

DELVING DEEP

Techniques We Wished We Had
Known as Emerging Scholars

Edited by Paula O’Kane,
John R. Busenbark, Aaron F. McKenny
and Sotirios Paroutis

RESEARCH METHODOLOGY IN
STRATEGY AND MANAGEMENT

VOLUME 15

DELVING DEEP

RESEARCH METHODOLOGY IN STRATEGY AND MANAGEMENT

Series Editors: Paula O’Kane, Aaron F. McKenny,
Sotirios Paroutis, and John R. Busenbark

PREVIOUSLY PUBLISHED VOLUMES

- Volume 1: Research Methodology in Strategy and Management, 2004
- Volume 2: Research Methodology in Strategy and Management, 2005
- Volume 3: Research Methodology in Strategy and Management, 2006
- Volume 4: Research Methodology in Strategy and Management, 2007
- Volume 5: Research Methodology in Strategy and Management, 2008
- Volume 6: Building Methodological Bridge, 2011
- Volume 7: West Meets East: Toward Methodological Exchange, 2012
- Volume 8: West Meets East: Building Theoretical Bridges, 2012
- Volume 9: Social Entrepreneurship and Research Methods, 2014
- Volume 10: Advancing Research Methodology in the African Context: Techniques, Methods, and Designs, 2014
- Volume 11: Standing on the Shoulders of Giants, 2019
- Volume 12: Advancing Methodological Thought and Practice, 2020
- Volume 13: Research in Times of Crisis: Research Methods in the Time of COVID-19, 2021
- Volume 14: Methods to Improve Our Field, 2023

RESEARCH METHODOLOGY IN STRATEGY AND
MANAGEMENT, VOLUME 15

DELVING DEEP: TECHNIQUES WE WISHED WE HAD KNOWN AS EMERGING SCHOLARS

EDITED BY

PAULA O'KANE

University of Otago, New Zealand

JOHN R. BUSENBARK

University of Notre Dame, USA

AARON F. MCKENNY

Indiana University Bloomington, USA

AND

SOTIRIOS PAROUTIS

University of Warwick, UK



United Kingdom – North America – Japan
India – Malaysia – China

Emerald Publishing Limited
Emerald Publishing, Floor 5, Northspring, 21-23 Wellington Street, Leeds LS1 4DL

First edition 2025

Editorial matter and selection © 2025 Paula O’Kane, John R. Busenbark, Aaron F. McKenny and Sotirios Paroutis.

Individual chapters © 2025 The authors.

Published under exclusive licence by Emerald Publishing Limited.

Reprints and permissions service

Contact: www.copyright.com

No part of this book may be reproduced, stored in a retrieval system, transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without either the prior written permission of the publisher or a licence permitting restricted copying issued in the UK by The Copyright Licensing Agency and in the USA by The Copyright Clearance Center. Any opinions expressed in the chapters are those of the authors. Whilst Emerald makes every effort to ensure the quality and accuracy of its content, Emerald makes no representation implied or otherwise, as to the chapters’ suitability and application and disclaims any warranties, express or implied, to their use.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN: 978-1-83797-027-8 (Print)

ISBN: 978-1-83797-026-1 (Online)

ISBN: 978-1-83797-028-5 (Epub)

ISSN: 1479-8387 (Series)

Erratum: It has come to the attention of the publisher that the chapter Shepherd, D.A., O’Kane, P., and Paroutis, S. (2024), “A Qualitative Researcher’s Journey: An Interview With Professor Dean A. Shepherd, Paula O’Kane, and Sotirios Paroutis”, O’Kane, P., Busenbark, J.R., McKenny, A.F. and Paroutis, S. (Ed.) *Delving Deep (Research Methodology in Strategy and Management, Vol. 15)*, Emerald Publishing Limited, Leeds, pp. 85–94. <https://doi.org/10.1108/S1479-838720240000015005>, did not include Dean A. Shepherd as an author due to an editorial error. Our guidelines make it clear that anyone who made a significant contribution to the work must be included as an author. Dean A. Shepherd has been added as an author of this chapter in both the published version and metadata, and the correct author information is as follows: Dean A. Shepherd, Paula O’Kane, Sotirios Paroutis. The publisher sincerely apologizes for this mistake.



