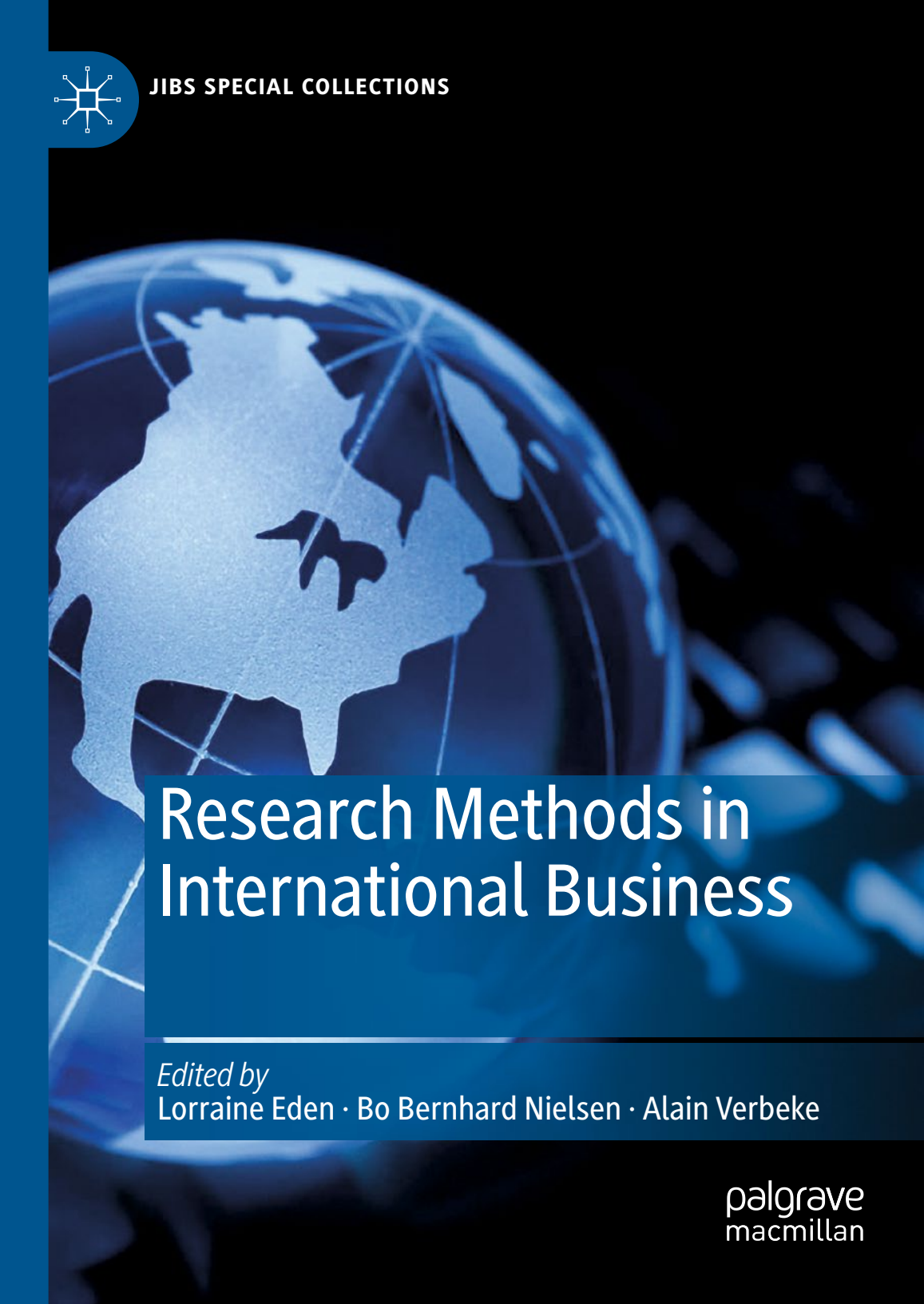
JIBS SPECIAL COLLECTIONS
Research Methods in International Business
Edited by
Lorraine Eden · Bo Bernhard Nielsen · Alain Verbeke
palgrave
macmillan

JIBS Special Collections

Series Editor
Academy of International Business

Technical Series Editor: Alain Verbeke is a Professor of International Business Strategy and holds the McCaig Research Chair in Management at the Haskayne School of Business, University of Calgary (Canada). He was elected as the Inaugural Alan M. Rugman Memorial Fellow at the Henley Business School, University of Reading (UK), and is Editor-in-Chief of the *Journal of International Business Studies.* He is also an Adjunct Professor at the Solvay Business School, Vrije Universiteit Brussel (Belgium). He is an elected Fellow of the Academy of International Business and has served as Secretary and Treasurer of the Fellows.

Journal of International Business Studies **(*JIBS*)** is an official publication of the Academy of International Business and is the top-ranked journal in the field of international business. The goal of *JIBS* is to publish insightful, innovative and impactful research on international business. *JIBS* is multidisciplinary in scope and interdisciplinary in content and methodology. For more information, visit www.jibs.net.

Academy of International Business (AIB) is the leading association of scholars and specialists in the field of international business. A global community for the creation and dissemination of knowledge about international business and policy issues, the AIB transcends the boundaries of single academic disciplines and managerial functions to enhance business education and practice. For more information, visit aib.msu.edu.

More information about this series at
http://www.palgrave.com/gp/series/15231

Lorraine Eden
Bo Bernhard Nielsen • Alain Verbeke
Editors

Research Methods in International Business

palgrave
macmillan

Editors
Lorraine Eden
Department of Management
Texas A&M University
College Station, TX, USA

Alain Verbeke
University of Calgary
Calgary, AB, Canada

University of Reading
Reading, UK

Vrije Universiteit Brussel
Brussels, Belgium

Bo Bernhard Nielsen
The University of Sydney Business School
The University of Sydney
Sydney, NSW, Australia

Copenhagen Business School
Frederiksberg, Denmark

ISSN 2397-8775 ISSN 2397-8783 (electronic)
JIBS Special Collections
ISBN 978-3-030-22112-6 ISBN 978-3-030-22113-3 (eBook)
https://doi.org/10.1007/978-3-030-22113-3

Contents

List of Figures

List of Tables

Part I

Introduction

1

Research Methods in International Business: Challenges and Advances

Bo Bernhard Nielsen, Lorraine Eden, and Alain Verbeke

We thank Agnieszka Chidlow, Henrik Gundelach, Stewart R. Miller, and Catherine Welch for their comments and help on an earlier draft of this chapter.

