the key decision-maker in the organization who has the ability to make go/no go decisions independently, that will streamline the process substantially.

When trying to identify possible collaborators, another recommendation is to consider finding business partners that do not require other people to execute the experiment, but instead allow the researcher greater involvement. Often the probability of getting the green light to run a field experiment can be increased substantially when the burden on the business partner is minimized or even eliminated completely. Case in point, in the field experiment examining toilet paper consumption in Wu et al. (2017) mentioned previously, the researchers provided all of the supplies to the fitness studio and did all of the measurement—the fitness studio employees only had to restock the toilet paper when needed, as per usual. As such, the only real hurdle to overcome was getting permission to change the type of toilet paper that the fitness studio used for the duration of the experiment; there was no additional burden placed on the employees. Generally speaking, the more responsibility that the researchers take on in running a field experiment, the greater the likelihood that business partners will be willing to collaborate.

Alternatively, and carrying much smaller risk, another option is to simply run a field experiment independently, on your own—on campus, in local community areas, or other places that provide the appropriate context for the behavior of interest (however, be sure to get permission if you are doing this in any location that is not considered a public space). As discussed above with respect to experiments in-the-field, university campuses, in particular, are typically quite amenable to allowing researchers (associated with their universities) to run field experiments. Kristofferson, White, and Peloza (2014), for example, ran a field experiment in the main student union building at the University of British Columbia. Specifically, research assistants intercepted students inside the student union, giving them free poppy pins for Remembrance Day. Whereas some students received the pin inside an envelope, others were just given the pin directly in the open. Unbeknownst to the students, another research assistant down the hall subsequently asked them for a donation on behalf of Canada’s War Veterans and recorded whether they made a donation and how much they donated. Because students were unaware that they were part of a research study, unaware that some students received the pin in an envelope while others did not, and unaware that their donation decisions were being observed, all of the criteria for a field experiment were met, even though the study was executed independently by members of the research team. Similarly, White, Simpson, and Argo (2014) ran a field experiment on their own outside a coffee shop in the business school, where they had a confederate give students a free cup of coffee. The confederate told all of the students that the cup was compostable but told students that either business (in-group) or computer science (out-group) students are the most effective in composting efforts on campus. Another research assistant surreptitiously recorded whether the students put their cup in the nearby compostable bin. Again, the researchers were able to collect compelling experimental data from the field, without any of the hassles associated with collaborating with an outside business partner.

In planning your field experiment, be sure to make your research design as simple as possible. There will inevitably be many things in the field that do not go according to plan, and having the simplest design possible will increase the chances that your manipulation(s) are executed properly, resulting in cleaner data. Simpler designs also make it easier for the company and the employees that
are involved in the experiment to all be on the same page and digest the moving parts. To the degree possible, try to design the experiment to require participation from the smallest possible number of employees; the more people involved with implementing the manipulation(s) or measuring the DVs, the greater the likelihood of inconsistency. When a team of employees is required to execute the experiment, be sure to provide them with simple, clear instructions that outline every step of the experiment and data collection process. In addition, training the entire group to follow the exact same protocol is essential. Remember, the typical employee has never been involved in running an experiment and will need thorough training. Take nothing for granted, and, when in doubt, it is better to overexplain than leave important details out of the experimental protocol.

Of note is the general risk of conducting field experiments with external third-party partners (e.g., companies). There is often a mismatch between what companies usually want (what works—their goals) and what consumer behavior researchers usually want (why does it work, under which specific conditions and contexts—your goals). This misalignment often leads to the abrupt cancellation of a project, lack of ability or willingness to share the critical data, or even legal blocks on publishing results. To try and mitigate some of the risk, it is recommended to (a) make sure the method(s) that you choose address both your goals and the goals of the partner, and (b) obtain written commitment and, when applicable, legal clearance prior to investing in the experiment. Also, be sure to make all expectations transparent and agreed upon before starting the experiment. There have been many unfortunate instances where people have run a field experiment and then encountered serious data licensing/publication issues after everything has already been completed. It goes without saying that data that cannot be published are obviously not at all helpful in increasing the meaningfulness and validity of research.

One last point to emphasize: It is possible to have participants in a field experiment complete a short survey or answer questions in order to provide additional dependent measures after they have engaged in a consumption experience. As mentioned previously, although there is disagreement among researchers on what constitutes a field experiment (e.g., Gneezy, 2017), we agree with Morales et al. (2017) who proposed that the most critical factor determining whether a consumer research study can be considered a field experiment is that participants do not know they are part of a research study at the time the manipulation is occurring or while they are engaging in real consumption behavior. As such, participants may be asked to answer questions without invalidating the field experiment classification because the consumption experience itself is still “clean,” in that participants are unaware that anything about the consumption experience is different from those in which they typically engage. In Study 1 of Russell, Russell, Morales, and Lehu (2017), for example, moviegoers went to a theater to watch a movie, completely unaware that an experiment was taking place. Once they got inside the theater, depending on their condition, they watched only the movie; the standard industry movie previews and then the movie; or both commercials and the standard industry movie previews and then the movie. Once they left the theater, moviegoers were asked to answer a brief survey about how much they enjoyed different aspects of the consumption experience. Again, because they did not know they were part of a study while they were watching the movie or that other moviegoers were watching previews and/or commercials as well, this study meets the criteria of a field experiment, even though a research survey was administered. Notably, by allowing participants to answer explicit questions as part of a field experiment, though not a behavioral DV, some of the limitations inherent in field studies (i.e., limited process evidence, only one DV) can be overcome, and we recommend—when feasible—doing so for precisely this reason.