INVESTOR IN PEOPLE

CONTENTS

About the Editors vii

About the Contributors ix

**Building an Empirical Body of Evidence: Developing Rapport
With Reviewers and Overcoming Skepticism in Strategic
Management Research** 1
Timothy J. Quigley

**Philosophy and Management Research: A Crucial yet Neglected
Connection** 23
Eric W. K. Tsang

**The SMART Tool: Encouraging Standardized and Replicable
Research in Management to Enhance Credibility** 43
*Andrew B. Blake, Oleg V. Petrenko, Timothy J. Quigley, Aaron
D. Hill and Amrit Panda*

**How to Cross the Uncanny Valley: Developing Management
Laboratory Studies Using Virtual Reality** 57
Timothy D. Hubbard and Michael Villano

**A Qualitative Researcher's Journey: An Interview With Professor
Dean A. Shepherd, Paula O'Kane, and Sotirios Paroutis** 85
Dean A. Shepherd, Paula O'Kane, Sotirios Paroutis.

**Executive Personality Assessment With Large Language Models:
Updating an Existing Tool and Advancing Similar Measures in
Strategy and Management Research** 95
*Joseph S. Harrison, Steven Boivie, Timothy D. Hubbard and Oleg
V. Petrenko*

This page intentionally left blank

ABOUT THE EDITORS

Paula O’Kane is a Senior Lecturer in Human Resource Management at the University of Otago, Dunedin (New Zealand). She has expertise in interpretative qualitative research across a range of areas including performance management, social media, and the future of work. She has extensive expertise in Computer Aided Qualitative Data Analysis and has published in *Research Methodology in Strategy and Management*, as well as *Organizational Research Methods*, *Human Resource Management Journal*, *Studies in Higher Education*, among others.

John R. Busenbark is the Mary Jo and Richard M. Kovacevich Associate Professor of Management and Organization in the Mendoza College of Business at the University of Notre Dame, Indiana, USA.

Aaron F. McKenny is an Assistant Professor of Management and Entrepreneurship at the Kelley School of Business, Indiana University, and was the recipient of the 2020 *Sage Publications/RMD Lawrence R. James Early Career Award* from the Academy of Management Research Methods Division. His methods research emphasizes the dissemination and advancement of content analysis in organizational research and has been published in journals such as *Organizational Research Methods*, *Journal of Applied Psychology*, *Journal of Management*, *Annual Review of Organizational Psychology and Organizational Behavior*, *Research Methods in Strategy and Management*, and *Family Business Review*. He is a coeditor of the *Research Methodology in Strategy and Management* series and sits on the review boards of *Academy of Management Review*, *Organizational Research Methods*, *Journal of Management*, *Journal of Business Venturing*, *Entrepreneurship Theory and Practice*, and *Family Business Review*.

Sotirios Paroutis is a Professor of Strategic Management at the Warwick Business School (United Kingdom). He primarily uses qualitative methods to study the practices and processes organizational actors employ when dealing with strategic tensions. His work has been published in *Research Methodology in Strategy and Management*, *Strategic Management Journal*, *Organization Studies*, and *Long Range Planning*, among others.

This page intentionally left blank

ABOUT THE CONTRIBUTORS

Andrew B. Blake is an Assistant Professor of Management in the Rawls College of Business at Texas Tech University. He received his PhD from the University of Arkansas. Andrew's research broadly focuses on the intersection of social psychology and the micro-foundations of strategy, emphasizing attention on the impact that CEOs and entrepreneurs have on organizational outcomes.

Steven Boivie is the Carroll & Dorothy Conn Chair in New Ventures Leadership at the Mays Business School at Texas A&M University. He received his PhD in Strategic Management from the University of Texas at Austin, his Master's degree from Brigham Young University, and his Bachelor's degree from Utah State University. He is primarily interested in how behavioral and social forces affect human actors at the top of the organization, and he conducts research in the areas of corporate governance, top executives, and directors. His research has been published in the *Academy of Management Journal*, *Strategic Management Journal*, *Organization Science*, *Journal of Applied Psychology*, *Academy of Management Annals*, and *Journal of Management*. Steve's research has also been mentioned in a number of press outlets including the *Wall Street Journal*, *Harvard Business Review* online, [Forbes.com](https://www.forbes.com), *IR Magazine*, *Bloomberg*, *The Economist*, and more.

