Contemporary psychological research can be dizzying in its complexity, and this complexity results in patterns of variation and covariation among the observations from a set of studies that requires careful treatment in meta-analysis. For example, individual studies in a given domain can vary considerably in terms of their dependent measures and moderators; examine multiple conditions that result from the experimental manipulation of those moderators and give rise to multiple dependent effects of interest (e.g., simple effects and interaction effects); employ a mix of study designs (e.g., unmoderated versus moderated, between-subjects versus within-subjects, univariate versus multivariate); and feature different contexts, treatment manipulations, and measurement scales. Further, individual papers feature multiple studies that, while different, are quite similar particularly in comparison to studies featured in other papers.

[to cite]:

[url]:
http://www.acrwebsite.org/volumes/1700002/la/v4_pdf/LA-04

[copyright notice]:
This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
Special Session Summary

Big Data and Consumer Behavior: Challenges & Solutions for Experimental and Observational Analyses

Blakeley B. McShane, Northwestern University, USA

Blakeley B. McShane, Northwestern University, USA
Ulfi Bockenholt, Northwestern University, USA

Paper #2: Multilevel Multivariate Meta-analysis with Application to Choice Overload
Blakeley B. McShane, Northwestern University, USA
Ulfi Bockenholt, Northwestern University, USA

Paper #3: Deconstructing Hedonic Experiences using Tensor Factorization
Karsten T. Hansen, University of California, San Diego, USA
Vishal Singh, New York University, USA

Paper #4: Online Reviews: Some Empirical Generalizations
Karsten T. Hansen, University of California, San Diego, USA
Vishal Singh, New York University, USA

SESSION OVERVIEW

This session will discuss the challenges posed by the growing size and complexity of consumer research data as well as suggest solutions with a focus on understanding these issues through carefully motivated methodology and application to case studies. Paper #1 and Paper #2 provide guidance on the opportunities provided by taking a big data view for experimental data while Paper #3 and Paper #4 provide similar guidance for observational data. All four papers demonstrate how the growing size and complexity of consumer research data can be exploited to yield new insights for theory development and theory testing.

More specifically in terms of experimental data, when the number of subjects per study, the number of studies per paper, and the number of papers are all increasing, the importance of synthesizing a single paper in order to distill the results in a digestible format is increasingly important. Paper #1 discusses novel single paper meta-analysis methodology that achieves exactly this and illustrates it via application to papers recently published in JCR and JMR. Paper #2 goes an important step beyond Paper #1 and introduces novel meta-analytic methodology that accommodates the increasing complexity of consumer research data that include multiple dependent measures, moderators, effects of interest, study designs, contexts, manipulations, and measurement scale. It applies this technique to data from fifty-seven studies of choice overload, an important area of consumer research inquiry, yielding both new findings as well as implications for future studies.

Focusing on observational unstructured data, Paper #3 proposes a framework for deconstructing hedonic experiences based on users’ subjective descriptions of experiences (in the form of text). Traditionally, in the ethnographic consumer research, scholars did this manually by simply reading the transcript. This paper shows how modern computational techniques empower researchers to automate this process in order to achieve massive scale and illustrates this with an application to the travel experiences of over 50,000 individuals. Paper #4 analyzes over 80 million amazon.com reviews. The focus of this paper is determining what makes reviews helpful to other consumers. Theory suggests that negative information is more diagnostic and helpful than positive information. The paper also explores the effect of expertise, review length and other characteristics on the perceived helpfulness of the review.

Single Paper Meta-analysis: Benefits for Study Summary, Theory-testing, and Replicability

Meta-analysis is a well-established statistical technique that synthesizes two or more studies of a common phenomenon. Because multiple studies provide more information about the common phenomenon than any single one of them, meta-analysis can offer a number of benefits. For example, insofar as the studies measure the common phenomenon with some degree of error, a meta-analysis, which pools the results from the studies via a weighted average, will yield an estimate that is on average more accurate than that of any individual study. In addition, the uncertainty in the meta-analytic estimate will typically be smaller than the uncertainty in the estimate of any individual study thereby inter alia increasing statistical power relative to individual studies and providing a means of resolution when individual studies yield so-called conflicting results. Further, meta-analysis allows for the investigation of differences among studies, for example by quantifying the impact of study-level covariates or the degree of between-study variation.

These benefits have been widely realized in behavioral research in traditional meta-analyses of studies that appear in multiple papers. However, they have only very seldom been realized in meta-analyses of studies that appear in a single paper. Indeed, a typical behavioral research paper features multiple studies of a common phenomenon that are analyzed solely in isolation. Because the studies are of a common phenomenon, this practice is inefficient and foregoes important benefits that can be obtained only by analyzing them jointly in a single paper meta-analysis (SPM).

