Promoting Data Richness in Consumer Research: How to Develop and Evaluate Articles with Multiple Data Sources

SIMON J. BLANCHARD
JACOB GOLDENBERG
KOEN PAUWELS
DAVID A. SCHWEIDEL

As stated in the mission of the *Journal of Consumer Research* (*JCR*) (2022) and a recent editorial (Schmitt et al. 2022), *JCR* is a multi-disciplinary journal where consumer research provides insights about consumers and consumption in the marketplace in a way that meaningfully extends the knowledge from one of our core disciplines (e.g., psychology, sociology, economics) about a consumer-oriented construct. Unfortunately, the labels "consumer research" and "consumer behavior" have come to connote far more than the focus of the work—just as, somewhere along the way, "consumer behavior" and "quant" came to imply a particular type of data source (and associated analysis methods) that is primarily used to study theory and phenomena of interest (experiments vs. "field data"). Why this strong association between consumer-relevant questions, data, and methodology? One reason may be that the field rewards specialization. Another may be due to the incentive structure in business schools (Stremersch, Winer, and Camacho 2021). Nevertheless, the rigid lines dividing the artificially created sub-disciplines are our own making, for better and worse. One way to address this divide and consequently expand the reach of our research beyond those who specialize in our particular sub-disciplines is to use more than one type of data source when addressing a consumer research question. Such data richness is the key theme of this article.

Navigating methodological boundaries is at the center of the field and relevant to journals that are or strive to become "big tents" for research scholars. Consider the decisions that such scholars must make during a research project. They must not only choose data sources (e.g., lab experiments, databases) but also analyze them (analytical tools) in a way that is compatible with their intended contribution—be it theoretical, methodological, or substantive (figure 1). Although some combinations of data sources and analytical tools have become putative (e.g., ANOVA/regressions for experiments, econometrics for natural experiment field data, machine learning for large unstructured datasets), they are not necessary nor sufficient to achieve or achieve prevent publication in any specific journal. With a wide range of data sources and analytical tools available to researchers, we should strive to embrace diversity, bringing researchers from different disciplines into the fold, where the only litmus test, as clearly expressed in *JCR*'s mission, should be that the research puts the consumer front and center.

Experimental methods might be uniquely positioned to speak about the nature of the psychological processes undertaken that would be otherwise inaccessible facets of the customer journey, including sensory information, attention,

Simon J. Blanchard (simon.blanchard@georgetown.edu) is a provost's distinguished Associate Professor and the Beyer Family Associate Professor at the McDonough School of Business, Georgetown University, Washington, DC, USA. Jacob Goldenberg (jgoldenberg@idc.ac.il) is a professor of marketing at the Arison School of Business Administration at Reichman University (IDC), Kanfei Nesharim St, Hertsliya, Israel, and a visiting professor at Columbia Business School, Columbia University, USA. Koen Pauwels (k.pauwels@northeastern.edu) is a distinguished professor of marketing at the D'Amore-McKim School of Business, Northeastern University, Boston, MA, USA, and an adjunct professor at BI Norwegian Business School, Norway. David A. Schweidel (dschweidel@emory.edu) is the Rebecca Cheney McGreevy Endowed Chair in Marketing and a professor of marketing at the Goizueta Business School, Emory University, Atlanta, GA, USA. Please address correspondence to Simon J. Blanchard. The authors contributed equally to this article. They would like to thank Andrew Stephen, Stephen Spiller, and Shane Wang for their helpful comments.

This curation was invited by editors Bernd Schmitt, June Cotte, Markus Giesler, Andrew Stephen, and Stacy Wood.

Advance Access publication 27 April 2022

 • Vol. 00 • 2022
https://doi.org/10.1093/jcr/ucac018

FIGURE 1

FOUNDATIONAL ELEMENTS OF RESEARCH PROJECTS

Data Sources Used

Field Experiments

Lab Experiments
- Scenario-based studies
- Incentive-aligned experiments
- Conjoint-measurement tasks
- Simulations

Observational Data
- Databases
- Natural experiments
- Participant-observation
- Archival data

Surveys
- Primary data surveys
- Secondary data surveys
- Interviews, Informants, Focus Groups

Analytical Tools Used

Agent-based Model

Analytical Model

Consumer Culture Theory

Machine Learning (e.g., computer vision, natural language processing)

Optimization

Statistical or Econometric Model (e.g., ANOVA, regression, time series)

Structural Model

Intended Contribution

Theoretical

Methodological

Substantive

memory, information processing, emotional states, personality, and other contextual factors that influence consumer decisions in-between measurements in observational data. That is, highly controlled experimental methods can help provide compelling evidence for internal validity. Moreover, experimental data may be particularly beneficial when a particular problem is so nascent that few data sources exist or when observational data are unavailable to researchers (e.g., a domain with legal restrictions).

However, one must recognize that observational data can uniquely speak to otherwise inaccessible facets of the customer journey, including antecedents of purchasing (e.g., online search, clickstream behavior, and the consumption of social media content), realized purchase decisions (e.g., panel data derived from loyalty programs), and post-purchase decisions (e.g., posting to social media platforms). Likewise, legal and ethical restrictions may render experiments infeasible (e.g., health effects of smoking over one's lifetime). In many cases, non-experimental methods are required to track consumer behavior over time, study intricate dynamics and long-term effects, and provide compelling evidence for external validity and the possible effect sizes.

The present article investigates criteria and dimensions for what constitutes data-rich articles, that is, articles using more than one data source and their prevalence in the *JCR* over 2018–2021. Second, we take the perspective of experimentalists (empirical researchers), pondering the benefits and risks of adding observational data (experiments) to a series of experiments (observational datasets) and providing examples of best practices and pitfalls. Third, we present recommendations to reviewers who are faced with the challenge of evaluating data-rich articles. Finally, we conclude with general recommendations regarding the state of data-rich research practices.

THE STATE OF DATA RICHNESS IN *JCR*

To investigate how combinations of data sources are captured in *JCR*, we began our data gathering efforts by collecting the citations for all 215 articles published from 2018 to 2021 in *JCR*. After excluding commentaries, conceptual articles without any data (e.g., frameworks), and curations, we arrived at a final sample of 203 articles. Then, we briefly surveyed the articles published in 2021 to develop a tentative taxonomy of data sources that have been recently at *JCR* before, with the help of a research assistant, manually coding each article from 2018 to 2021 as to whether each of the data sources was used at least once. As challenges to our definitions became apparent (such as choosing the right level of depth to ensure the right level of abstraction[1] and distinguishing data source from the analytical tool), we considered other typologies offered in the field (Grewal, Gupta, and Hamilton 2020; Baumgartner, Blanchard, and Sprott 2022), which often conflate data source and analytical method. We discussed among ourselves and finally arrived at the following groupings of data sources: laboratory experiments, field experiments,

1 For example, choosing to group qualitative research methods was particularly difficult as such articles often, by definition, require the use of multiple data sources. The decision to group them into a single category was made because the totality of articles was only 17, not our belief that the data sources are homogeneous.