Joseph S. Harrison is an Associate Professor of Strategy in the Haslam College of Business at the University of Tennessee. He received his PhD in Strategic Management from Texas A&M University. His research focuses on social and behavioral aspects of strategic leaders and using content analytic techniques to understand organizational phenomena. Through his research, he has codeveloped multiple machine learning applications to extract meaning from unstructured text and provide insights into executive personality, corporate governance, risk, and performance. His work has been published in *Academy of Management Journal*, *Academy of Management Review*, *Journal of Management*, *Organization Science*, and *Strategic Management Journal*, among others. It has also been highlighted in press outlets like *Harvard Business Review* online, [Fortune.com](https://www.fortune.com), and the TEK2Day podcast.

Aaron D. Hill is an Associate Professor in the Management Department of the Warrington College of Business at the University of Florida. He received his PhD at Oklahoma State University. His research focuses on strategic leadership and governance, examining the implications of executives for organizational outcomes.

Timothy D. Hubbard is an Assistant Professor of Management in the Mendoza College of Business, University of Notre Dame. He received his PhD from the University of Georgia. His research focuses on behavioral strategy – the cognitive and social factors that influence strategic decisions and their outcomes. He serves as a Director of the University of Notre Dame’s Virtual Reality Laboratory. His research appears in the *Academy of Management Journal*, the *Strategic Management Journal*, the *Journal of Management*, the *Journal of Management Studies*, and the *Academy of Management Discoveries*, among others. He teaches strategic management and organizational theory.

Amrit Panda is a PhD student in the Management Department in the Rawls College of Business at Texas Tech University. He received his Master’s in Human Resource Management from the Tata Institute of Social Sciences, India. His research interests lie broadly around upper echelons and competitive behavior of firms. Specifically, he is interested in exploring how CEOs and top management teams shape competitive strategy and its subsequent effect on firm performance and other strategic outcomes.

Oleg V. Petrenko is an Assistant Professor in the Strategy, Entrepreneurship, Venture, and Innovation Department at the Sam M. Walton College of Business at the University of Arkansas. He received his PhD from Oklahoma State University. His research broadly focuses on the micro-foundations of strategy in large corporations and new ventures. Specifically, he examines the psychology of executives and entrepreneurs and how it impacts strategic decision-making and firm performance. Oleg’s research focuses on creating knowledge that will help the leaders of tomorrow be successful at leading their organizations and taking on new challenges.

Timothy J. Quigley is the Georgia Athletic Associate Professor of Management in the Terry College of Business at the University of Georgia. He received his PhD from Pennsylvania State University and is broadly interested in CEOs, CEO succession, and how CEOs impact organization outcomes. Tim is currently an Associate Editor at *Strategic Management Journal*.

Dean A. Shepherd, Bachelor of Applied Science (RMIT), MBA and PhD (Bond University, Australia), is the Ray and Milann Siegfried Professor of Entrepreneurship at the Mendoza College of Business, Notre Dame University. He has held positions at the Kelley School of Business, Indiana University, the Leeds School of Business, University of Colorado, and the Lally School of Management, Rensselaer Polytechnic Institute. His research and teaching are in the field of entrepreneurship; he investigates the decision-making involved in leveraging cognitive and other resources to act on opportunities, responding with resilience to adversity, learning from experimentation (including failure), and the dark side of entrepreneurship. He has authored over 20 books and 180 publications in top management and entrepreneurship journals with over 80,000 Google sites. Dean is a Fellow of the Academy of Management. From the Entrepreneurs Division of the Academy of Management, Dean has been awarded the “Dedication to

Entrepreneurship Award”, the “Mentor Award”, and the “Foundational Paper Award” (twice). He is the past Editor-in-Chief of the *Journal Business Venturing*. He is on the editorial boards of the *Academy of Management Journal*, the *Academy of Management Review*, *Entrepreneurship Theory and Practice*, *Strategic Entrepreneurship Journal*, and the *Journal of Management Studies*.