In this paper, we introduce meta-analytic methodology that is specially tailored to the SPM of the set of studies that appear in a typical behavioral research paper. Our SPM methodology is user-friendly because it requires only basic summary information (e.g., means, standard deviations, and sample sizes); importantly, despite requiring only this basic summary information, the model underlying our SPM methodology is equivalent, by a principle known as statistical sufficiency, to that underlying the “gold standard” meta-analytic approach, namely an appropriately-specified hierarchical (or multilevel) model fit to the individual-level observations [Stewart and Tierney, 2002, Simmonds et al., 2005, Cooper and Patall, 2009, Haidich, 2010]. In addition, our SPM methodology is widely-applicable; indeed, a literature review reveals that it could have been used in 86% of the behavioral research papers published in the three most recent volumes of the Journal of Consumer Research (Volumes 40-42).

Our SPM methodology provides important benefits for study summary, theory-testing, and replicability that we illustrate via three case studies that include papers recently published in the Journal of Consumer Research and the Journal of Marketing Research and that,
as we further note in our discussion, are either not provided by or are provided only in part by alternative approaches. Specifically, it provides a graphical and quantitative summary of the studies. The graphical summary facilitates the communication and comparison of results within and across studies thus simplifying assessments of convergence while the quantitative summary provides a more precise estimate of each effect of interest as well as the uncertainty in this estimate. This increased precision is important for theory-testing because it allows for more powerful tests of posited effects. Further, these more powerful tests can deepen theory-testing by motivating new decompositions of the effects that investigate alternative explanations. Additionally, it provides an estimate of and accounts for between-study variation. This estimate of between-study variation can suggest unaccounted for moderators that have the potential to enrich theory while accounting for between-study variation improves calibration of Type I and Type II error. Finally, it provides sample size analyses for future studies and future sets of studies that account for the uncertainty associated with effect estimates as well as between-study variation thus enhancing replicability.

Our SPM methodology has two additional benefits. First, because it requires only basic summary information and this information is often reported in papers, it allows readers as well as authors to conduct an SPM and obtain the benefits for study summary, theory-testing, and replicability discussed above. Second, because the reporting of an SPM is extremely concise, it allows authors to, if they desire, include in the SPM studies they have that are related to those reported in the paper but which themselves were not reported in the paper; this allows authors to provide further evidence about the phenomenon of interest without taking up a great deal of journal space and thus should enhance replicability.

Because our SPM methodology is user-friendly, widely applicable, and provides these manifold benefits, we advocate that authors of typical behavioral research papers use it to supplement the single-study analyses that independently discuss the multiple studies in the body of their papers as well as the “qualitative meta-analysis” that verbally synthesizes the studies in the general discussion of their papers. When used as such, this requires only a minor modification of current practice.

To facilitate this, we provide an easy-to-use website that implements our SPM methodology. The website is available at http://www.singlepapermetaanalysis.com/

It includes a detailed tutorial that shows how to replicate the case studies presented in this paper and how to apply it to new papers.

**Multilevel Multivariate Meta-analysis with Application to Choice Overload**

Contemporary psychological research can be dizzying in its complexity, and this complexity results in patterns of variation and covariation among the observations from a set of studies that requires careful treatment in meta-analysis. For example, individual studies in a given domain can vary considerably in terms of their dependent measures and moderators; examine multiple conditions that result from the experimental manipulation of those moderators and give rise to multiple dependent effects of interest (e.g., simple effects and interaction effects); employ a mix of study designs (e.g., unmoderated versus moderated, between-subjects versus within-subjects, univariate versus multivariate); and feature different contexts, treatment manipulations, and measurement scales. Further, individual papers feature multiple studies that, while different, are quite similar particularly in comparison to studies featured in other papers.

However, the meta-analytic techniques typically employed in practice introduce a host of simplifications that fail to account for this complexity. For example, a common approach involves collapsing the observations from multiple conditions to form a single effect of interest; converting the effects to a common, standardized scale such as the Cohen’s d scale (i.e., the difference between two means divided by the pooled standard deviation of the individual-level observations); and modeling the standardized effects via a linear mixed model with one or two variance component parameters. If differences in dependent measures or moderators are accounted for, this is typically done only via fixed main effects. These simplifications can result in among other things the neglect of differences in dependent measures and moderators and miscalibrated inference.