B. B. Nielsen
The University of Sydney Business School, The University of Sydney, Sydney, NSW, Australia

Copenhagen Business School, Frederiksberg, Denmark
e-mail: bo.nielsen@sydney.edu.au

L. Eden (✉)
Department of Management, Texas A&M University, College Station, TX, USA
e-mail: leden@tamu.edu

A. Verbeke
University of Calgary, Calgary, AB, Canada

University of Reading, Reading, UK

Vrije Universiteit Brussel, Brussels, Belgium
e-mail: alain.verbeke@haskayne.ucalgary.ca

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_1

Introduction

> It is because cross-national research is so consuming of time and other resources that we ought to be willing to settle for less than the ideal research designs. It is also in this context that we should be more forgiving of researchers who might seem opportunistic in selecting the countries and problems for cross-cultural research. (Sekaran 1983: 69)

Most, perhaps all, international business (IB) scholars would now agree that the above statement, made more than 35 years ago, is no longer valid. The challenges and costs associated with obtaining adequate samples and/or measures do not free a researcher from the responsibility of crafting a well-thought-out research design and adopting rigorous research methods. Gone are the days when IB researchers could be excused for a relative lack of methodological rigor due to "real-world constraints", such as the absence of requisite financial resources to deploy the most advanced methodological approaches in complex international settings (Yang et al. 2006). For today's scholars, staying up to date—that is, understanding and using the best available and most appropriate research methods—clearly matters.

There have been several calls for expanding research settings and using more advanced methods to analyse complex, cross-border phenomena. A recent review of the methodological trajectories found in *JIBS* over the past 50 years (1970–2019) shows a dramatic rise in sophistication of the methods deployed (Nielsen et al. 2019), which is a testament to a maturing field. In fact, most IB journals with high Web-of-Science impact scores now formally or informally adhere to rigorous methodological standards and impose requirements on authors similar to the ones prevailing at the leading journals in other business disciplines such as marketing, organizational studies, and strategy (e.g., Hahn and Ang 2017; Meyer et al. 2017).

Moreover, during the past two decades, much has been written on sound methodological practices and rigor in IB research, some of which is documented in this book (see also, e.g., Nielsen and Raswant 2018; Nielsen et al. 2019). While *Organizational Research Methods* remains the primary outlet for pure methods articles, several of our scholarly journals,

particularly *JIBS* and the *Academy of Management Journal*, have regularly published editorials and articles on best practices in research methods. Some of the best *JIBS* editorials on research methods have been included in this book.

IB scholars, like academics everywhere, study basic research methods in undergraduate and graduate classes such as statistics, econometrics, qualitative methods, and research design. We learn or teach ourselves how to use STATA, SPSS, SAS, or NVivo. These activities happen early in our careers, so we need to update our knowledge of best research methods practices on a regular basis. To partially address this need, within professional associations such as the Academy of International Business (AIB) and the Academy of Management (AOM), much attention is paid to sharing knowledge on best research practices through, for example, doctoral and junior faculty consortia. The Research Methods Shared Interest Group (RM-SIG) was recently established within the AIB to promote the advancement, quality, diversity, and understanding of research methodologies by IB scholars. The RM-SIG is just one example of a broad range of initiatives being undertaken by the global community of IB scholars to both help keep scholars abreast of the latest research methods and to push the field forward in the methods sphere.

We view our book as part of this ongoing initiative, which has spread across all business and social science disciplines, of improving the overall quality of methods used in business research. Our specific focus in this book is best practices in IB research methods. We take stock of some key challenges faced by the field in the realm of research design and methods deployed, and we also discuss recent advances in overcoming these challenges. We view our book as a unique, up-to-date reference source on good and best practices. By identifying and assembling a set of exemplary *JIBS* articles together with commentaries and reflections on these articles, we hope to share with the IB research community at large what now constitutes these best practices.

Our objectives for this introductory chapter are four-fold. First, we think that it is important to reiterate how and why high-quality research methods matter to IB scholars. Our second goal is to identify a number of base-line systemic methodological challenges facing scholars in IB research. A third goal is to introduce the various *JIBS* articles included in

the book, which were selected because they represent sound and best practices to help overcome these challenges. We also briefly introduce the insightful commentaries and reflections provided by leading scholars on each of the *JIBS* pieces. Our last goal is to provide recommendations to IB scholars in the hope that the field will continue its positive trajectory and evolve into a net exporter of research methodology.

How and Why High-Quality Research Methods Matter?

The answer to the questions "*How* and *why* high-quality methods matter in international business research?" may be self-evident; however, we believe it is worth reiterating the benefits of using high-quality research methods—and the costs of using inadequate, outdated, or sloppy methods—to the field of IB studies.

First, the benefits. Acting with academic integrity means being consistent with "the values of honesty, trust, fairness, respect and responsibility in learning, teaching and research" (Bretag 2019). Acting with integrity in our research requires using high-quality research methods that promote the "truth" and minimize error. Our research is often motivated by puzzles we see around us in the real world. We use and test theories in order to better understand the world in which we live. We want our research to be credible and useful to other scholars, policy makers, managers, and the public. As is the case with any area of research, the choice and application of specific research methods largely determine the quality of subsequent knowledge creation, as well as the intellectual contribution made to the field. A well-crafted methodological approach can go a long way towards establishing a study's rigor and relevance, thereby enhancing its potential impact, both in terms of scholarly advancement and improvement of managerial and policy practice. Only by practising the state-of-the-art in terms of research design, including sampling, measurement, analysis, and interpretation of results, will other scholars have confidence in the field's findings. In fact, sound methodology—"research that implements sound scientific methods and processes in both quantitative and

qualitative or both theoretical and empirical domains"—is one of the seven core guiding principles of the movement for Responsible Research in Business and Management (https://rrbm.network/).

However, scholars often view the benefits from research integrity as accruing only in the long term and primarily to society as a whole. In the short term, pressures to publish and the desire for tenure and promotion may be much more salient. Eden, Lund Dean, and Vaaler (2018: 21–22) argue that academia is full of "research pitfalls for the unwary" that can derail even well-intentioned faculty who believe they are acting with academic integrity. Doctoral students and junior faculty members are especially susceptible to these pitfalls due to the challenges they face as new entrants to academia—the liabilities of newness, resource dependence, and outsiderness.

The polar opposite of academic integrity is "scientists behaving badly" by engaging in academic misconduct/dishonesty (Eden 2010; Bedeian, Taylor and Miller 2010). IB scholars, similar to scholars throughout the social and physical sciences, are familiar with the three main types of academic misconduct: falsification (manipulating or distorting data or results), fabrication (inventing data or cases), and plagiarism (copying without attribution). In FFP (falsification, fabrication, and plagiarism) cases, researchers fail to tell the truth in scientific communications about their research (Butler et al. 2017). Such academic misconduct corrupts the research process and damages public trust in scientific literature. Research misconduct occasionally leads to retraction of the published work, and there is empirical evidence that the majority of retracted journal articles were retracted due to misconduct by the authors (Fang et al. 2012). Retraction also carries with it significant financial and personal consequences for the authors and substantial ripple effects on one's colleagues, students, prior collaborators, and home institutions; see, for example, the types and estimates of costs in Stern, Casadevall, Steen and Fang (2014); Michalek, Hutson, Wicher and Trump (2010); Tourish and Craig (2018); and Hussinger and Pellens (2019).