Eric W. K. Tsang is the Dallas World Salute Distinguished Professor at the Jindal School of Management, University of Texas at Dallas, and is a Fellow of the Academy of International Business. He received his PhD from the University of Cambridge. His main research interests include organizational learning, strategic alliances, corporate social responsibility, entrepreneurship, and philosophical analysis of methodological issues. He has published widely in leading business journals, such as *Academy of Management Journal*, *Academy of Management Review*, *Journal of International Business Studies*, *Marketing Science*, *MIS Quarterly*, and *Strategic Management Journal*, and is the author of the books *The Philosophy of Management Research* and *Explaining Management Phenomena: A Philosophical Treatise*. His publications have been well cited by fellow researchers. For example, his article “Social Capital, Networks, and Knowledge Transfer” (with Andrew Inkpen) is the winner of the 2015 Academy of Management Review Decade Award.

Michael Villano is a Research Assistant Professor of Psychology in the College of Arts and Letters at the University of Notre Dame.

This page intentionally left blank

BUILDING AN EMPIRICAL BODY OF EVIDENCE: DEVELOPING RAPPORT WITH REVIEWERS AND OVERCOMING SKEPTICISM IN STRATEGIC MANAGEMENT RESEARCH

Timothy J. Quigley

University of Georgia, USA

ABSTRACT

As the field of strategic management has evolved, expectations for the empirical evidence presented in manuscripts have risen substantially. Rather than a single model testing a hypothesis with a p-value below a standard threshold being sufficient, reviewers, editors, and eventual readers now demand additional evidence including multiple tests, advanced statistical models, alternative specifications, interpretation of practical rather than just statistical significance, and more. Reviewers appear to be increasingly skeptical and often raise a seemingly endless number of questions. In this chapter, I outline the idea of a body of evidence and suggest ways authors can build their evidence by anticipating reviewer questions and structuring manuscripts accordingly. Doing so allows authors to overcome skepticism by building positive rapport and trust with reviewers and the ultimate readers of their work. I conclude by discussing the review process where I offer suggestions about how reviewers and editors might adapt to this changing landscape. I specifically argue that all studies are flawed. Rather than asking for a single study to do more to address small inconsistencies or puzzling results, I suggest gatekeepers in the review process should consider the possibility that publishing and allowing research conversations to flourish might result in greater knowledge generation over time.

Delving Deep

Research Methodology in Strategy and Management, Volume 15, 1–22

Copyright © 2025 Timothy J. Quigley

Published under exclusive licence by Emerald Publishing Limited

ISSN: 1479-8387/doi:[10.1108/S1479-838720240000015001](https://doi.org/10.1108/S1479-838720240000015001)

Keywords: Building a body of evidence; empirical research; trust; robustness; authoring; reviewing

INTRODUCTION

When you submit an empirical manuscript to a journal, you enter an asynchronous dialogue with a review team composed of an action editor and a set of anonymous reviewers. In doing so, you have two primary tasks. First, you must convince the review team that your ideas have merit. There is some debate about what this entails. The oft cited Davis treatise “That’s Interesting” argues that papers should be interesting, novel, or counterintuitive and that they should create a “movement of the mind” by confronting taken-for-granted assumptions (Davis, 1971). Others stress the importance of addressing problems that are relevant to society while noting that an extreme focus on novelty might predispose our research toward chance findings and encourage troubling behavior like p-hacking and HARKING (hypothesizing after results are known) (Bettis et al., 2016; Tihanyi, 2020). As Bettis et al. note, it seems odd that the world would only be arranged so “that all phenomena of research importance are counterintuitive (p. 260).” Rather, they argue, we should “return the word interesting to its standard English language meaning of something that you want to learn more about (p. 260).” To this end, Huff (1999, p. 3) conveys the importance of positioning research within one or more ongoing scholarly “conversations,” while Colquitt and George (2011) proffer that research should address big problems, change conversations, encourage new discussions, and provide insights for practice.

I’m of the opinion that generating interest is part of the creative process of research. It’s hard to fully define, up front, all the ways something can be interesting, but we all know it when we see it. In a classic case of equifinality, there exist numerous creative ways to generate interest from readers (and, before that, reviewers) but, without it, your chances of success fade quickly as reviewers find conceptual cause to recommend rejection before even considering your empirical efforts. Nevertheless, this first imperative is conceptual in nature, focused mostly on the topic, framing, and motivation for a study.