In this paper, we introduce multilevel multivariate meta-analysis methodology that better accounts for the complexity of contemporary psychological research data. In particular, our methodology directly models the observations from a set of studies in a manner that accounts for the variation and covariation induced by the factors that observations differ in their dependent measures and moderators and are nested within, for example, papers, studies, groups of subjects, and study conditions. We also introduce two theoretically interesting and extremely parsimonious special cases of our methodology.

Our methodology generalizes prior multivariate meta-analysis models in three important respects, namely to simultaneously accommodate (i) not two dependent measures but an arbitrary number of dependent measures; (ii) not a single effect of interest (arising from, for example, a two condition study) but an arbitrary number of study conditions that result from the experimental manipulation of moderators and give rise to multiple dependent effects of interest; and (iii) not two levels but an arbitrary number of levels that account for the variation and covariation induced by the fact that the observations are nested (e.g., within papers, studies, groups of subjects, and study conditions).

These extensions are important because they are motivated by and respectful of key features of contemporary psychological research data—in particular data from papers and studies of the choice overload hypothesis—and thereby allow our methodology to better account for the variation and covariation induced by the fact that observations differ in their dependent measures and moderators and are nested. Choice overload has already been the subject of two prominent meta-analyses [Scheibe, et al., 2010, Chernev et al., 2015]. These meta-analyses employed different variations of the simplifications to the data and model discussed above and arrived at contradictory conclusions: Scheibe et al. [2010] “found a mean effect size of virtually zero” whereas Chernev et al. [2015] found that “the overall effect of assortment size on choice overload is significant.”

To resolve this difference, we apply our methodology to the set of fifty-seven studies from twenty-one papers originally examined by Chernev et al. [2015]. By avoiding the simplifications employed in the two prior meta-analyses, our methodology more fully accounts for the complexity of choice overload data and provides richer insight.

For example, it shows that choice overload varies substantially as a function of the six dependent measures (i.e., assortment choice, choice deferral, option selection, regret, satisfaction, and switching likelihood) and four moderators (i.e., choice set complexity, decision task difficulty, preference uncertainty, and decision goal) examined in the domain and that there are potentially interesting and theoretically important interactions among them. For example, choice overload occurs for the high level of the decision task difficulty moderat-
tor when option selection or satisfaction is the dependent measure; however, it is reversed for the low level of the decision task difficulty moderator when option selection is the dependent measure while it is nullified when satisfaction is the dependent measure. It also occurs when there are no moderators regardless of the dependent measure while it fails to occur when there are moderators–regardless of the moderator level–when choice deferral is the dependent measure. In sum, while choice overload reliably occurs for some dependent measure / moderator combinations and it is reliably reversed for others, for still others the evidence is quite mixed.

It also shows that the various dependent measures have differing levels of variation (heterogeneity). Specifically, choice deferral—a prominent and important dependent measure in the choice overload literature (e.g., it was the primary dependent measure featured in the original studies of choice overload by Iyengar and Lepper [2000])—is the dependent measure with by far the largest level of variation.

Finally, it shows levels up to and including the highest (i.e., the fifth, or paper, level) are necessary to capture the variation and covariation induced by the nesting structure. Failure to account for this results in miscalibrated inference.

Because our methodology is motivated by and respectful of key features of contemporary psychological research data—specifically in that it accommodates an arbitrary number of dependent measures, study conditions, and levels in the nesting structure—it is quite general and widely applicable and we expect it to yield rich insight in future applications. To facilitate its application, we provide an easy-to-use website that implements our methodology. The website is available at https://blakemcshane.shinyapps.io/mlmvmeta/

It includes a detailed tutorial that shows how to replicate the choice overload analysis presented in this paper.

**Deconstructing Hedonic Experiences using Tensor Factorization**

In this talk we propose to use tensor factorization to the problem of deconstructing users’ hedonic experiences. This involves (i) delineating the number of unique experiences, (ii) qualitatively describing each experience, and (iii) finding the valence of each experience (positive, negative or neutral). We also show how to link user characteristics to the likelihood of undergoing each experience.

While the proposed framework is general enough to include a wide variety of different data types, this talk will focus on text data. In other words, the source data is a corpus of documents where each document contains a given user’s experience. For example, a document could be a user writing about a vacation to Hawaii or staying at a Las Vegas resort.

Tensor factorization – or more precisely non-negative tensor factorization - is based on a simple idea, namely that the whole is the sum of the parts and the parts are non-negative. The non-negative aspect of the deconstruction turns out to lead to decompositions that have much cleaner interpretations than traditional data reduction techniques such as principal components analysis or factor analysis. This has proven to be an extremely useful approach in areas such as image reconstruction, computer vision, music analytics, and text/document classification.