Because of the huge costs involved when scientists behave badly, most of our universities, journals, and professional associations now have Codes of Ethics that outline, prohibit, and punish research misconduct. The Academy of International Business (AIB), for example, now has

three ethics codes, one each for the AIB journals, members, and leadership (https://www.aib.world/about/ethics/). In fact, to the best of our knowledge, *JIBS* was the first scholarly business journal to have its own code of ethics (Eden 2010). Many of our journals and professional organizations also belong to COPE, the Committee on Publication Ethics (https://publicationethics.org/), which provides detailed process maps for handling various types of academic misconduct.

While the costs of academic dishonesty are well understood, there are also huge costs to scholarly inquiry from engaging in the grey area between academic integrity and misconduct, that is, in what has been called "sloppy science" or "questionable research practices" (QRPs) (Bouter et al. 2016). QRPs are research methods that "operate in the ambiguous space between what one might consider best practices and academic misconduct" (Banks et al. 2016a: 6). QRPs can occur in the design, analysis, or reporting stages of research. The typical motivation for QRPs is the desire of authors to present evidence favouring their hypotheses and to increase the likelihood of publication in a high-impact journal (Edwards and Roy 2017; Eden et al. 2018).

Banks et al. (2016a) and Banks et al. (2016b) identify six types of QRPs: selectively reporting hypotheses, excluding data post hoc, HARKing (hypothesizing after the results are known), selectively including control variables, falsifying data, and poor reporting of *p* values. Bouter et al. (2016) provide a list of 60 major and minor research misbehaviours, which they group into 4 areas based on research stage: study design, data collection, reporting, and collaboration. The authors conclude that selective reporting and citing, together with flaws in quality assurance and mentoring, are the top four examples of researchers cutting corners and engaging in sloppy research practices.

QRPs have high costs; they can "harm the development of theory, evidence-based practice, and perceptions of the rigor and relevance of science" (Banks et al. 2016b: 323). Incorrect statistical procedures can lead to flawed validity estimates, as shown in Antonakis and Dietz (2011). Engaging in low-quality or unethical research methods makes it impossible for other scholars to reproduce and replicate our results, leading to an overall distrust in scholarly publications (Rynes et al. 2018: 2995).

Evidence that scholars in the social and physical sciences, including in our business schools, do engage in QRPs and academic dishonesty is widespread, ranging from lists of retracted articles on Retraction Watch (http://retractionwatch.org) to stories in the *New York Times* and *Nature*. The evidence suggests that while only a small percentage of researchers may engage in academic misconduct (FFP), sloppy science is much more widespread (Bouter et al. 2016; Butler et al. 2017; Hall and Martin 2019).

We believe that the spread of QRPs in academia is partly because not everyone shares our interest in and passion for research methods. For many, if not most, academics, learning the "ins and outs" (the "dos and don'ts") of particular research methods is difficult, a bit like "bad medicine"—you know it must be good for you because it tastes terrible. Math and statistics anxiety exist even among doctoral students and perhaps even more so among full professors! Over our careers, we have seen many examples of faculty who learned one research method early in their careers and relied on that method for all their projects, rather than learning new, more appropriate methods. We have also seen "slicing and dicing" of research projects where the workload was parsed out among co-authors, and no supervision of, or interaction with, the co-author assigned to write the research methods and results sections was provided. When scholars cut corners due to math/statistics anxiety and/or laziness, they open the door to questionable research practices.

QRP is clearly and issue facing all social scientists, not only IB researchers. Science is in the middle of a "reproducibility and replicability crisis" and "international business is not immune", as Aguinis, Cascio, and Ramani (2017: 653) argue. Once scholars begin to have doubts about the findings of scientific research, we have started on the path towards viewing all research results with a jaundiced eye as "fake news". We know that the conclusions drawn from an empirical study are only as solid as the methodological practices that underlie the research. If we want to raise the quality and impact of IB research, we need to bring not only rigor but also transparency and credible expectations of reliability back into our research methods. We turn now to a discussion of how to do this.

Methodological Challenges

Our second goal in this chapter is to identify a number of base-line systemic methodological challenges facing scholars in IB research. These methodological challenges, we argue, seem to plague IB research because of the types of research questions asked and the cross-border contexts studied.

For example, consider distance as a concept in IB research. A great number of measures and methodologies have attempted to capture the complexity of distance between nations (even accounting for within country variations), and the implications for firms operating within and across different types of distance (e.g., geographic, cultural, economic, institutional). Many IB research questions in the realm of distance can only be answered by taking into account multiple levels of analysis beyond the country level, including individuals (e.g., senior executives), headquarters, subsidiaries, strategic groups, industries, and so on (Nielsen and Nielsen 2010). However, the methods used to examine multilevel phenomena in IB studies (such as variance decomposition) have often been relatively unsophisticated and may have left key questions unanswered (Peterson et al. 2012). Moreover, of special interest to IB research that seeks practical relevance are the executives who formulate, implement, and monitor strategic initiatives related to IB operations (Tihanyi et al. 2000; Nielsen and Nielsen 2011). The methodological approaches deployed to tap into executives' motives, preferences, values, and ultimate decisions (such as the usage of demographic proxies) have, at times, been limited in their capacity to describe and explain these complex phenomena (Lawrence 1997).

Below we identify some of the most salient methodological challenges facing IB researchers. This list of challenges is by no means exhaustive and we invite the community of IB scholars to contribute to this list and—more importantly—to provide input into possible solutions and best practices, for instance, by actively partaking in the AIB RM-SIG activities and/or contributing to the website (https://rmsig.aib.world/) or newsletter.

The discussion below is organized around three major research phases in terms of methodological choices: (1) Problem definition and research questions; (2) Research design and data collection; and (3) Data analysis and interpretation of results.

Phase 1: Problem Definition and Research Questions

As suggested above, IB refers to a complex set of phenomena, which require attention to both similarities and differences between domestic and foreign operations at multiple levels of analysis. We see the following key methodological challenges in the realm of problem definition.