TABLE 1

TYPOLOGY OF DATA SOURCES IN *JCR*

- **Laboratory experiment**: A study conducted under highly controlled conditions, under which the experimenter controls the independent variable and nearly all sources of extraneous variation. This includes most studies conducted in a lab or through a panel. It also includes incentivized experiments (e.g., participants keep the product they choose at a lab, on mTurk, or doing an on-campus tabling survey).
- **Field experiment**: A study done in a real environment of participants, but although the experimenter has control over the independent variable, it cannot control all sources of extraneous variation. Typical use cases are offering different menus to subsets at a real establishment to paying customers (not first endowed with a participation payment) or conducting an experiment using Facebook ads.
- **Observational data**: A study done in the real environment of participants, but the researcher has no control over both the independent variable and the experimental setting. Most secondary data (whether their analysis involves complex econometrics or not) fall under this category, but it excludes surveys. Typical subcategories are databases, natural experiments, archival data (e.g., social media data scraping), and participant-observation.
- **Survey**: A study conducted under highly controlled conditions, during which participants are asked a series of questions and which the experimenter has no control over the independent variable. This does not include instruments with manipulation but includes both primary and secondary data surveys. It also includes interviews (e.g., informants) and focus groups.
- **Meta-analysis**: A study integrating results from previously published research (usually multiple articles).

observational data, surveys, meta-analysis, and qualitative methods (see table 1 for definitions).

Contrary to the lab, in a field experiment, the experimenter controls the independent variable but not all contextual sources of extraneous variation (e.g., competitor entry, situational factors). In observational data, the researcher has no *a priori* experimental control over both the independent variable and the context, even if characteristics of the context can be used to assess control analytically. However, the dataset analysis can achieve *a posteriori* experimental control by constructing pseudo-control groups (e.g., matching techniques) or in the context of a natural experiment, one can employ analytical techniques such as regression-discontinuity or difference-in-difference estimation. Often, researchers search for tradeoffs between two opposite trends where increased control (e.g., from matching through natural experiments, field experiments, and the lab) strengthens the support for the theory but decreases the external validity and the estimation of the effect size.

We also note that the primary defining characteristic for surveys is whether there is direct questioning of a participant without *a priori* experimental assignment to a condition. As such, we include in surveys both primary data collection (i.e., survey collected for the research project) and secondary data collection (i.e., those collected by third-party). The difference between a dataset (observational data) and a survey can be summarized as to whether the non-experimental data are obtained through direct questioning (survey) or direct observation (dataset).

It is also important to reiterate that a specific analytical tool is neither necessary nor sufficient to label an article as data rich. For example, Thomas (2019) uses a single data source (observational data) to investigate whether television was responsible for a new generation of smokers. Although he provided evidence of internal validity by taking advantage of a natural experiment and using analytical methods, the article would not be considered data rich for our purposes. The same applies to an article that uses Bayesian analyses and convolutional neural networks on experimental data (Blanchard, Dyachenko, and Kettle 2020) or one that only applies text mining to a large unstructured data set. By decoupling the data source and analytical tool used, we define data-rich articles as those which incorporate two (or more) data sources.

Table 2 presents the co-occurrences of the various methods (in a single article) across data sources and some summary statistics. Across the entire sample ($n = 203$), the average number of data sources is 1.48, with 40.39% of articles ($n = 82$) including more than one type of data source. Of course, the majority include two sources ($n = 62$), but some include three ($n = 13$) and some four ($n = 2$). Not surprisingly, the most commonly used data source in *JCR* is lab experiments, with 86.21% of *JCR* articles including at least one lab experiment and 38.86% of articles that include a laboratory experiment and another data source.[2]

From table 2, we can make several additional observations. First, we note that the most common pairing is laboratory experiments with observational data ($n = 34$). Specifically, 19.43% of articles with lab experiments also have at least one observational data source, and 15.43% have a field experiment. Although our category of observational data captures multiple kinds, we have found that the most common include database analyses (Kappes, Gladstone, and Hershfield 2021) and scraped data (Reich, Kupor, and Smith 2018) and are generally used to provide evidence that a substantive relationship found in the lab also replicates in the field. We also find that field experiments are even more likely to be used as a complement to laboratory evidence, as all 27 articles with a field experiment also included laboratory experiments. Concerning surveys ($n = 40$), we find that the majority of articles with

2 While frequency of simple occurrence may vary between journals, the co-occurrences may be a more stable statistic across journals. We encourage future research on this topic.

TABLE 2

TYPOLOGY OF DATA SOURCES IN *JCR* (2018–2021)

	Co-occurrence					Data source statistics	
Method	Lab. exp.	Obs. data	Survey	Field exp.	Meta-ana.	Used at least once (%)	% that are data rich
Laboratory experiment	175	34	21	27	1	86.21	38.86
Observational data	34	55	25	4	0	27.09	87.27
Survey	21	25	40	2	0	19.70	87.50
Field experiments	27	4	2	27	0	13.30	100.00
Meta-analysis	1	0	0	0	3	1.48	33.33
Entire sample							40.39

a survey also include laboratory experiments (52.50%; $n = 21$) and that all but two articles include at least one other source of data (both were scale-development articles).

Finally, in figure 2, we illustrate the dynamics among data sources used. First, we note that the prevalence of data-rich articles seems fairly high and stable over time (percentage of articles which include more than one type of data source; 2018: 41%, 2019: 36%; 2020: 43%; 2021: 44%). Second, we note that although lab experiments remain included in most *JCR* articles (min: 78%; max 95%) and field experiments remain included in very few (min: 8%, max: 16%), there is a notable increase in the number of articles incorporating observational data (from 16% in 2018 to 50% in 2021). As to methods, we also see an increase: from an average of 1.44 in 2018, to 1.42 in 2019, to 1.5 in 2020, and finally 1.61 in 2021. We may expect this trend to continue.

Although it appears that publishing data-rich articles is now common, and the average number of data sources used has increased, combining data sources still raises concerns. For one, it is common to believe that such articles require greater investments of time and effort and that reviewers may be less efficient or even more critical in their assessments of methods that are not part of their toolbox. We address these issues in turn.

DEVELOPING DATA RICHNESS TO BOLSTER AN ARTICLE'S CONTRIBUTION

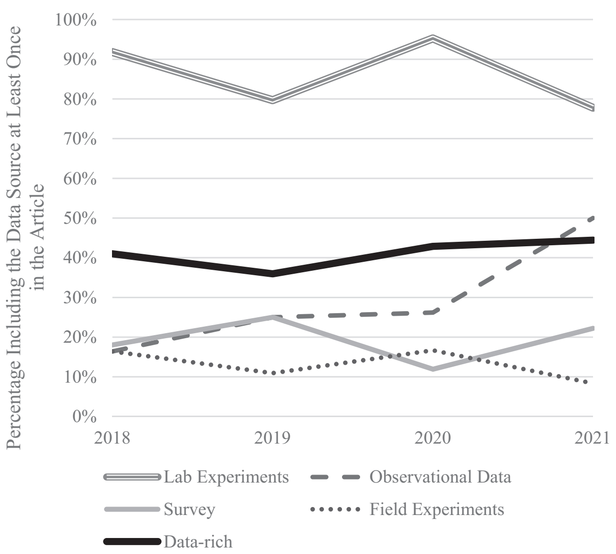
While there are advantages conveyed by data-rich research, it is often less parsimonious by nature. For example, the philosopher William of Ockham, connecting simplicity to scientific truth, argued that "plurality should not be posited without necessity." This statement led to the articulation of more specific principles, including "Of two competing theories, the simpler explanation of an entity is to be preferred" and "Entities are not to be multiplied beyond necessity." These principles are part of a scientific school of thought that aspires to achieve elegance in theories by making them as simple as possible. Yet, parsimony is not a standard in consumer research. As such, Ockham's razor may be less relevant given two fundamental requirements for empirical consumer research: (1) the validation of empirics must be evaluated in the context of the intended contribution (i.e., theoretical, methodological, or substantive) and (2) single-study research designs vary in the kind of validation that they enable.