In our proposed framework “the whole” is a (text) description of a user’s experiences in a certain context (e.g., travel, product usage, personal reflection, etc.). The “parts” are constituent experiences making up the total. For example, a person might write an online review for travel web site of a visit to a resort in Mexico, detailing the various elements of the overall experience. Alternatively, it could be a recent college graduate reflecting on the experience of spending four years in college.

How exactly should one delineate the individual component experiences and their associated valence? This can be done manually by simply reading the text and interpreting the underlying themes. This is the standard practice in the qualitative or ethnographic consumer research tradition. In the travel scenario, examples of themes might be product and service related characteristics like the quality of the food at the resort or the rooms, but also more subjective experiences such as reflecting on spending time with loved ones or walking in nature. Extracting underlying themes or “base”-experiences from an overall description is, however, not possible when a larger number of user experiences are involved.

In this work, we document the usefulness of the proposed framework by deconstructing the experiences of over 50,000 users, each writing about their travel experience to a certain destination. This involves (i) quantifying the total number of unique experiences, (ii) the qualitative nature/interpretation of each experience (i.e., what is the experience?), and (iii) the valence of each unique experience (does it tend to increase or decrease the valuation of the overall experience). We should how users naturally can be clustered based on their propensity to write about the constituent experiences and how user characteristics (such as demographics, location or attitudes) affect this propensity.

This analysis allows us to address questions such as: Are there more negative than positive hedonic experiences? Do users agree on what the positive and negative experiences are? For example, there might be consensus among users on what the negative experiences are, but wild disagreement about the nature of the positive experiences. Broader implications for consumer behavior research and usefulness for managerial decision-making are also discussed. We also discuss various extensions to the framework such as allowing for changes in experiences over time and how to incorporate experimental variation in the framework.

**Online Reviews: Some Empirical Generalizations**

Online consumer reviews (OCR) provide opportunities for consumers to share their experiences and opinions about products/services and are regarded among potential buyers as one of the most trusted sources of information. In this article, we analyze the information content in over 80 million reviews from amazon.com. The database is the universe of all reviews on the site (e.g., books, electronics, music, etc.) and spans over 15 years (1997-2014).

Building on previous literature, we construct a wide array of attributes to quantify the information content of reviews: length and timing of review, its syntactic and semantic features, star rating, as well as price and other characteristic of the product. Our analysis focuses on analyzing the ‘helpfulness’ of review based on votes by readers. In particular, we focus attention on the helpfulness ratio of a review, defined as the number of helpful votes (“I found this review helpful”) divided by the number of total votes.

Theory suggests that consumers perceive negative information as more diagnostic and persuasive. For example, Rozin & Royzman [2001] argues that giving greater weight to negative information and experience is a bias that both humans and animals share. However, we find the helpfulness of a review to increase monotonically with the “star” rating of the review. With minor exceptions (e.g. subcategory ‘electronic warranties’), this pattern is consistent across all product categories at amazon.com.

Interestingly, we find that this effect is strongly moderated by the overall quality of the product – the effect of star rating is much stronger for high-quality products than for low quality products. This
means that a one-star review of a product that in general is highly rated is considered extremely unhelpful. On the other hand, a five-star review of an overall poorly rated product doesn’t impact the perceived helpfulness of the review very much.

A more detailed analysis highlights important asymmetries when users deviate from the “consensus” review. Previous researchers have argued that deviation from the consensus rating is considered unhelpful for both positive and negative deviations – with a more or less symmetric effect. For example, Danescu-Niculescu-Mizil, Cristian, et al. [2009] finds symmetric effects of deviating for the “norm” review in amazon.com book reviews. However, their analysis does not control for underlying product quality. When controlling for product quality, we find strong asymmetries when comparing helpfulness effects of positive and negative deviations. This is contrary to Danescu-Niculescu-Mizil, Cristian, et al. [2009].

In addition, we find that long reviews are in general considered more helpful than short reviews. Furthermore, we use information on the top 100K reviewers at amazon.com that provides us information on reviewer name (used for coding gender), location (city/state), and their reviews for all other products at amazon.com. We find systematic differences between heavy vs. occasional reviewers (what one might call “experts” vs. “novices”) and between males vs. females in their language use, ratings, and length of reviews.

Implications for online retailers and service providers are discussed. More broadly, we discuss the use of large commercial databases for research on consumer behavior. We argue that many important questions and theories in consumer behavior and psychology can be tested and analyzed using non-experimental methods and data. This provides an important addition to existing methodologies in consumer behavior such as surveys and, of course, experimental methods.

REFERENCES