1. *Is the problem truly international?*

   Isolating the international (cross-border/cultural) influence on the key relationship(s) in the study may require a deep understanding of both the domestic (i.e., the country of origin) and foreign (i.e., country of operations) business environments, in terms of political, institutional, economic, social, cultural, and behavioural characteristics. Methodologically, this may require input from researchers who are familiar with these environments and/or necessitate field trips to establish the nature of the "international" phenomenon.
2. *Are concepts and theories equivalent and comparable across contexts (cases, countries, cultures, etc.)?*

   Much current debate in IB revolves around the applicability of "standard" internationalization theories to emerging market firms (e.g., Santangelo and Meyer 2011; Cuervo-Cazurra 2012; Ramamurti 2012). To the extent that the applicability of theories and their key assumptions (e.g., the degree of confidence in the reliability of societal institutions) differ across national borders, IB researchers must embed such differences in their research design and develop suitable research questions with attention to equivalence and compatibility across contexts. It is now widely recognized that much IB research may inherently be about "contextualising business" (Teagarden et al. 2018), but at the same time scholars must guard against the possibility of alternative explanations or biases introduced by the very nature of the

context(s) being investigated. Identifying explicitly contextual influences and their potential impacts, both in the design and interpretation of outcomes of a particular study, are critical for determining the boundaries within which theories used might be applicable. For instance, the impact of a variable such as state ownership on a variety of outcomes (e.g., the probability of going international or of engaging in international mergers vis-à-vis other entry modes or the location of outward foreign direct investment) may be largely dependent on the political and institutional environment in the country of origin (Estrin et al. 2016; He et al. 2016a, b). State ownership in China versus Norway may have completely different implications for explaining and predicting international expansion moves.

3. *What types of research questions are being asked?*

   A significant challenge is to establish a clear linkage between the specificity versus general nature of the research questions asked, and the related ambition to explain empirical phenomena and extend theory. IB scholars sometimes claim they will try to answer general research questions, such as: Where do firms locate their international operations? What entry mode choice is the best given the nature of the knowledge assets involved? Do firms benefit from internationalization? Unfortunately, the research design, including inter alia sample limitations and a restricted set of variables, then sometimes leads scholars to overestimate the generalizability of their results and to make exaggerated claims as to their contributions to theory. There is nothing wrong with relatively narrow, phenomenon-driven empirical research, but such research is unlikely to answer general research questions with important implications for theory. In this realm, IB scholars should always remember that their research questions should drive data collection and choice of methodology—not the other way around.

Phase 2: Research Design and Data Collection

Research design and data collection efforts are also susceptible to a number of challenges that are especially salient in IB research. The research

design must ensure the equivalence and comparability of primary and secondary data, which may be related to different environmental contexts. Here again, we see three main challenges.

1. *What is an appropriate sample?*
   In many instances, particularly in developing countries, reliable information about the target population may not be available from secondary sources. Government data may be unavailable or highly biased. Lists of targeted respondents may not be available commercially (e.g., small samples of certain respondents, such as entrepreneurial women in some cultures). In general terms, sampling is often performed with the implicit assumption that all sampled firms or individuals in a nation share the same underlying characteristics, such as national cultural characteristics, but this is often untrue. For example, in an extreme case, a US-based company entering China might be managed by Chinese immigrants, and a potential joint venture partner in China considered by this US firm might be managed by US-trained executives or even US natives. The point is to avoid sampling in IB studies on the basis of convenience, without properly evaluating whether assumed characteristics of the sample actually hold. In the presence of inadequate sampling, any purported contributions to IB theory development must be viewed with suspicion. In the realm of cross-cultural studies, Ryen, Truman, Mertens, and Humphries (2000) and Marschan-Piekkari and Welch (2004) highlight various challenges associated with deploying qualitative techniques in developing countries; for example, respondents in cross-national surveys may interpret specific ideas or concepts put forward by researchers from developed countries in a culturally specific manner, rendering any comparisons among nations untenable.
2. *What is an appropriate sample size*?
   The answer may be much more complex than suggested by rules of thumb or generally accepted conventions in statistical analysis. The simple reason is that samples in IB research may have much stronger heterogeneity in terms of relevant variables than in domestic settings. For example, when assessing the impact of cultural distance on governance mode choice (such as a joint venture versus a wholly owned

operation), it does not suffice to take into account the normal distance variables, such as geographic distance, that would be considered in domestic settings. A variety of distance parameters should be considered (e.g., institutional and cultural/psychic distance). In addition, the actual impact of these additional distance variables will depend on firm-level parameters such as the composition of the top management team (e.g., international experience and cultural diversity). As a result of the greater number of relevant variables, requisite sample size should also increase, in this case to isolate the discrete effects of a larger number of explanatory variables on governance choice.

3. *How to avoid non-sampling errors?*

In IB research spanning multiple cultures, both measurement non-equivalence and variations in interviewer quality can lead to non-sampling errors. The increasing availability of large international surveys has opened a wide avenue of new possibilities for researchers interested in cross-national and even longitudinal comparisons. Such surveys build on constructs measured mostly by multiple indicators, with the explicit goal of making comparisons across different countries, regions, and time points. However, past research has shown that the same scales can have different reliabilities in different cultures. Davis et al. (1981) demonstrated that two sources of "measure unreliability", namely the assessment method deployed and the nature of the construct, can confound the comparability of cross-cultural findings. Thus, substantive relationships among constructs must be adjusted for unequal reliabilities before valid inferences can be drawn. Hence, it is critical to assess whether questions "travel" effectively across national and cultural borders (Jowell et al. 2007).

It is also important to test empirically the extent to which survey responses are cross-nationally equivalent, rather than erroneously assuming equivalence. IB scholars are therefore advised to test the assumption of measurement equivalence empirically, for instance, by applying a generalized latent variable approach (Rabe-Hesketh and Skrondal 2004) or optimal scaling procedures (Mullen 1995). Other possible strategies include (1) identifying subgroups of countries and concepts where measurement equivalence holds, and continuing with cross-country

comparisons within these subgroups; (2) determining how severe the violation of measurement equivalence is, and whether it might still allow meaningful comparisons across countries; and (3) at a minimum, trying to explain the individual, societal, or historical sources of measurement non-equivalence, and the potential impact thereof on results (Davidov et al. 2014).