It is well known that research designs influence our ability to validate causal inference (Cook and Campbell 1979). Whereas some research designs enable high internal validity (i.e., confidence that the treatment influences the outcome), others offer external validity (i.e., the generalizability of the treatment outcomes to other situations). Depending on an article's intended contribution, the kind of evidence needed may vary, and as such, should the likely composition of the article's basis for empirical evidence. If the authors' primary intended contribution is theoretical in consumer decision-making processes (e.g., a new process, a novel moderator), they can be expected to use data that enable a high level of internal validity. If authors' primary intended contribution is substantive (e.g., studying how a theorized effect across purchase contexts, quantifying the real-world impact of a strategy), they can be expected to use data sources that enable a high degree of external validity.

In our view, the authors' work should thus begin with a clear exposition of their intended contribution. For authors seeking to make a strong theoretical contribution, the primary validation needed should be interval validity. Lab experiments may be sufficient as the experimenter controls both the treatment and context. To the extent that authors can show across studies that their scenarios generalize (e.g., showing the findings are not due to a very specific execution of the stimuli or the choice of a peculiar product category) and use real behavior or incentive-aligned studies, authors should be able to achieve the necessary level of theoretical contribution while providing reasonable evidence of external validity. For authors with a strong substantive contribution, the primary validation needed should be external validity, and observational data may prove to be sufficient even though the experimenter does not perfect control the treatment and context. Finally, to the extent that

FIGURE 2

EVOLUTION OF DATA SOURCE USAGE (SUMS UP TO MORE THAN 100%)



NOTE.—Not plotted is the *average* number of data sources used. The average is increasing from 1.44 (2018) and 1.42 (2019) to 1.5 (2020) and 1.61 in (2021).

authors can either take advantage of the nature of the data generating mechanism (e.g., a natural experiment occurred) or use appropriate econometrics, authors should be able to achieve the necessary level of substantive contribution while providing reasonable evidence of internal validity by ruling out alternative explanations (Goldfarb, Tucker, and Wang 2022).

Although a single data source should prove to be sufficient in many cases, there are many situations where authors conceive an intended contribution that spans the theoretical and the substantive. Such situations may occur in a variety of ways. For example, authors who use only lab experiments would like to make statements regarding the generalizability of their effect (e.g., effect sizes in the real-world applicability across domains). Authors who have a single observational dataset would like to make statements regarding an underlying decision-making process that cannot be measured using the available data. In such situations, adding a different kind of data source may offer an advantage because it can be used to provide evidence that would be difficult (if not impossible) to obtain otherwise. At the same time, it is often fraught with difficulty because, while authors are familiar with handling one source of data, they often do not have the same facility in working with the "other" source. In the remainder of this section, we first take the perspective of authors who want to augment an article's existing data sources to improve external validity and second the perspective of authors who want to augment an article's existing data sources to improve internal validity. We will focus on two recent articles (Spiller and Belogolova 2017; Wang et al. 2021) as thread examples of how to construct a sequence of studies to take advantage of data richness and how to navigate the review process. Of course, many other articles could have been chosen for our purposes. Table 3 highlights other articles whose assemblage of studies enabled the authors to cross the proverbial publication finish line.

Adding Data Sources to Improve the Substantive Contribution

External validity is concerned with generalizability to other situations, which includes demonstrating that the phenomenon occurs "in the real world" (i.e., ecological validity) that the causal inferences generalize across contexts and situations (e.g., companies, types of products) and allow us to speak to the significance of the effect in the context of many other effects occurring in said real world. Field experiments and observational data are more natural complements to research already presenting strong internal validity, depending on the kind of external validity sought to achieve the desired substantive contribution. For example, is the goal to provide evidence that the effect can generalize to the field? That it applies across other purchase situations? To substantiate the size of the effect in the field? When contemplating adding a data source to improve external validity, we recommend that authors focus on:

1. Can it be used to help generalize to the field *and* across consumption contexts?
2. Can the data (or analytical tools) provide a sufficient level of internal validity?

Observational data (e.g., natural experiments, datasets, scraped social media) are well-suited to remedy concerns of ecological validity especially when it is difficult to find a partner for field experiments. However, observational data are not a panacea. The first area of key importance is: among all potential observational data (e.g., brand tracking, purchase, consumption, marketing, competitive, financial) and analysis methods, why are the data selected appropriate to speak to the kind of external validity sought?[3]

Consider Spiller and Belogolova (2017), who examined consumer beliefs about quality versus taste. As their contribution was primarily intended as theoretical, the authors first produced a series of single-context experiments (e.g., about olive oil) to establish the internal validity of the effect. Then, they sought to provide evidence of external

3 Our examples focus on the inclusion of observational data instead of field experiments, as they tend to be more commonly accessible and a recent editorial provides example guidance on the successful use of field experiments (Nelson, Simester, and Sudhir 2020).

TABLE 3

EIGHT RECENTLY PUBLISHED DATA-RICH ARTICLES (AND HOW DATA-RICHNESS HELPED)