Should you succeed in the first task of generating interest – a topic which receives ongoing attention from editors and seminal scholars alike – consideration moves to the second task. Here, your aim is to convince reviewers that your empirical results are robust, that they support your conceptual arguments and hypotheses, that the related inferences reasonably reflect what can be concluded from your data and analyses, and that these likely reflect the state of things in the population of interest. Frank (2000), the progenitor of the increasingly ubiquitous impact threshold of a confounding variable (ITCV) test, often refers to this as engaging in conversation with skeptics of empirical work. Frank surmises that it is incumbent on authors to assume readers are skeptical of the study’s methodological procedures and to conduct (and explain) sufficient empirical techniques to assuage that trepidation.

The purpose of this chapter is to further address this second task, which has received far less attention than its initial counterpart. In doing so, I will offer the concept of “*building a body of evidence*” as your objective in this dialogue with methodological skeptics. No research project is perfect. No set of conclusions can be supported uniformly by every empirical test. Rather, your goal is to present a convincing body of evidence to persuade readers that your results are robust, and that you questioned your empirical inferences by subjecting supportive findings to the same scrutiny as you would if the results didn’t work out. Doing this conveys your findings are trustworthy and likely reflective of something happening in the world. Here, the creative process continues, and the savvy scholar can use descriptive statistics, multiple tests, alternative specifications, and other means to craft a robust story that builds further interest in a paper through the strength and comprehensiveness of the empirical evidence. Done well, it creates a dialogue with reviewers and, once published, the readers as well.

Those who have been through the review process might quibble with the characterization of a paper submission as the “beginning of a dialogue.” As the “gatekeepers of science” (Crane, 1967, p. 195), reviewers clearly have an active voice in the process. They are chosen from among a pool of experts in the field and, as Frank (2000) points out, they generally assume the role of a skeptic who must be convinced through a collection of persuasive evidence that the arguments merit publication and empirical findings are legitimate. Reviewers exercise a certain privilege where they get to tell authors what they like and what they don’t like without ever having to directly face the author(s) they are assessing. Unless the manuscript is given a revise and resubmit, authors don’t have much of a chance to engage in a “dialogue” by responding to criticism. If a paper is accepted, authors become known while reviewers enjoy anonymity. The potential problems within this process are well documented (Miller, 2006; Starbuck, 2003).

At first glance, I can see doubters of my view instinctively making the case that there’s no “dialogue” in this process, but hear me out. My argument is that as reviewers digest a paper, they are stimulated by the writing to ask a series of seemingly rhetorical questions. While they turn the pages of your manuscript, reviewers repeatedly ask “but what about. . .,” “why didn’t you. . .,” “but did you consider. . .,” or “what happens if . . .” I say “rhetorical” because, while reviewers (and later, readers) would like an answer, the authors are not physically present to offer one. Knowing that reviewers engage in such self-dialogue, an astute scholar can try to anticipate these questions and use them to assemble a set of logical arguments, alternative specifications, and supplemental analyses that effectively addresses those rhetorical questions *as the reviewer raises them silently in their own mind*. Done successfully, an author can have an asynchronous dialogue across time and space with reviewers. Yet, doing so requires that authors have some foresight created largely through a willingness to solicit critical feedback and be critical of one’s own work.

To be clear, however, this chapter is not about simply piling up multiple statistical tests to overwhelm reviewers. They will see through that and doing so, at best, proves superfluous and cumbersome for all parties involved. Further, this is not about extensive use of jargon, technical terms, or seemingly advanced methodological wizardry to make a point. Of course, many papers call for

complex statistical methods and they should be used (and clearly explained) when needed. However, their use can, at times, create more problems than they solve (Certo et al., 2016; Semadeni et al., 2014). Rather, building a body of evidence is about telling a story through descriptive statistics, formal empirical tests, alternative specifications, supplemental tests, and examples that flow logically from your conceptual arguments.