Phase 3: Data Analysis and Interpretation of Results

The internal validity of IB research improves if the outcomes of a study have fewer rival explanations. The approach adopted to analyse data and interpret results should address and control for such alternative explanations (Cuervo-Cazurra et al. 2016). Here, IB researchers can formulate plausible, rival hypotheses that could explain the results (see, e.g., Nielsen and Raswant 2018). Even with rival hypotheses in play, we see the following four challenges:

1. *How to address outliers?*

   IB research may be particularly susceptible to the impact of outliers for two reasons. First, outlier outcomes are often included in IB studies, even though the economic actors responsible for the outlier results were not considered ex ante as being members of the target population. One example is the presence of first-generation immigrant managers or firm owners in studies of national companies' choices of foreign locations for international activities. The results of these studies may still be relevant, but if the purpose of the study is, for example, to assess the impact of national cultural distances (measured as the supposed distances between one home country and a number of potential host countries) on location choices, then the aggregate responses of immigrant managers and owners will very possibly function as outlier outcomes. The reason is that their decision-making on location may not be influenced at all by the national cultural distances considered, for example, if their first-choice foreign location is their country (or region) of origin. Second, again in the realm of distance, many multinational enterprises (MNEs) employ expatriates, who are

likely to have characteristics different from those shared by the general population of managers, whether in their home country or in the host country where they work. It would therefore be a mistake to assess, for example, head office–subsidiary interactions based on characteristics of the home and host countries at play, when expatriate executives play key roles in these interactions. Various statistical techniques are available to assist IB scholars in identifying multivariate outliers (e.g., see Mullen et al. 1995).

2. *How to choose the level(s) of analysis?*

Many IB phenomena are by default multilevel in nature. For instance, MNEs are nested within home and host country contexts. By the same token, subsidiaries are nested within MNE "hierarchies", typically the headquarter(s) of the parent company. A number of scholars have emphasized that it is imperative to approach IB phenomena at a variety of levels of theory and analysis (Arregle et al. 2006; Peterson et al. 2012, Goerzen et al. 2013), with due attention paid to nesting or cross-level effects (Andersson et al. 2014). Failure to account for the multilevel structure of hierarchically nested data is likely to yield statistical problems. Such problems arise from improperly disaggregating datasets, thereby violating the assumption of independence among observations and ignoring inter-class correlations that increase the risk of type I and type II errors (Snijders and Bosker 2011; Raudenbush and Bryk 2002). If these problems arise, random coefficients modelling (RCM) offers three substantial advantages over traditional statistical models (Raudenbush and Bryk 2002): (1) improved estimation of effects within each level; (2) possibility to formulate and test hypotheses about cross-level effects; and (3) portioning the variance and co-variance components among levels.

In addition, IB phenomena are often influenced by contexts that are interwoven in a more complex fashion than "simply" being hierarchically nested. On the one hand, MNE subsidiaries are nested within their parent companies, but also within national contexts (e.g., home/host country contexts) in a hierarchical way. On the other hand, MNEs (both parent companies and their subsidiaries) are cross-nested within home and host countries, as well as within industries, but the

countries involved are not nested within industries or vice versa. Empirically (as well as conceptually), it is therefore important to recognize heterogeneity at the firm, industry, and country levels, as well as cross-nested embeddedness. Here, a special application of RCM—namely cross-classified random coefficients modelling or CCRCM—may help isolate the effects of the cross-cutting hierarchies (e.g., country and industry) on the dependent variable (e.g., firm performance), thereby avoiding model under-specification and biased results (Fielding and Goldstein 2006). Though still uncommon in IB research (for a recent example, see Estrin et al. 2017), scholars are strongly encouraged to account for the nested structure of the IB phenomena they study, and for non-hierarchical embeddedness in particular, when theorizing about—and testing the effects of—context on firm (industry, team, or individual) behaviour.

Also related to the presence of multiple levels is the challenge of ecological fallacies. These refer to the unqualified usage in one level of analysis of the variable scores that were derived from analysis at another level. As one example, an ecological fallacy comes into play when a researcher uses culture-level scores (e.g., based on Hofstede's cultural dimensions or GLOBE measures) without conducting individual-level analyses to interpret individual behaviour. Conversely, a problem of "atomistic" fallacy arises when a researcher constructs culture-related indices based on individual-level measurements (attitudes, values, behaviours), without conducting societal-level cultural analysis (Schwartz 1994). Culture can be important for many IB decisions and outcomes, whether as a distance measure (Beugelsdijk et al. 2018) or a contextual control variable (Nielsen and Raswant 2018). Yet, ecological fallacy challenges are seldom addressed fully, despite ample evidence that they matter (Brewer and Venaik 2014; Hofstede 2001: 16; House et al. 2004: 99).

3. *How to avoid personal bias in interpreting and reporting results?*

In IB studies, a researcher working out of a particular context (such as a national culture or a set of national economic institutions relevant to IB transactions) must often interpret data gathered in various other contexts. The researcher's own context-dependent biases may then affect her or his interpretation of the outcomes. We noted above that

concepts may not easily "travel" across borders, and that theories and methods are not necessarily "equivalent" across contexts. One should therefore avoid assuming too easily the universality of concepts, theories, and methods. In addition, researchers themselves may potentially introduce another bias based on their personal ethnocentrism and other context-determined preferences.

These biases often remain undetected, especially when scholars build upon extant streams of equally biased research, sometimes amplified by individuals and "clubs" of like-minded scholars adopting the same methods and involved in editorial reviewing processes. Individuals may actually have a preference for—and may thereby be instrumental to—long waves of biased research being published. Meade and Brislin (1973) suggested a partial solution to this problem, relevant especially in the context of multinational research teams. They suggested that researchers from each country should independently interpret the results obtained, so that inter-interpreter reliability can be assessed.

On a positive note, the average number of authors and national diversity in terms of authors of *JIBS* articles[1] has increased substantially over the past 50 years. In the 1970s, the average *JIBS* article had 1.48 authors with 17 percent of first authors being from a country other than the United States. In the 2000s (2000–2009) these numbers increased to 2.33 authors per article with 55 percent of first authors being from a country other than the US. Since 2010 these numbers have further increased to 2.88 authors per article and 62 percent non-US first authors (Nielsen et al. 2019).

Even if an author has reflected adequately on the questions outlined above, there is still *work* to be done, since a wide range of methods-related decisions must still be made. To aid the reader in making those decisions, we now turn to some methodological advances in conducting IB research. In the next section, we examine 11 *JIBS* publications that were designed to promote a level of sophistication at par with or ahead of other business disciplines but keeping in mind the specificities of IB research. We augment these pieces with new Commentaries and Reflections on how the field has advanced since these *JIBS* articles were first published.

Methodological Advances

Our third goal in writing this introductory chapter is to introduce the remaining chapters in this book. These chapters were selected because they represent good and best practices to help overcome the methodological challenges we identified above. Each chapter included in this volume represents a significant advance in IB methods, given the field's unique features.

