Journal of Applied Research in Memory and Cognition 7 (2018) 11–15

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Commentary

Connecting Laboratory and Field Research in Judgment and Decision Making: Causality and the Breadth of External Validity

Daniel M. Bartels*, Reid Hastie, Oleg Urminsky

University of Chicago Booth School of Business, United States

Markman (2018) provides a thought-provoking perspective on the relationships between lab and field research, using his own research program on regulatory fit as an instructive example (e.g., Otto, Markman, Gureckis, & Love, 2010; Worthy, Maddox, & Markman, 2007). He discusses characteristics of typical lab and field studies, identifies some examples of productive relationships, and analyzes the failure to communicate between lab and field researchers, focusing on the relationship between academic cognitive psychology research programs and naturalistic decision making programs (e.g., Lipshitz, Klein, Orasanu, & Salas, 2001; abbreviated as NDM hereafter). He also provides useful advice about how to facilitate interactions between researchers of each type.

We elaborate on Markman's advice on how to promote lab–field interactions in research by exploring distinctions in the goals and benefits of different kinds of field research and noting some additional ways of effectively relating lab and field settings. We add two important distinctions to Markman's conceptual analysis: narrow versus broad external validity and descriptive-observational versus causal-experimental field studies.

Narrow Versus Broad External Validity

Markman refers to a tradeoff between internal and external validity (Campbell & Stanley, 1963; Cook & Campbell, 1979). As we interpret it, internal validity refers to the validity of causal claims specifically in the setting where they are discovered. Markman indicates that external validity refers to the usefulness of a description or theoretical construct to elucidate and provide control over behavioral phenomena in a specific naturally occurring situation. NDM studies of firefighters, military personnel, or airplane crews *in situ* exhibit high external validity because their conclusions are valid in one specific naturally occurring, non-laboratory setting.

Markman "types" NDM studies as high in external validity and low in internal validity and implies that much of laboratory-based judgment and decision-making research is low in external validity but high in internal validity. We think taking a broader view of research that engages with the field may lead to different conclusions about the nature and inevitability of tradeoffs between lab and field.

There is an important distinction between narrow and broad external validity: the generalizability of a conclusion or finding only to one specific naturally-occurring target situation versus generalizability to many situations beyond the one where the finding was originally discovered. We would conjecture that much NDM research has aimed to answer questions about decisions in one specific non-laboratory setting like firefighting incidents, commercial aircraft cockpits, the bridge of a Navy cruiser, or the radar room on a Navy destroyer. The motivation and funding for much of this research was aimed to improve individual and team decision processes in these specific settings.

This type of narrow external validity is an appropriate objective for many applied, domain-specific research programs. Whether these kinds of findings generalize will depend on how similar the factors related to the causal processes are in the original study setting and the setting that is the target of generalization. By conducting studies in the settings and with the actors to which they wanted to draw conclusions, NDM researchers followed the optimal strategy to produce conclusions with narrow external validity.

Author Note

* Correspondence concerning this article should be addressed to Daniel M. Bartels, University of Chicago Booth School of Business, United States. Contact: bartels@uchicago.edu

In contrast, most scientific laboratory research is motivated to discover more basic—and, for this reason, potentially more broadly valid—principles of human nature. This kind of research aims for broad external validity. The vehicle for generalization is not similarity between the situation where the research is conducted and the target setting. Rather, generalization will be supported if an identified causal mechanism represents a general characteristic of human nature that extends to settings beyond the original studies, and therefore can be broadly applied to understand and control behavior (Mook, 1983; Pearl & Bareinboim, 2014). If that scientific goal is achieved, it will be possible to generalize the theoretical conclusions to many different settings. It is less likely that research conducted in only one naturally occurring setting will achieve high levels of this type of broad external validity.

Many major successes in the scientific enterprise of inducing general causal principles for decision behaviors have been produced by using a variety of controlled tasks as the base of the analysis. Some prime examples would be Kahneman and Tversky's (1979) prospect theory (see also Tversky & Kahneman, 1992); Anderson's (1996) information integration theory; Payne, Bettman, and Johnson's (1993) adaptive decision maker; Brunswik and Hammond's lens model (Hammond & Stewart, 2001); Edwards's intuitive statistician (Weiss & Weiss, 2009); Gigerenzer's (2000) simple adaptive heuristics; and many others.