Article	Primary intended contribution	Data sources used	Benefits from data-richness
Sussman, Paley, and Alter (2021)	Theoretical: Consumers select larger portions of infrequent foods because they believe they have a smaller impact on their weight.	Archival data from a popular calorie tracking app and five experiments.	The observational data allowed the analysis of actual consumption over time, which was not possible with experiments. Then, the experiments allowed not only to investigate the process but also to test manipulations that influence perceived infrequency. Neither would be possible in observational data for their research question.
Kappes, Gladstone and Hershfield (2021)	Substantive: consumers hold beliefs that spending implies wealth, and that they impact financial decisions.	Survey administered to 2000 app users, experiments, and lab survey (scale).	Although the field survey only measured a single item, it allowed to show that consumers do hold the beliefs that spending implies with objectives financial measures. The panel survey helped validate the multi-item scale, and the lab experiment showed beliefs could be manipulated. Every study served a clear purpose.
Streicher, Estes, and Büttner (2021)	Theoretical: attentional breadth affects impulsive purchasing through an exploratory mindset.	Field experiment and observation at supermarket, and lab experiment with eye-tracking.	The authors combined two different field studies: one of observing consumers' intended shopping and one who gave a manipulation through a hypothetical task. Jointly, they provided a mix of external and internal validity effect in a way that would be possible in a single study.
Goor, Keinan, and Ordabayeva (2021)	Theoretical: status threat leads to displaying success in alternative domains.	Quora posts, bumper stickers, field experiment at marathon, and lab experiments.	There are many ways to instantiate status threats, and many correlates to status threats. The variation across field data (e.g., car value, mentions of successful peer) helped make the results appear less ad hoc than if a single field study had been used.
Packard and Berger (2021)	Substantive: using concrete language influences perceptions of helpfulness.	Archival of customer service phone calls and lab experiments.	The authors used NLP not only to quantify concreteness but also control for numerous observable language and voice factors. For both field studies, the authors also provided in a Web Appendix Lasso and Ridge regressions to show the estimates remain stable even while being agnostic to model selection. This was essential for a contribution of showing the importance of concreteness.
Kalra, Liu, and Zhang (2020)	Substantive: how the round number bias affects preference for target retirement funds and wealth.	Dataset of defined contribution plans and laboratory experiments.	To quantify the bias impact, the authors had to rule out alternative explanations but lacked exogenous variation in the treatment (funds ending in 5 or 0) or individual parameters for risk aversion. Their use of simulations and experiments provided compelling evidence, yet lab experiments allowed them to carefully rule out selection effects due to differences in age cohorts.
Garbinsky et al. (2020)	Theoretical: Introducing the financial infidelity scale.	Surveys, lab experiments, money management app.	Following traditional steps in scale development, the authors conducted numerous surveys of the population of interest (couples) and even did controlled intercepts on campus. The survey data within a couple's money management app allowed them to substantiate the relationship with a behavioral outcome which, although not causal, was only possible in the field.
Kim, Barasz, and John (2019)	Theoretical: Introducing acceptability of information flows and showing it influences how ad transparency affects effectiveness.	Surveys, lab experiments, field experiments.	Having demonstrated the theoretical relationships in the lab, the authors sought to demonstrate in the field. A particular challenge was that key theoretical moderators could not be manipulated or controlled for in the field (e.g., trust in the website), which the authors transparently acknowledged. Yet, their careful use of pre-tests allowed the authors to substantiate mitigate concerns about how the chosen websites theory inconsistent levels of moderators.

validity. Their approach was multi-pronged as they used two additional data sources: surveys and observational data. First, they addressed generalizability concerns by conducting three large-scale surveys in which perceived objectivity (i.e., beliefs about whether a choice is a matter of quality or taste) was measured along with key dependent variables relating to the willingness to pay and self-referencing.[4] Although such surveys did carry inherent limitations due to the measurement of treatment and the outcome, they went beyond correlations and included numerous alternative specifications, including ones that incorporated product-pair and chosen option fixed effects, and random intercepts at the participant level. These three collapsed studies, and careful analyses, eventually provided strong evidence to show that the relationship observed in follow-up experiments where the treatment is manipulated (e.g., studies 2A and B) was not due to the idiosyncratic choice of the product category. Nevertheless, their hypothetical scenarios (even if they had been incentive-aligned or "real behavior") could not possibly enable the authors to speak about generalizability to the field nor quantify the possible effect in the real world. Therefore, second, they gathered observational data of consumers and critics reviewing movies (study 3) to investigate the effect in the field and proxying the degree to which individuals believe the quality is a matter of taste by measuring self-referencing.

For an article that sought to make a theoretical contribution about our understanding of the process and consequences of heterogeneity regarding beliefs of quality versus taste, it was primordial that the authors established a high degree of internal validity and their lab experiments enabled them to do so. Moreover, even though their large-scale cross-category lab experiments enabled them to show that the effect generalized to other consumption contexts, the authors still lacked evidence that the effect would generalize to the field.[5] With that in mind, the observational data from movie evaluations provided evidence for ecological validity. However, it provided only some confidence concerning internal validity in these particular data.

In this story, it is important to understand the limitations of the data concerning contribution. Had Spiller and Belogolova (2017) not provided numerous lab experiments that varied the consumption context across brands and product categories (studies 1A–C), the addition of the observational data about movie reviews would have seemed ad-hoc and the inability to econometrically or structurally control for alternative explanations could have led to disbelief about internal validity. Yet, as the article had clearly exposed its aim and evidence for a theoretical contribution and that the field data were used to provide a sufficient degree of confidence for ecological validity, the entire package proved to be sufficient.

We advise the following with regard to the inclusion of observational data to improve external validity. First, the authors should make sure to indicate *how* the additional observational data are intended to fit as part of the contribution of the article, depending on the goal. For example, suppose the authors wish to present evidence that the effect generalizes to the field. In that case, the authors need to expend a considerable amount of space not only saying so but also arguing how the measures used in the observational data (no longer under the control of the experimenter) are good proxies for the treatment administered in the lab (e.g., self-referencing in Spiller and Belogolova 2017). They also need to argue why alternative explanations are possible but unlikely. For example, authors need to expect sample representativeness (e.g., how well movie reviews capture matters of quality versus taste) and attrition (the kind of reviewers might change over time and influence one another).

Second, the authors cannot outright dismiss concerns of internal validity in observational data solely to the existence of other data (e.g., experiments). Whereas it may be tempting to ignore concerns of internal validity when using observational data, it is important to recall that some level of internal validity is still required at the data level if these data are going to be used to make claims about the generalizability of a relationship to the field or make comments about possible effect sizes. As such, observational data are more often than not ineffective when introduced as "only for correlational evidence" without consideration of alternative explanations due to the lack of random assignment and manipulation control typical for an experiment. The research needs to anticipate questions such as: is there simultaneity (i.e., Y and X variables drive each other over time instead of X causing Y; e.g., critics justify scores using self-reference in Spiller and Belogolova 2017)? Do the authors observe a positive correlation because of omitted variable bias (e.g., less reliable critics use more self-reference language)?[6]

Even when intended only to boost external validity, the use of observational data requires careful consideration of analysis techniques to ensure that claims can be made with sufficient internal validity. Methods such as regression and ANOVA, likely appropriate when the experimenter controls the treatment assignment and the context, may not be appropriate for observational data. For one, the relationship between X and Y may be curvilinear, for example, inverted U-shape (with an optimum level of X) or diminishing returns. For another, a high standard error for coefficients may be due to multicollinearity among observed variables. When used for data captured over time, regression also

4 See Kim, Barasz, and John (2019), Garbinsky et al. (2020), and Kappes, Gladstone, and Hershfield (2021) for helpful examples of research using primary data surveys as part of its empirical evidence.

5 See Garbinsky et al. (2020), Goor, Keinan, and Ordabayeva (2021), and Sussman, Paley, and Alter (2021) for other excellent similar examples who used observational data to augment external validity in an article primarily aiming at a theoretical contribution.