We have organized the remainder of this book into 11 distinct Parts. Each Part has one to three chapters. The first chapter is an original *JIBS* article on a particular methods topic published between 2010 and 2019. Each *JIBS* article is followed by an insightful Commentary from one or more content experts who deliver forward-looking observations on the importance of the methodological challenges considered and on the most effective ways to respond to such challenges. Four Parts also include a third chapter, a Further Reflections note prepared by one or more of the authors of the original *JIBS* article.[2]

The three co-editors selected the 11 original *JIBS* articles included in this book after a detailed, lengthy, and iterative search process. We chose 2010 as our starting year so as to include 10 years of *JIBS* publications. We used several criteria for selection, both quantitative ones such as citation counts, and qualitative ones when re-reading the articles ourselves where we assessed their contribution on three dimensions. First, we wanted each article to represent a different methodological challenge in IB research. Second, an important selection criterion was our assessment of each article's likely contribution to raising the rigor and relevance of contemporary IB scholarship. Third, we also used as a selection criterion the need to respect diversity and plurality in methodological focus, thereby acknowledging the importance of both qualitative and quantitative methods, as well as mixed-methods approaches.

Parts II through IV (Chaps. 2, 3, 4, 5, 6, 7, and 8) in this book are concerned with recurring methodological challenges in contemporary IB research and offer best practices to overcome these challenges. Parts V through VII (Chaps. 9, 10, 11, 12, 13, 14, and 15) deal with methodological challenges and advances in qualitative research in IB. Parts VIII

through X (Chaps. 16, 17, 18, 19, 20, 21, and 22) discuss methodological challenges in quantitative methods and suggest ways to deal with these challenges. The volume concludes with Parts XI and XII (Chaps. 23, 24, 25, 26, and 27), which focus on frontier methodological challenges in IB research.

In the rest of this section we summarize the main ideas presented in each of the 11 Commentaries and 4 Reflections chapters. Since these 15 chapters also provide summaries of the original 11 *JIBS* articles, we do not include them here.

Part II (Chaps. 2 and 3) deals with the reproducibility and replicability of research findings. In his Commentary on Chap. 2, "Science's Reproducibility and Replicability Crisis: International Business Is Not Immune" by Aguinis, Cascio, and Ramani (2017), Andrew Delios argues in Chap. 3 that the solutions and recommendations provided to improve our empirical methods—mainly the use of meta-analysis—miss the opportunity to question more fundamentally whether existing research protocols should continue to be standard operating procedures or replaced. In Delios' view, the so-called replication crisis is not a crisis; "it is a reality and a logical off-shoot of the accepted research standards we have in the field of IB research". He suggests that our decision as a community of scholars is not whether we should engage in replication or reproducibility studies but rather "whether we want to make the investments necessary to re-think the core of our methods and to address the long-standing systemic challenges to conducting good, repeatable empirical research in international business".

Part III (Chaps. 4 and 5) builds further on this theme and suggests best practices with respect to conducting, reporting, and discussing the results of quantitative hypothesis testing, so as to increase rigor in IB research. Agnieszka Chidlow, William Greene, and Stewart Miller discuss and augment the insights from Chap. 4, "What's in a *p*? Reassessing Best Practices for Conducting and Reporting Hypothesis-Testing Research" by Meyer, Witteloostuijn, and Beugelsdijk (2017). Their Commentary suggests a more rational approach to reporting the actual level of significance by placing the burden of interpretive skill on the researcher since there is no "right" or "wrong" level of significance in hypotheses testing. Scholars are encouraged to give higher priority to selecting appropriate levels of

significance for a given problem instead of the misleading culture of the "old asterisks habit". The idiosyncratic features of many IB phenomena call into question conventionally accepted significance levels "because different classes of research may require different levels of alpha". The commenters also discuss the pros and cons of modern technology in ensuring credible and ethical research designs and execution. The authors ultimately place the burden on the entire IB scholarly community—authors, co-authors, reviewers, editors, and PhD supervisors—to avoid QRPs such as HARKing and *p*-hacking.

Part IV (Chaps. 6, 7, and 8) completes the discussion of recurring challenges and best practices by addressing alternative explanations to improve the validity and generalizability (i.e., "trustworthiness") of empirical research in IB. Jonathan Doh comments on Chap. 6, "Can I Trust Your Findings? Ruling Out Alternative Explanations in International Business Research" by Cuervo-Cazurra, Andersson, Brannen, Nielsen, and Reuber (2016). The original *JIBS* article provides guidance on how to ensure that authors establish the "correct" relationships and mechanisms so that readers can rely on their findings. Doh in his Commentary (Chap. 7) makes several important observations, including that "too often scholars are fearful of revealing any findings or cases that are contrary to their overall hypotheses (whether formal or informal) and may somewhat subconsciously or unknowingly suppress this countervailing information". He also reiterates points made by previous commenters that "the core challenge in IB concerns some of the generally accepted norms, practices, and assumptions that undergird what we consider to be an acceptable empirical exposition". Doh closes with a plea for more attention to societal "grand challenges" and argues that while such research "may require interdisciplinary approaches, multilevel methods, and consideration of a diverse range of societal actors and influences, it offers an organizing principle for IB research that seeks to achieve relevance, rigor, and real-world contribution".

Chapter 8 provides Further Reflections on the original *JIBS* article (Chap. 6) written by three of the authors (Brannen, Cuervo-Cazurra, and Reuber). They provide two observations; (1) how difficult it is for scholars to take up the challenge of tackling significant, bold, real-world phenomena with an open-minded, interdisciplinary multi-methods

approach, and (2) how challenging it is to review mixed-methods articles. To help with the latter, the authors make several astute suggestions for reviewers of mixed-methods articles, including allowing for multiple story lines to develop while paying particular attention to how data are used to build evidence. Their hope is to increase methodological ambidexterity among IB scholars.

Parts V through VII (Chaps. 9, 10, 11, 12, 13, 14, and 15) deal with methodological challenges in qualitative research in IB. Part V (Chaps. 9, 10, and 11) focuses on how to theorize from qualitative research, especially case studies, and the critical role of context. In Chap. 10, Kathleen Eisenhardt makes several important points in her Commentary on Chap. 9, "Theorising from Case Studies: Towards a Pluralist Future for International Business Research" by Welch, Piekkari, Plakoyiannaki, and Paavilainen-Mäntymäki (2011). Specifically, Eisenhardt argues that "while helpful, the article's central typology and 2x2 create artificial distinctions" and goes on to suggest that "its interpretation of theory building cases combines cherry-picked phrases with an eighteenth-century view of positivism". In Eisenhardt's view, the role of context was and is always central to qualitative research, and cases can be used to both develop and test theory. She views cases as independent experiments where replication logic is germane and where one should seek to develop an underlying theoretical logic. Eisenhardt concludes that there is an emerging recognition of the similarity across inductive methods as well as the relevance of specific methods for different types of research questions and contexts. She advocates for more attention to the role of language in defining and naming constructs, which may take on different meanings in different cultural or linguistic contexts. New technologies, such as machine learning and big data, may offer promise for the future with regard to meeting some of the challenges of case study research.