**Utilizing More Appropriate Participants**

Utilizing the most appropriate populations should be incorporated into any research design. Thus, one of the first questions that you will need to ask yourself is whether your question of interest is
specific to a certain population or is generalizable to a broader group. Imagine, for example, that you are testing a hypothesis about older adults and purchase decisions. In order to claim that any findings are relevant to older adults in general, the experiment must be done using that specific population. There are many approaches that will allow you to gain access to your population of interest. For example, if you are running a laboratory experiment and interested only in a specific race or gender, you can design your experiment such that it starts with a screener and will only include people in your study who match the population you are studying. Another option is to recruit participants directly. For example, if you wanted to study college-age students who suffer from social anxiety, you could place flyers around campus specifically recruiting your population of interest. Yet another option is to utilize online panels (e.g., Turk Prime, Qualtrics panels, Prolific) that allow you to specifically target participants on a range of variables, such as demographic variables (e.g., age, income, race) or even profession (though we recommend including verification measures inside your study despite panel company assurances of screeners). Alumni networks are another way to gain access to a large number of individuals in a given profession.

In some cases, it may make sense to invest in more labor- and capital-intensive ways to secure more appropriate participants. The Brazilian School of Public and Business Administration at Fundação Getulio Vargas (FGV) in Rio de Janeiro, Brazil, for example, built a second behavioral lab, in addition to its lab in the main building at FGV, in order to conduct research specifically on low-income consumers. Recognizing that most university labs are unable to include low-income consumers as participants in research studies, FGV decided to expand its participant pool by building the Behavioral Lab–Maré, which is located in a low-income urban area in the North Zone of Rio de Janeiro. By collecting data in this lab, researchers at FGV have been able to examine research questions focused on social class and the social identity of low-income consumers effectively (Jacob, Vieites, Goldzsmidt, & Andrade, 2018; Vieites, Goldzsmidt, & Andrade, 2018; see also, Carnegie Mellon’s Data Truck, Carnegie Mellon University, n.d.).

**Conclusion**

There has been a push in recent years to increase field experiments in consumer research in an effort to increase external validity. Field experiments are one great approach to enhancing impact but are often difficult and time consuming. However, they are not the only way to increase the realism and impact of research. In fact, many of the canonical experiments in consumer research do not involve field experiments at all. Rather than automatically assuming that a field experiment is the only way to enhance the validity of your results, it is important to first identify what question you are asking and then consider what method would be best to address that question. In other words, carefully consider which methods make the most sense to answer your research question and then, using those method(s), determine how you can most effectively enhance the realism and external validity of the research. Here we offer a range of methods and practical advice on how to efficiently and effectively enhance external validity. Experiments in-the-field are one method, in particular, that are often overlooked but can be a great approach to collecting meaningful data that may result in less hassle and lower risk than running a large-scale field experiment. Running field studies on your own, as an independent research team, is another very useful approach. Thus, before investing all of the resources required to collaborate with a business partner to run a large-scale field experiment, we encourage the reader to first and foremost consider what it is that makes results meaningful and impactful. Specifically, think deeply about the specific research question you are asking and the type of data you need to most effectively answer that question, while maximizing internal validity, replicability, external validity, and generalizability. In some cases, a field experiment may in fact provide the best data for a particular research question, but there are also cases where secondary and archival data, observational data, quasi-experiments, natural experiments, experiments in-the-field, or even
lab experiments with realistic IVs and DVs make for the most meaningful and impactful empirical package. We hope this chapter has helped to identify more precisely how and when these different methods of collecting data from the field can help to maximize the validity of consumer research.

Notes
1 Natural experiments and quasi-experiments have many similar elements, and some researchers consider natural experiments to be a type of quasi-experiment, sometimes even referring to them as “naturally occurring quasi-experiments.” The key difference between them is that the variation in a natural experiment happens “naturally” (e.g., a natural weather event), whereas the variation in a quasi-experiment is manipulated, usually by the researcher for research purposes (e.g., a medical intervention). Notably, there are some scenarios (e.g., a policy implementation) that could arguably be classified as either.
2 Some researchers correctly point out that, in today’s consumption environment, consumers are often aware they are constantly being monitored (e.g., online purchase patterns). The exact point at which knowledge of being monitored affects one’s behavior is an empirical question, yet to be evaluated.

References
Beyond the Lab