We also believe that broad external validity will be promoted by research in field settings, but that general conclusions will require studies that span a variety of relevant, naturally occurring domains. Determining whether prospect theory captures broadly generalizable causal mechanisms, for example, required testing its predictions in a broad range of naturally occurring (or laboratory) contexts, such as goal striving (Allen, Dechow, Pope, & Wu, 2016), consumer purchasing (Bell & Lattin, 2000), financial trading (Haigh & List, 2005), taxi driving (Camerer, Babcock, Loewenstein, & Thaler, 1997; Thakral & Tô, 2017), and gambling (Camerer, 2000).

Descriptive Versus Causal Research in the Field

A second key distinction is between research in the field that is primarily descriptive and based on observations versus research that attempts to identify causal relationships in the field based on interventions or complex statistical modeling (e.g., Angrist, Imbens, & Rubin, 1996). Most of the studies conducted in the NDM research tradition have been aimed at description rather than at directly testing causal mechanisms.

As Markman points out, descriptive research "encourages researchers to think about global theories of how choices are made that take into account context, expertise, time pressure, and team performance," and even descriptive findings can call into question simplifying assumptions often made in the lab. For example, Markman cites important findings (fire-fighting, military tactics) where the traditional one-shot, fixed choice-set decision conceptual framework, developed to illuminate laboratory-based research (as well as many everyday examples in medical, legal, and consumer decision making), does not seem to apply. We agree that these are instructive cases in which research in the field can inform conclusions about limitations on the scope of theories and models developed for decisions in different settings. We are a bit less optimistic than Markman, however, about the power of NDM approaches to consistently generate overarching, integrative theories, given their practical focus on specific applications.

A key point here, that we reiterate from Markman's target article, is that NDM research, as well as immersive observational methods popular in sociology and anthropology, are primarily descriptive and are therefore unlikely to produce strong conclusions about causal relationships. (Some critics endorse a more extreme position, that descriptive methods lack methodological power to resolve theoretical conflicts and to falsify precise causal hypotheses.) So, many—but not all—field research programs entail a tradeoff between internal and external validity. Some approaches to field research test the causal effects of real-world interventions, which can provide both internal and external validity.

There is a long tradition of real-world research, rooted in medical research, developmental economics, and social psychology, that relies on "field experiments" or "randomized controlled trials" (Campbell, 1991; Harrison & List, 2004). Field experiments can identify causal effects of an intervention, testing the causal predictions of theories, whether originally based on lab or field research. Field experiments often lack the degree of control and opportunities for process measurement achievable in the lab, limiting the ability of the researcher to dissect the multiple psychological processes that may mediate or moderate an outcome of interest.

When predictions are not confirmed in the field, it can be difficult to zero-in on the specific theoretical claim than should be revised. When faced with lack of confirmation, a researcher whose theory was not confirmed will often note that theoretical preconditions for the hypothesis may not have been instantiated in the less controlled field setting (e.g., maybe people were not attentive to the theoretically significant information, or other incentives conflicted with performing the theoretically relevant task).

Nevertheless, field experiments can be extremely useful for questioning simplifying assumptions in the lab and highlighting the need for more global theories that take the full set of relevant factors into account to make externally valid, broadly generalizable predictions. Field experiments are often ideal for identifying the causal effect of a change (Benartzi et al., 2017) under real-world conditions, and therefore can be very powerful for testing theories that do make strong, falsifiable predictions.

To pick one historical example, in the 1850s when the physician John Snow hypothesized that water contamination could cause cholera, he did not bemoan the lack of internal validity in the field and the lack of external validity in the lab. Instead, he removed the handle of the Broad Street water pump in London and ended the local cholera epidemic, providing compelling experimental evidence for a causal link. Snow's experiment was not definitive in terms of the causal mechanism (cholera was identified in the lab 35 years later), but this crude field experiment provided high levels of both internal and external validity.