Eisenhardt's Commentary is followed by Further Reflections (Chap. 11) provided by the four original authors of the 2011 article. In their reflections piece, the authors outline three ways in which the themes of their 2011 article have been further developed since its publication. Firstly, they point to an increase in studies taking an abduction approach, which emphasizes a theoretical starting point and offers qualitative researchers a vocabulary to articulate how they iterate between theory

and data. Secondly, the authors emphasize a need for more holistic explanations, dissolving the dichotomy between qualitative and quantitative research. Such approaches may include rarely used methodologies, such as longitudinal single cases, historical methods, and use of retrospective data, which may help researchers trace causal mechanisms over time and develop process explanations addressing how social change emerges and evolves. Finally, the authors see a general trend toward combining contextualization with causal explanation, which holds great promise for future IB research.

Part VI (Chaps. 12 and 13) investigates the linkages between historical and qualitative analyses and suggests that more attention be paid to longitudinal qualitative research in IB. Catherine Welch comments on Chap. 12, "Bridging History and Reductionism: A Key Role for Longitudinal Qualitative Research" by Burgelman (2011). Her Commentary in Chap. 13 provides not only new insights on the methodological challenges of conducting longitudinal qualitative research in IB, but also on the potential value added that might result from this approach. She highlights how a longitudinal, qualitative research approach would allow IB scholars to "go beyond reductionist forms of explanations to account for complex causality, system effects, context dependence, non-linear processes and the indeterminacy of the social world". The implications of what she calls "Burgelman's vision" are far reaching: it requires us to "rethink the research questions we pose, the analytical techniques we use, the nature of the theories we develop, and the way we view our role as social theorists". She contrasts the "standard" approach to qualitative research with the alternative offered by the longitudinal qualitative research "vision" of Burgelman and concludes that to realize this would entail a paradigmatic shift yet to be implemented in IB research. However, there is hope; these two paradigms may be mutually supportive as "longitudinal qualitative research may form a 'bridge' between history and reductionist research, and a 'stepping stone' to formal mathematical models". In Welch's words: "IB researchers have the opportunity to diversify and enrich the methods we use and the theories we develop".

Part VII (Chaps. 14 and 15) discusses the relevance and applicability of Fuzzy-set Qualitative Comparative Analysis (fsQCA) for advancing IB theory. Stav Fainshmidt comments on Chap. 14, "Predicting Stakeholder

Orientation in the Multinational Enterprise: A Mid-Range Theory" by Crilly (2011). Fainshmidt's Commentary in Chap. 15 discusses some of the key judgement calls that researchers using fsQCA must make throughout the analytical process. Applying fsQCA may help IB scholars to straddle both qualitative and quantitative analyses in an iterative manner, which helps to pinpoint causal mechanisms as well as generalize and contextualize qualitative findings that often span multiple levels of analysis. Yet such analysis requires important judgement calls regarding (among other things): (1) calibration, (2) frequency, and (3) consistency. Fainshmidt points to several other analyses that may augment the ones proposed by Crilly, such as varying the frequency threshold, evaluating the impact of alternative calibration approaches, and revisiting decisions related to counterfactuals. He also suggests using the proportional reduction in inconsistency (PRI) statistic. This should help researchers who utilize fsQCA, to correct for the potential contribution of paradoxical cases and to identify paradoxical rows in the truth table, thereby producing a more accurate solution. He cautions, however, that these analyses "should be considered in light of the decisions made in the main analysis and the context of the study or data at hand".

Parts VIII through X (Chaps. 16, 17, 18, 19, 20, 21, and 22) discuss methodological challenges in quantitative methods and suggest ways to deal with them. Part VIII (Chaps. 16 and 17) draws attention to the difficulty of adequately theorizing and accurately empirically testing interaction effects within and across levels of analysis in IB research. Jose Cortina discusses Chap. 16, "Explaining Interaction Effects Within and Across Levels of Analysis", by Andersson, Cuervo-Cazurra, and Nielsen (2014) and makes three important observations in his Commentary in Chap. 17. First, Cortina notes how conceptual diagrams, which are intended to aid comprehension, often have the opposite effect because they do not represent the statistical model being tested. Specifically, he points to the importance of including an arrow between the moderator (Z) and the dependent variable (Y) because "the coefficient for the product must reflect rate of change in Y per unit increase in the product holding both of its components constant". Second, Cortina provides mathematical evidence for the fact that there is no moderator-predictor distinction. Therefore, he argues, researchers should clarify why it makes

sense to say that the effect of X on Y *depends on the level of Z* as opposed to the effect of Z on Y *depends on the level of X*. Finally, Cortina proposes restricted variance interaction (reasoning) as a potential tool that may help researchers move from a general notion regarding Z moderating the X-Y relationship to a variable-specific justification for a particular interaction pattern.

Part IX (Chaps. 18 and 19) reflects on another critical methodological challenge facing many IB researchers, namely that of endogeneity. Myles Shaver discusses Chap. 18, "Endogeneity in International Business Research" by Reeb, Sakakibara, and Mahmood (2012); his Commentary in Chap. 19 presents three points to complement the issues raised by the authors. Shaver starts by reminding us about what exactly causality is and how to establish causal identification. He illustrates the importance of paying attention to *both* guiding theory *and* alternative explanations (or theories) that would or could lead to the same relationship between the variables we study. Shaver points out that authors must take several steps toward causal identification and must view this as a process, rather than deploy a supposedly "simple" fix. He offers three specific observations (presented here in a different order): (1) the difficulty of establishing causal identification is linked to the *nature* of the data we collect; (2) causal identification is best established through a cumulative body of research and a plurality of approaches; and (3) establishing causal identification requires both well-crafted theories and well-crafted alternative theoretical mechanisms.

Part X (Chaps. 20, 21, and 22) addresses another important issue in quantitative IB research, namely that of common method variance (CMV). Harold Doty and Marina Astakhova discuss uncommon methods variance (UMV) in their Commentary on Chap. 20, "Common Method Variance in International Business Research" by Chang, Van Witteloostuijn, and Eden (2010). Their Commentary in Chap. 21 offers four guidelines that will help reviewers evaluate the extent to which CMV threatens the validity of a study's findings. These guidelines encourage reviewers (and authors) to ask critical questions related to: (1) the extent to which single source or self-report measures may be the most theoretically appropriate measurement approach in a particular study; (2) how the content of the constructs may help judge the potential for biased

results; (3) how likely it is that the observed correlations are biased given reported reliabilities; and (4) whether the larger nomological network would appear to make sense.

In a Further Reflection (Chap. 22) on their *JIBS* article, Van Witteloostuijn, Eden, and Chang reiterate the importance of making appropriate ex ante research design decisions in order to avoid or minimize such issues. They also provide compelling evidence for the continued importance and relevance of the single-respondent–one-shot survey design in many instances. Therefore, ex ante approaches to CMV issues are preferable to ex post remedies. The authors end by musing over the extent to which CMV issues may also apply to other research designs than single-respondent–one-shot survey designs, arguing for the importance of replication studies—particularly ones that utilize different research designs and methods.