These ideas were primarily developed through my own journey trying to publish my research, sparring with reviewers (who are, more often than not, correct), through work as an associate editor, and by collaborating with and teaching applied econometrics to doctoral students. In developing these ideas, some colleagues insisted much of it should already be well understood by most strategic management scholars. Perhaps that's true. On the other hand, considering how many scholars misunderstand or have difficulty explaining what a p -value is (Aschwandten, 2015) or fail to grasp the differences between fixed- and random-effects models (Certo et al., 2017), among many other seemingly innocuous and ostensibly well-grasped empirical procedures, I suspect there is a considerable gap between what should be commonly understood and what actually is. In my experience, I find many scholars, both new and seasoned, fail to take advantage of many of the ideas outlined here when submitting their work. Too many accept statistical significance as "proof" of support for their ideas and hastily push toward submission never considering if that result was luck or the work of statistical anomalies traced to outliers or coding errors. Thus, while the primary audience of this chapter is likely to be those who are relatively new to the field of strategic management (doctoral students and newer faculty), it is my hope that many of the ideas will resonate with and be useful to more seasoned scholars as well.

Moreover, and perhaps even more crucially for our field, I hope these ideas can inform all of us as reviewers. No study is perfect. Rather, every study has lingering weaknesses or slight inconsistencies. Most papers conflict, to some degree, with prior theory or empirical findings. Often there's a result that doesn't quite line up with the rest of the empirical evidence. Research should be judged on the collective merit of the arguments and evidence offered, with some level of acceptance that flaws and inconsistencies will always exist. A body of evidence can paint a broad picture about what might be. Within that body of evidence, inconsistencies or weaknesses create avenues for continued dialogue (Huff, 1999), new research, and eventual breakthroughs (Hollenbeck & Wright, 2017). From this perspective, then, a secondary audience of this chapter might be the more seasoned scholars serving as reviewers or editors and in the privileged position of determining what gets published. Here my hope is that these ideas serve as a counterbalance against our natural tendency to nitpick every small flaw while failing to see how a given study builds on past conversations while opening the door to new ones yet to come.

A BODY OF EVIDENCE

What is a "body of evidence"? Borrowing from criminal justice, when pursuing a criminal case, a prosecutor must present evidence that proves, beyond a reasonable doubt, that a defendant is guilty of a crime. In pursuing this work, the

astute prosecutor must identify and lay to rest any alternative and reasonably plausible theory about the events that occurred. This can be accomplished by presenting substantial physical evidence establishing critical facts of the case, the use of witness testimony, and the opinions of experts that might speak to the plausibility of alternative explanations or the chances for error when using scientific techniques to analyze evidence.

As an example, imagine a case of theft where a suspect's fingerprints were found at the scene of a crime where valuables were stolen. Simply demonstrating the existence of the defendant's fingerprints is probably not sufficient to convict. One must show that items were actually stolen and not lost, that it was likely the defendant who stole them, and that it is unlikely the crime was committed by someone else. An astute defense attorney, knowing many other fingerprints were present at the scene, will raise this issue to cast doubt. To counteract this, a skilled prosecutor might anticipate the defense will raise this issue. Rather than wait, the prosecutor might proactively demonstrate they belong to individuals known to be routinely at the scene while also documenting that each has an alibi that reasonably eliminates them from consideration. In doing so, the prosecutor is showing that they questioned their own case.

The parallels aren't perfect. In court, testimony happens in real time and witnesses from each side can be immediately cross examined. In attempting to publish a manuscript, the process is asynchronous, and authors get to respond formally only if a revise and resubmit is offered. Still, an author is akin to the prosecutor presenting a case to the jury (action editor) while a reviewer is the defense attorney for science – a guardian for truth aiming to keep bad science from entering the hallowed pages of our respected journals. A reviewer's job, then, is to identify pockets of reasonable doubt in a manuscript. An author, in turn, must present a body of compelling logical arguments and empirical evidence that provides ample support for the claims made. If an author can effectively foresee the lines of inquiry that are likely to come from reviewers (like the astute prosecutor above), authors can also participate in the cross examination by proactively answering reviewer questions as they turn the pages of a manuscript.