6 Panel data structures, when enabling fixed effect estimation, can be helpful to address concerns regarding time-invariant omitted variable (e.g., stable personality traits). See Kettle et al. (2016) for an example.

assumes that variables are stationary and thus rule out permanent effects (Dekimpe and Hanssens 2000; Slotegraaf and Pauwels 2008). Such issues can be addressed through measurement (e.g., partial least squares), by linking errors across consumer decisions (e.g., seemingly unrelated regression), by explicitly modeling dual causality over time (Pauwels 2018), or by reducing variance and random error by combining models (such as ensemble models, random forests, and XG Boost in machine learning).[7] We recommend that researchers interested in using observational data to not only provide evidence of ecological validity but external validity more broadly (e.g., counterfactuals, quantify effect sizes in the field) carefully consider whether their observational data and analytical techniques used to enable them to have sufficient internal validity to make such statements. For instance, cross-sectional data (e.g., "consumers who do X, experience Y") face the internal validity threat of other ways consumers differ. In contrast, time-series data (e.g., "consumers that started doing X, experienced Y") need to show whether the periods did not differ in other ways. Particularly, as the multi-source research moves from the theoretical to the substantive, we believe it is important for authors to recognize that the combination of multiple data sources does not provide shelter from discussions that pertain to the internal validity of observational data.[8]

Adding Data Sources to Improve the Theoretical Contribution

Internal validity is provided when the researcher has used controls to determine that the outcome is due to changes in the treatment. Because the researcher has already controlled the treatment and situation in data from laboratory experiments, their design tends to enable internal validity and causal statements. In field experiments, the researcher loses some control over the situation (i.e., by controlling the treatment but not the situation) such that she gives up some internal validity to external validity.[9] In the observational data, the researcher does not control the treatment or the situation at the design level. Internal validity is harder to justify but requires the researcher to provide evidence that, although alternative explanations are possible, they are unlikely. They may do so entirely sufficiently by taking advantage of the unique nature of the observational data (i.e., a natural experiment, see Thomas 2019) or through careful econometric analyses. Indeed, an appropriate research path is to begin with a phenomenon observed in the observational data and then unpack it by leveraging exogenous shocks that produce variation for a quasi-experiment (Goldfarb et al. 2022). Yet, laboratory experiments can be a natural complement to research already presenting strong external validity, depending on the nature of the intended contribution. Is the goal to improve internal validity by having an experiment replicate the effect under more controlled environments? Or is the goal to provide evidence for an underlying process that cannot be observed in the field? When contemplating adding a data source to improve external validity, we recommend that authors focus on two areas:

1. Is the goal to provide additional internal validity for the effect or the process?
2. Can the experimental design provide a sufficient level of internal validity?

Wang et al. (2021) wished to substantiate the importance of three vocal tone features (focus, low stress, and stable emotions) for persuasion. As they had begun their investigation using crowdfunding data campaigns from Kickstarter only in the music category, the authors needed to replicate their analyses across another category to ensure that the findings would not be limited to the unique context of music startups on crowdfunding websites. Moreover, the data had been carefully augmented with numerous controls such as alternative vocal tone variables, project characteristics, and the history of the artists. The findings were significant, and reliable across the two categories.

Suppose that the authors had stopped here. While the authors may have sufficient external validity and findings for a substantive contribution around the importance of vocal tones, the empirical analyses presented in studies 1 (music) and 2 (technology) would be insufficient to make the desired theoretical contribution about the specific role that the three emotions play in persuasion. Not only does the analysis of the observational data as presented fail to shed light on the underlying mechanism, but the analysis as presented does not provide sufficient internal validity to make the alternative explanations improbable. In study 3, Wang et al. (2021) replicate their findings in the lab. In addition to looking at the impact of vocal characteristics on persuasiveness, they use this study as an opportunity to collect information on perceived competence, positing (and showing) that this mediates the relationship between vocal characteristics and persuasiveness. Lacking study 3, the authors could only speculate about the underlying mechanism because such a measure was not readily available in the secondary data to which they had access. In this case, the addition of the laboratory study replicates the analysis of data collected in the field and allows the authors to

7 For example, Packard and Berger (2021) provided atheoretical robustness checks for variable selection (using Lasso and Ridge) in their appendix.

8 Kalra, Liu, and Zhang (2020) provide an excellent example of how alternative explanations can be ruled out both theoretically and through robustness tests.

9 An excellent example is in Kim, Barasz, and John (2019) who could manipulate their treatment in the field but could not control for an important theoretical moderator (trust).

understand better the relationship between vocal characteristics and persuasion that they observe.

We advise the following concerning the inclusion of experimental data to improve internal validity. First, suppose the authors wish to present evidence that the effect holds under tightly controlled condition (i.e., the observational data do not effectively rule out alternative explanations). In that case, the authors need to explain which sources of extraneous variation the experiment can help rule out. In the case of Wang et al. (2021), a concern was that numerous persuader factors (e.g., liking) and trust could determine both the use of tone and persuasion outcomes. As such, they opted to manipulate tone while holding personality treats constant. To do so, they build on the research on acoustics to carefully control for pitch length and text. The specific manipulation was able to control for alternative explanations by design—the main strength of experimental work. A common mistake at this stage is to think that selectively (or even randomly) sampling from the observational data suffices. Had Wang et al. (2021) chosen some projects from the observational data that vary in tone and asked participants to evaluate their persuasiveness, they would have only controlled for the situation but not the treatment and the usefulness of the lab experiment data would have been limited. As many measures found in the field can be manipulated in the lab, it is of utmost importance that the added experiments control not only for situational factors but also do not introduce novel confounds or keep confounds in the treatment.[10]

Second, if the authors wish to provide empirical evidence of the underlying psychological process through the experiment, it is important that they think not only about how they will manipulate the treatment but also how they will design the experiment in such a way that it can provide process evidence. For example, in the case of Wang et al. (2021), the proposed process was three audio features (focus, low stress, and stable emotions) improved persuasion through perceived competence and not alternative process explanations such as trust and processing fluency. This required the authors to anticipate alternative process explanations and examine the process evidence for competence while accounting for covariation in trust and processing fluency. Common mistakes at this stage are failing to also measure for probable expected alternative explanations (e.g., the authors had only measured perceived competence, not trust and fluency), failing to select scales whose psychometric properties have been shown carefully, and failing to properly take measurement error into account while investigating statistical evidence for the process.

Even when intended only to boost internal validity, the use of laboratory experiments requires careful consideration of research design and analysis techniques. The use of an experimental research paradigm is not a shortcut to careful design that maintains internal validity. Authors are also encouraged to *a priori* determine the necessary sample sizes for their experimental studies (Meyvis and Van Osselaer 2018) and preregister their analyses while minimizing "researcher degrees of freedom" (Simmons, Nelson, and Simonsohn 2011). In addition, if the observational data are generated from a particular subpopulation and the authors wish to replicate the data using experiments and panels such as mTurk to provide evidence of internal validity (e.g., finding gamers, people with credit card debt), it is recommended that they follow the two-stage panel approach recommended by Sharpe Wessling, Huber, and Netzer (2017) to minimize risks of misrepresentation as opposed to relying on statistical controls using measures that may be biased.