It is not entirely clear to me what the standard of evidence should be. Depending on the maturity of the research topic under consideration and how the findings might affect the health, welfare, or safety of the people, businesses, society, or other relevant entity that is the focus of the study, one could easily argue for a standard of "more likely than not" a hypothesis is true, "beyond a reasonable doubt," or some other standard. Of course, setting that standard is under the purview of reviewers and editors. But, what seems clear to me is that one cannot simply offer a hypothesis, report the result of a single regression model with stars in the appropriate place documenting a p -value of less than 0.05, and then declare victory. More is needed, and this can be accomplished by building a body of evidence.

Below I will discuss some of the elements that can encompass a body of evidence. As summarized in [Table 1](#), these include descriptions of the sample and measures, descriptive statistics, formal hypothesis tests, alternative specifications, supplemental analyses, and examples. Notably, papers do not need to include all of these nor is this an exhaustive list of possible forms of evidence. A body of evidence should

Table 1. Components of a Body of Evidence.

Category	Examples	Purpose
Sample and measures	• Complete description of sample including accounting for lost observations • Clear description of measures • Support for construct validity even for existing measures	• Transparency, future replication, clear understanding that sample reflects population of interest • Ensuring measures meaningfully capture constructs
Descriptive statistics	• Correlations of key relationships • Means with group comparisons • Distributions	• Generate a sense that the data are behaving as expected without fancy, multivariate analyses
Formal tests and effect sizes	• Traditional regression results • Related beta coefficients • Confidence intervals • Marginal effects • Practical or economic impact	• Provide formal statistical support for claimed relationships and demonstration of magnitude
Alternative specifications	• Alternative measures • Varied sampling techniques • Alternative estimators	• Provide robustness to demonstrate relationships were not found due to luck or that the relationship is only found in a very particular sample with certain measures that was picked “because it worked”
Alternative explanations and supplemental tests	• Alternative logical tests that must be true (or not) if your results are true • Tests of the limits of your theoretical arguments	• Test boundary conditions and attempt to rule out alternative arguments and causal mechanisms • Explore novel findings that might generate new conversations
Examples	• Anecdotes • Case studies • Quotes	• Capture the phenomenon of study in the real world • Demonstrate relationship exists and is recognizable
Other creative approaches	• Limited only by creativity of author(s)	• Further tell the story of the paper

be assembled to tell a story, build from existing research conversations (Huff, 1999), flow logically from the theoretical arguments and formal results, and can include an array of creative approaches not covered here. Before getting into the various forms of evidence, I address why formal tests, alone, are often insufficient evidence to fully support theoretical arguments. I then discuss the various forms of evidence. To conclude, I discuss the value of having this dialogue with reviewers and offer suggestions for how an author might gain the foresight needed to anticipate review questions before submitting a paper for review. I also briefly address how reviewers might think about this process as well.

INSUFFICIENCY OF FORMAL TESTS AND THE FOLLY OF OVERINTERPRETING *P*-VALUES

Recall that, for a null-hypothesis test, a *p*-value is simply the odds of finding a relationship as large as was found (or larger) in a given sample when there is no

corresponding relationship in the underlying population (Bettis et al., 2016; Kennedy, 2008). Many misinterpret p -values as one or all of the following: the odds a finding is wrong; or one minus the p -value as the odds the finding is true; or an indication of the relative strength of a finding ($p = 0.01$ is “stronger” than $p = 0.05$). These are all commonly applied and wildly incorrect. To further emphasize this, run the Stata code available in Appendix 1. The code generates 100 random “ y ” variables and 100 random “ x ” variables and then generates 100 regressions where a single x ($x_1, x_2, x_3, \dots, x_{100}$) is used to predict the corresponding y ($y_1, y_2, y_3, \dots, y_{100}$). Imagine each x - y pair is a random draw from some large population of interest. Each randomly drawn sample includes 100 observations. In the final step of the Stata code, the betas and p -values for x -variable in the 100 regressions are captured and tallied. The final line outputs a value tabulating the number of times out of the 100 models when the p -value for the x beta was less than or equal to 0.05.

Given the data are randomly generated, the “true” relationship in the underlying population is zero. Yet, by definition, the expected outcome here is five. That is, we’d expect five cases out of 100 where the p -value is less than or equal to 0.05 even if our data were purely random. If you run this code repeatedly (say 100 times), the average number of cases with p -value less than or equal to five will be approximately 5 but individual runs will return values that range well above and below 5. Why does this happen? When randomly generating a sample from the population, sometimes through dumb luck we get a sample that shows a relationship that doesn’t exist in the population – a Type 1 error. If our standard is a p -value that is less than or equal to 0.05, then, even with completely random data, we will see statistically significant coefficients approximately five percent of the time. With a critical value of 0.10, the expectation would be 10. In the version I ran as I am typing this, there were 8 with p less than or equal to 0.05 and 12 less than or equal to 0.10 (see Table 2 for a list of these cases). The two lowest p -values were 0.000 and 0.005 corresponding to betas of -0.389 and $+0.297$, respectively. If the two lowest p -values provide “significant”

Table 2. “Significant” ($p \leq 0.10$) Betas and p -Values in 100 Regressions With Random Data.

Case	p -Value	Beta	Confidence Interval	
1.	0.000	-0.389	-0.583	-0.195
2.	0.005	0.297	0.090	0.504
3.	0.014	-0.238	-0.426	-0.050
4.	0.020	-0.237	-0.436	-0.039
5.	0.028	0.205	0.023	0.386
6.	0.031	0.218	0.020	0.416
7.	0.032	-0.193	-0.368	-0.017
8.	0.040	0.184	0.008	0.360
9.	0.068	0.173	-0.013	0.359
10.	0.072	0.162	-0.015	0.340
11.	0.073	0.196	-0.019	0.412
12.	0.078	-0.192	-0.406	0.022

estimates for the population that are similarly sized in magnitude but in the opposite direction, it should be clear why the p -value cannot speak to the “strength” of a result.

Understanding how this simple example applies to research highlights a critical reason why we need to build a body of evidence in our papers. With each study we complete, we take a scoop of randomly selected data (we hope it is randomly selected, and for sake of argument here, let’s assume it is) from an underlying population and compute statistical tests from that sample. Using the results of these tests on the sample, we estimate or infer – why it’s called inferential statistics – what the value of that parameter is in the population. No empirical estimator is perfect. Each has numerous assumptions, and we almost always violate at least some of them. Even if we didn’t, it’s important to acknowledge that our fancy statistical tools only provide estimates of what is occurring in the population along with a confidence interval of that estimate. But this is not 100% certainty. That is, there’s no guarantee the real value lies within the specified range because estimates will offer false positives a nontrivial portion of the time.

As illustrated with my sampling distribution example via the code in Appendix 1, if the 100 pairs of x and y represented 100 scholars simultaneously but independently pursuing similar research questions with a randomly chosen sample from the same underlying population, we would expect about 5 of them to have supported results with a p -value less than or equal to 0.05 even if the relationship didn’t exist. If one “significant” test was enough, we would have to concede that our journals are disproportionately filled with the lucky few who found the spurious or random results supportive of hypothesized relationships that do not exist in the population (Goldfarb & King, 2016). We can begin to overcome this problem by building a body of evidence. Authors must accept that reviewers are rightly asking, at least in part, “was this paper one of the 5-in-100 that found a result from luck or is this result real?” Of course, in a cruel bit of irony regarding what it takes to be interesting, with increasingly counterintuitive or novel hypotheses (Davis, 1971), we would expect a reviewer to be increasingly skeptical.

Sadly, p -values have been given “almost mythical properties far removed from the mundane probabilistic definition” (Bettis et al., 2016, p. 259) that underlies their correct meaning. A p -value of 0.049 is taken to mean that a corresponding hypothesis is “true” yet one that is 0.051 means it is most certainly “false.” Of course, in most studies of modest sample sizes, these two p -values are certainly not practically or even statistically different from each other. Yet, in many cases, one gains the scientist a publication while the other goes into the file drawer. While beyond the scope of this chapter, alternative approaches do exist. Replications allow us to assess the validity and boundary conditions of prior work, and Bayesian analyses allow us to calculate the degree of belief in an outcome based on our prior knowledge of the topic. Further, the random paper that gets published when the results were spurious should not doom science to the fate desired by skeptics. As Bettis et al. (2016) note, “one study proves little or nothing... Instead, it establishes initial confirming evidence” (p. 260). In assessing the true nature of the world, we should assess the collective body of scientific evidence.