10 A similar approach was used in Packard and Berger (2021).

NAVIGATING THE REVIEW PROCESS OF DATA-RICH ARTICLES

Although scholars recognize the potential benefits of gathering types of data sources for their research, data-rich articles are commonly believed to suffer from a structural disadvantage in the review process. As a result, editors often assign data-rich articles to a multi-disciplinary team of reviewers (e.g., reviewers used to experimental design vs. observational data). However, a more methodologically diverse review team is more likely to result in a divergence of recommendations, leading AEs and editors to struggle with identifying a singular achievable path for revision that would please the entirety of the review team. In this section, we highlight what we consider to be important issues that warrant the review team's attention.

Assignment of the Review Team

When editors first examine a manuscript, they perform a cursory read to identify how it is positioned, focusing on identifying a team familiar with the type of data source, methodological methods, and intended contribution domains (figure 1). For single data-source articles whose intended contribution aligns well with a large number of available reviewers and associate editors, the editor may assign an associate editor familiar with the theoretical framework (e.g., goal pursuit) or with the substantive domain (e.g., donations) as the type of data-source chosen (e.g., lab experiments or field experiments) may align well with what has previously been done in the literature. Then, the editor might consult the associate editor about developing a list of reviewers who are each familiar with the theories used, the research designs used, and the analytical methods used. However, as data-rich articles tend to include multiple data sources and analytical techniques, it is unlikely that the editor can assemble a team of reviewers who *each* possess expertise with all theories and research design and analytical methods used throughout the entirety of the manuscript. The editor may assemble the review team with different roles in mind depending on the intended contribution and the role that each data source plays in the manuscript. To illustrate, let us consider potential review team assignments for the two articles previously discussed: Spiller and Belogolova (2017) and Wang et al. (2021).[11]

Spiller and Belogolova (2017). In Spiller and Belogolova (2017), the editor reading the manuscript for the first time likely recognized that the manuscript first and foremost focused on making a theoretical contribution about which judgments are objective and which are subjective. As the editor likely considered a review team familiar with the theories used (e.g., perceived objectivity, fairness, self-referencing), he also considered the kind of evidence provided by the authors to assert the validity of their research. Noting both numerous (9) large experiments on mTurk and an observational data source from moving ratings (study 3), it is likely that the editor would assign an associate editor that is familiar with not only the theories used, but also with the most frequently used research design: laboratory-experiments. Then, the editors would ponder the role of the observational data in arriving at the intended contribution. Is the observational data meant to present the evidence of ecological validity (i.e., that the effect can be reasonably expected to occur in the field) or meant to substantiate the magnitude of the effect in the field (e.g., enable counterfactuals, substantiate effect sizes) in a given substantive area (e.g., movie ratings)? Given that Spiller and Belogolova (2017) positioned observational data as a way to provide external (ecological) validity to a primarily theoretical article, editors likely assigned a trusted reviewer who is generally comfortable with observational data analyses (without having to be a movie expert) and can evaluate the observational data in light of its intended contribution.

As a key takeaway, the positioning of observational data analyses is key to the assignment of the review team and, ultimately, how the manuscript is evaluated. For example, had the authors stated an intended contribution about substantiating the monetary consequences of beliefs about quality and taste in the context of movie reviews (e.g., self-referencing explains 20% of rating differences in reviews), the authors would have likely been assigned a reviewer who would have focused on whether the data enabled the authors to make such causal claims and whether their econometric analyses allowed for such counterfactuals. By not overclaiming their observational data and focusing on its role in providing ecological validity, the authors clearly defined how the manuscript should be evaluated. It also facilitated the assemblage of a review team that likely focused on the theoretical contribution while acknowledging that the observational data helped the authors to make a good case for external validity.

Wang et al. (2021). In Wang et al. (2021), the editor likely identified that the primary intended contribution was the identification of novel predictors (e.g., tone) of consumer behavior in the field. As such, the editor likely sought an associate editor with expertise with observational data and was comfortable with experimental design. Yet, given that the authors' data sources included observational and lab experiments (data sources), built on the persuasion literature (theory), and used text mining to generate predictors (methods) in the crowdfunding area (substantive), it is unlikely that the editorial team could assemble three reviewers who each possess expertise across all of the data, theory, method and substantive areas. Several approaches exist to assemble a team for such an article. First, the editors could add reviewers who complement their skills. For example, in an article that mostly relies on observational data, with editors most familiar with observational data, the editors could choose to invite (1) a reviewer mostly used to experimental design (data source) and an expert on persuasion (theory) while also inviting (2) a reviewer mostly used to observational data analyses (data source), text mining (methods) and perhaps even substantive domain expertise (crowdfunding). The risk of such an assignment is that reviewers might focus exclusively on different elements of the article (reviewer 1 focusing on contribution to the persuasion literature and internal validity of the experiments; reviewer 2 focusing on the substantive contribution to the crowdfunding domain and the adequacy of the text mining measures).

For Wang et al. (2021), how the studies were organized and introduced was critical to the expectation of how the review team would be formed and how different reviewers would evaluate it. Study 1 (technology category) introduced the measures and demonstrated the effect. Study 2 (replication in the music category) provided evidence that the findings were not due to the idiosyncratic selection of technology as the first category (i.e., increases external validity). Study 3 (lab experiment) was introduced and positioned as a way to provide additional internal validity (i.e., control the treatment) and rule out alternative explanations (i.e., investigate process) in a way that was not available in the data that the authors had. In doing so, they avoided stretching each study beyond what was possible and invited the review team to consider validity across multiple studies.

Several elements emerge across these two manuscripts in what makes a data-rich article undergo a successful review process. First, the authors clearly stated the role that each data source was meant to play in the context of their intended contribution. Doing so allowed the editorial team to assemble a set of reviewers who could evaluate the manuscript in a balanced way and evaluate the manuscript as a

11 Our discussions here are fictional and are not meant to convey the actual assignment of the review team for these two articles.

whole. Second, the authors stayed within what their data sources enabled them to claim. For example, although it might have been possible to do so with additional data, Spiller and Belogolova (2017) avoided claiming that their movie review data enabled the calculation of precise counterfactuals. Likewise, Wang et al. (2021) avoided in their studies 1 and 2 to make claims about the decision-making process and instead relied on the experimental data.

Reviewing Data-Rich Articles

Reviewing a data-rich article can be challenging. The use of multiple data sources can enable researchers to extend their contribution, and studies should be evaluated based on their intended contribution to the overall research paradigm. Rather than simply viewing the incorporation of additional data sources as a shield with which to defend the weaknesses associated with a single data source, we encourage viewing supplemental data sources as additions to the arsenal by which the overall contribution can be enhanced by providing additional insight to address elements that would not be possible otherwise (e.g., only the observational or field experimental data can be used to substantiate effect sizes in the field). This section provides recommendations as to how we suggest reviewers approach data-rich articles.

The Evaluation of Each Study (in Isolation). As discussed in the prior sections, adding a data source (e.g., observational data or lab experiments) does not eliminate validity concerns. Adding lab experiments to observational data does not license one to use lab experiments with confounded manipulations, just as adding observational to lab experiments does not license one to dismiss probably alternative explanations. Reviewers can (and should) expect that each study stands on its own, such that a minimum level of internal (external) validity is needed when laboratory experiments (observational data) to improve the internal (external) validity of a manuscript relying otherwise mostly upon only observational data (laboratory experiments). In other words, each part must have some study-level minimum level of validity.

For experiments, does the manipulation move more than the intended construct? A common mistake with researchers designing experiments to supplement observational data is that their manipulations are not pretested or taken from past literature. As such, authors open themselves to criticism that their manipulations either move something else (e.g., are you manipulating trust with changing in tonality?). To the extent that it is possible, relying on prior manipulations of given constructs is advisable. Second, authors should investigate whether the experimental design can introduce additional biases. For example, demand effects and scenarios that mention more than they intend to manipulate are problematic. Third, evaluate whether the outcome variable is measured via appropriate scales and free of measurement error. There are many standard scales with validated psychometric properties, and employing them is essential to claim internal (and construct) validity.

For observational data, the study should include evidence that would provide a basis for causality. That is, endogeneity concerns should be minimized. Endogeneity shows up as a correlation between the dependent variable and the error term, often because decision-makers act on variables unobserved to the researcher (Papies, Ebbes, and van Heerde 2017). Therefore, endogeneity concerns can be addressed by explicitly incorporating and modeling such actions (Gijsenberg, Van Heerde, and Verhoef 2015), or by the correct use of instrumental variables (Papies et al. 2017; Rossi 2014). Should this not be possible with the available data, the authors should clarify why this cannot be done.

Moreover, authors should justify how the observational data (or field experiments) provide a reasonable proxy for what is manipulated in the lab. For example, although Spiller and Belogolova (2017) could not find a dataset that directly measures perceived objectivity, they used the percentage of words that are first-person singular pronouns as their "measure of self-referencing and a proxy for (low) perceived objectivity." To what extent does self-referencing capture other elements than perceived objectivity? Reviewers may criticize this approach to measuring perceived objectivity, but only to the extent that they name an alternative explanation that correlates with self-referencing and predicts the relationship between self-referencing and the outcome.

The Evaluation of the Integration across Studies. First, reviewers should look for inconsistent *application of the treatment* across data sources. Often due to data availability constraints, there can be a mismatch between the variables measured, used, or manipulated in the lab. For example, when the variable in the field is a complex construct (e.g., trust), a simplified version is used in the lab because it is easier to manipulate.[12] The gap between the complex construct and the manipulation causes a disconnect between the two studies.

Second, reviewers should look for inconsistency in how *sampling* may be responsible for the effects. For example, senior managers (used in the field study) and MTurk participants or students (used in the lab) differ substantially, so if the phenomenon is not about a fundamental human behavior, each set of studies may have different potential explanations.

Third, reviewers should look for inconsistency in sampling *duration*. Sometimes a phenomenon in a lab test is

12 For example, Kappes, Gladstone, and Hershfield (2021) only used a single-item from their scale in the field but well justified the approach.

accelerated to minutes or seconds, while the phenomenon in the field study has a time scale of days, weeks, or even months. A mismatch in the time scale is not necessarily a fatal flaw—the reviewers must decide whether the accelerated scenario is consistent with the field dynamics. If not, they can ask the authors to provide more evidence for a match.

Fourth, reviewers should look for coherence in the applicability of *explanations*. For example, if the process is revealed in lab studies, some observational data may support the same mechanism. If such coherence is absent, it is often possible to ask authors to augment their observational data with process measures by codifying unstructured elements in their data or employing human judges to codify stimuli. Because such tests are sensitive, and the main goal is to show coherence, the effect sizes (or even significance) are often not as strong as the focal effects or hypotheses that the article presents. Given the importance of the link between the studies, reviewers may advise authors to write a special section presenting the connecting analyses to make it easier for reviewers and the readers.

Writing the Review. An advantage to data-rich articles is that they can allow authors to draw upon different sources to provide internal (vs. external validity) evidence. However, it is fairly common that authors attempt to stretch the implication of each study beyond what is feasible at that study's level. For example, suppose the authors use unwarranted causal statements from observational data without proper exogenous variation or the use of causal inference techniques. In that case, it is acceptable to point it out. Likewise, if the authors use lab experiments to generalize effect sizes to the field, it is acceptable to point it out. Yet, we encourage review teams to assess the entire package holistically. From our experience, doing so will require more patience with the authors (allowing for slower convergence).

A few additional points are warranted as advice to reviewers. First, reviewers are not assigned articles randomly. Suppose a reviewer feels that they are unqualified to review certain elements of a manuscript. In that case, they can either (1) do so to the best of their ability or (2) reach out to the editor to see whether they were invited for a specific assessment. Either way, reviewers should elaborate on the focus of their evaluation (if any) in private comments. Second, reviewers should refrain from reductive comments along the lines of "no observational data can provide causal evidence" or "using mTurk leads to garbage data." Third, confounds should be named. It is neither sufficient nor helpful to decry endogeneity (in observational) or confounded manipulation (in experiments) without stating the plausible alternative explanations (the fundamental review error; Lynch 1998). Ideally, an alternative (e.g., suggested analysis, coding of an instrument, alternative manipulation) should be provided to the authors. While this is true of the reviews of any article, it is particularly helpful in the case of data-rich manuscripts because of the necessary holistic assessment of the analyses presented in the manuscript. Absent specific guidance and discussion of how the proposed alternative would address shortcomings, the review may not provide the AE and editor with the full scope of reviewer concerns.

GENERAL DISCUSSION

In the present article, we have defined data-rich articles as employing more than one type of data source. We then explored the state of data richness in *JCR* by coding 203 empirical articles for the period of 2018–2021, from which we noted that data-rich articles have become more common in the journal. Although we see the increase as laudable, we believe that there still exist important misconceptions about conducting research using multiple data sources: (1) they are believed to be more difficult to conduct and (2) they face an undue burden in the review process. As such, we reviewed two data-rich articles (Spiller and Belogolova 2017; Wang et al. 2021) to illustrate how such articles can achieve the standards needed for publication and develop a series of recommendations for authors who wish to conduct similar data-rich research.

First, authors should carefully consider the role of the additional data sources as a function of the intended contribution—be it theoretical, methodological, or substantive. For example, whereas some data sources such as lab experiments (observational data) provide greater internal (external) validity by design, the addition of another data source such as a field experiment or observational data (lab experiment) is only helpful to the extent that it also helps an important substantive (theoretical) contribution by improving external (internal) validity.

Second, authors should recognize that assembling data sources into a single article does not license them to neglect proper treatment of validity. Laboratory and field experiments' manipulation should be free of confounds that diminish internal validity. Observational data and surveys should be analyzed with proper treatment of measurement error and take advantage of methods that enable better substantiation of internal validity. The use of multidisciplinary teams is encouraged.

Third, the authors should describe how the combination of data sources helps achieve the intended contribution. The review process for data-rich articles can be arduous when the editorial team is not sufficiently guided on how to evaluate the article. Just as stretching the internal (external) validity of observational data (lab experiments) can be misguided, so is the reverse.

Fourth, reviewers should evaluate the totality of the article as much as possible. We have argued that authors should focus on evaluating the individual adequacy of the contributions of the data sources at the study level and as a totality. As a totality, reviewers should pay particular attention to inconsistencies in the treatment administered, drastic differences in the sampling frame (population, duration), and the coherence in the explanations across data sources.

Yet, the question remains. Should an article be data rich by combining more than one source of data (and likely, method of analysis)? We would be remiss if we did not acknowledge that a single data source suffices in many cases. In some cases, authors can use observational data to rule out alternative explanations by convincingly demonstrating that while some explanations are possible, they are improbable. For an article primarily focusing on a substantive contribution, the observational data can be sufficient. Likewise, authors can use lab experiments with a high degree of ecological validity (i.e., rich scenarios, adding incentive alignment, and real-behavior lab experiments) to provide evidence that the study findings would generalize to the field. For an article primarily focusing on a developing theoretical contribution, such lab experiments can be sufficient. We believe, however, that nurturing the development of data-rich articles is important for the field not only because of its potential to break down our academic silos but also because it has the potential to improve our community's impact.

REFERENCES

Baumgartner, H., S. J. Blanchard, and D. Sprott (2022), "The Critical Role of Methodological Pluralism for Policy-Relevant Empirical Marketing Research," *Journal of Public Policy & Marketing*, 41 (3), DOI: 10.1177/07439156221092010.

Blanchard, S. J., T. L. Dyachenko, and K. L. Kettle (2020), "Locational Choices: Modeling Consumer Preferences for Proximity to Others in Reserved Seating Venues," *Journal of Marketing Research*, 57 (5), 878–99.

Cook, T. D., and D. T. Campbell (1979), *Quasi-Experimentation: Design & Analysis Issues for Field Settings*, Vol. 351. Boston: Houghton Mifflin.

Dekimpe, M. G., and D. M. Hanssens (2000), "Time-Series Models in Marketing: Past, Present and Future," *International Journal of Research in Marketing*, 17 (2-3), 183–93.

Garbinsky, E. N., J. J. Gladstone, H. Nikolova, and J. G. Olson (2020), "Love, Lies, and Money: Financial Infidelity in Romantic Relationships," *Journal of Consumer Research*, 47 (1), 1–24.

Gijsenberg, M. J., H. J. Van Heerde, and P. C. Verhoef (2015), "Losses Loom Longer Than Gains: Modeling the Impact of Service Crises on Perceived Service Quality over Time," *Journal of Marketing Research*, 52 (5), 642–56.

Goor, D., A. Keinan, and N. Ordabayeva (2021), "Status Pivoting," *Journal of Consumer Research*, 47 (6), 978–1002.

Goldfarb, A., C. Tucker, and Y. Wang (2022), "Conducting Research in Marketing with Quasi-Experiments," *Journal of Marketing*, 86 (3), 1–20.

Grewal, R., S. Gupta, and R. Hamilton (2020), "The Journal of Marketing Research Today: Spanning the Domains of Marketing Scholarship," *Journal of Marketing Research*, 57 (6), 985–98.

Kalra, A., X. Liu, and W. Zhang (2020), "The Zero Bias in Target Retirement Fund Choice," *Journal of Consumer Research*, 47 (4), 500–22.

Kappes, H. B., J. J. Gladstone, and H. E. Hershfield (2021), "Beliefs about Whether Spending Implies Wealth," *Journal of Consumer Research*, 48 (1), 1–21.

Kettle, K. L., R. Trudel, S. J. Blanchard, and G. Häubl (2016), "Repayment Concentration and Consumer Motivation to Get out of Debt," *Journal of Consumer Research*, 43 (3), 460–77.

Kim, T., K. Barasz, and L. K. John (2019), "Why Am I Seeing This Ad? The Effect of Ad Transparency on Ad Effectiveness," *Journal of Consumer Research*, 45 (5), 906–32.

Lynch John G. (1998), "Presidential Address Reviewing," in *NA—Advances in Consumer Research*, Vol. 25, ed. Joseph W. Alba and J. Wesley Hutchinson, Provo, UT: Association for Consumer Research, 1–6.

Meyvis, T., and S. M. Van Osselaer (2018), "Increasing the Power of Your Study by Increasing the Effect Size," *Journal of Consumer Research*, 44 (5), 1157–73.

Nelson, L., D. Simester, and K. Sudhir (2020), "Introduction to the Special Issue on Marketing Science and Field Experiments," *Marketing Science*, 39 (6), 1033–8.

Packard, G., and J. Berger (2021), "How Concrete Language Shapes Customer Satisfaction," *Journal of Consumer Research*, 47 (5), 787–806.

Papies, D., P. Ebbes, and H. J. van Heerde (2017), "Addressing Endogeneity in Marketing Models," in *Advanced Methods for Modeling Markets*, Cham: Springer, 581–627.

Pauwels, K. H. (2018), "Modeling Dynamic Relations among Marketing and Performance Metrics," *Foundations and Trends® in Marketing*, 11 (4), 215–301.

Reich, T., D. M. Kupor, and R. K. Smith (2018), "Made by Mistake: When Mistakes Increase Product Preference," *Journal of Consumer Research*, 44 (5), 1085–103.

Rossi, P. E. (2014), "Even the Rich Can Make Themselves Poor: A Critical Examination of IV Methods in Marketing Applications," *Marketing Science*, 33 (5), 655–72.

Schmitt, B. H., J, Cotte, M. Giesler, A. T. Stephen, and S. Wood (2022), "Relevance—Reloaded and Recoded," *Journal of Consumer Research*, 48 (5), 753–5.

Sharpe Wessling, K., J. Huber, and O. Netzer (2017), "MTurk Character Misrepresentation: Assessment and Solutions," *Journal of Consumer Research*, 44 (1), 211–30.

Simmons, J. P., L. D. Nelson, and U. Simonsohn (2011), "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant," *Psychological Science*, 22 (11), 1359–66.

Spiller, S. A., and L. Belogolova (2017), "On Consumer Beliefs about Quality and Taste," *Journal of Consumer Research*, 43 (6), 970–91.

Slotegraaf, R. J., and K. Pauwels (2008), "The Impact of Brand Equity and Innovation on the Long-Term Effectiveness of

Promotions," *Journal of Marketing Research*, 45 (3), 293–306.

Streicher, M. C., Z. Estes, and O. B. Büttner (2021), "Exploratory Shopping: Attention Affects in-Store Exploration and Unplanned Purchasing," *Journal of Consumer Research*, 48 (1), 51–76.

Stremersch, S., R. S. Winer, and N. Camacho (2021), "Faculty Research Incentives and Business School Health: A New Perspective from and for Marketing," *Journal of Marketing*, 85 (5), 1–21.

Sussman, A. B., A. Paley, and A. L. Alter (2021), "How and Why Our Eating Decisions Neglect Infrequently Consumed Foods," *Journal of Consumer Research*, 48 (2), 251–69.

Thomas, M. (2019), "Was Television Responsible for a New Generation of Smokers?," *Journal of Consumer Research*, 46 (4), 689–707.

Wang, X., S. Lu, X. I. Li, M. Khamitov, and N. Bendle (2021), "Audio Mining: The Role of Vocal Tone in Persuasion," *Journal of Consumer Research*, 48 (2), 189–211.