The volume concludes with Parts XI and XII (Chaps. 23, 24, 25, 26, and 27), which focus on frontier methodological challenges in IB research. Part XI (Chaps. 23, 24, and 25) delves into the complexities of modelling the multilevel nature of IB phenomena. Robert Vandenberg discusses opportunities and challenges specific to multilevel research in his insightful Commentary on Chap. 23, "Multilevel Models in International Business Research" by Peterson, Arregle, and Martin (2012). Vandenberg in his Chap. 24 Commentary points to four "hidden" jewels in the article that may not be apparent to the reader: (1) not addressing cross-level direct effects; (2) not using the term cross-level when addressing how a level 2 variable may moderate the slopes of an X-Y relationship within each level 2 unit; (3) introducing the concept of cross-classified cases; and (4) centring. He also addresses two methodological advancements since the publication of the article; (a) ML structural equation modelling; and (b) incorporating more than just two levels into multilevel analysis. Vandenberg has very strong views on the statistical possibilities (or impossibility as he argues) of testing particular types of relationships, such as cross-level direct effects, and he also urges authors not to use the term "cross-level interaction" at all. More constructively, he urges scholars to pay due attention to cross-classified cases, that is, data that are not hierarchically nested, which is often the case in IB research (see also our earlier discussion under the heading *How to choose the level(s)*

of analysis?). Finally, he also points to the importance of centring in multilevel analysis. He concludes by discussing the two recent advances (at least to IB scholars) pertaining to complex multilevel structural equation modelling (MLSEM) and models with more than 2 levels.

Further reflecting on the scope for the use of multilevel models (MLMs) in IB research, one of the original authors (Martin) in Chap. 25 reviews conditions and solutions for the estimation of MLMs where the dependent variable is not continuous. He points to several powerful and well-documented software packages that allow for the estimation of such models, but cautions that guidelines for appropriate use of MLM (particularly with non-continuous dependent variables) are less well documented and that sample size requirements are generally more demanding. With regard to sample size requirements, Martin reminds us that most power simulation studies use a single predictor (typically at level 2), thus rendering true power analysis difficult. He ends by offering two suggestions that may help researchers overcome the issues resulting from small sample size: (1) using repeated measures at level 1 and (2) using bootstrapping and a Bayesian estimator leveraging Gibbs sampling to reduce the number of unique subjects required.

Cultural distance is one of the most commonly used proxies to assess the general difficulties firms will face when operating across borders. Part XII (Chaps. 26 and 27) concludes this research methods volume by providing insightful recommendations about how to conduct distance research in IB via a detailed analysis of the various measures for cultural distance available. Mark Peterson and Yulia Muratova comment on Chap. 26, "Conceptualizing and Measuring Distance in International Business Research: Recurring Questions and Best Practice Guidelines" by Beugelsdijk, Ambos, and Nell (2018). The Peterson-Muratova Commentary in Chap. 27 assesses the recommendation to use the Mahalanobis distance correction in the context of studies of distances between a single home or host country and multiple other countries and finds strong support for its importance. Peterson and Muratova then offer two additional recommendations: (1) occasionally, distance from a single reference country can be meaningfully estimated for some of the country's salient cultural or institutional characteristics; and (2) when estimates for a reference country require more than the data can provide,

the best course of action may be to study cultural and institutional characteristics as variables representing other countries with which it does business.

Having briefly pointed out some of the highlights of the other 26 chapters in this book, we now turn to our last goal for this chapter: providing a few recommendations for best practices in research methods, which we hope will be useful for IB researchers.

A Few Suggestions for the Road Ahead

This book is a first, serious attempt to bring together various strands of state-of-the-art thinking on research methods in international business. The volume is intended as a solid reference book for scholars, ranging from research master's students to senior academics, as they reflect on the best research methods approaches that can reasonably be adopted given the nature of the IB phenomena studied. Deploying the practices suggested in this book will go a long way towards improving the image of IB research as a field of academic inquiry at par methodologically with the more conventional subject areas in business schools.

IB as a field of research is rapidly moving towards maturity. The level of methodological sophistication with which many IB scholars now study cross-border phenomena makes it likely that the IB context will increasingly prove to be fertile ground for developing innovative new methodologies that may inform other disciplines. If this occurs, IB will evolve from being a net methods importer towards becoming an exporter in its own right.

We end this chapter with four methods-related recommendations for IB scholars and reviewers alike, which follow from our long experience with assessing the work of colleagues and from our own work being evaluated by our peers.

1. ***Rules of thumb and widely accepted conventions as to what constitutes an acceptable methodological approach or interpretation of an outcome should never substitute for independent scholarly judgement and plain common sense.***

Many IB scholars legitimately claim that their work is sometimes unfairly judged because of "methods policing" by reviewers and editors, who follow simple heuristics and standard rules that should perhaps not apply to the specific study being assessed. One case in point is that of multicollinearity. As stated by Lindner, Puck, and Verbeke (2019): "Research in IB is affected by prevailing myths about the effects of multicollinearity that hamper effective testing of hypotheses**…** Econometric texts and theory tend to use 'clean' examples (i.e., simple ones with only one problem being present at a time) to analytically make a point about the effect of the violation of an assumption on point estimates or their variance. Yet, IB research usually deals with complex relationships and many interrelated variables. To the best of our knowledge, there is no investigation into how specific econometric and data problems affect regression outcomes in such 'messy' cases where, among potential other issues, multicollinearity exists". Lindner et al. (2019) show that many published *JIBS* articles explicitly calculate Variance Inflation Factors (VIFs) to address multicollinearity, which in turn helps to determine what variables to include in regression models. In more than 10 percent of the articles studied, variables were dropped when high VIFs had been observed. Yet, the authors demonstrate that high VIFs may be an inappropriate guide to eliminate otherwise relevant parameters from regressions.

Conversely, sometimes problematic empirical studies are actually published and become heavily cited because they align with a stream of studies adopting a similar methodological approach and meet a number of supposed quality standards, again as a function of heuristics and conventions, which may be ill advised at best. For example, a large literature exists on the multinationality–performance linkage, which—as far as theory is concerned—builds upon concepts related to the entire historical trajectory of firms, starting from their first international expansion move. But most empirical studies in this realm actually build on cross-sectional data or panel data covering only a limited time span (e.g., 10 years), rather than the firms' actual histories. An unacceptable discrepancy can therefore be observed between theory and data in a large number of articles, including arti-

cles published in leading academic journals, where reviewers and editors accept poor work, based on the standards set in earlier poor work (Verbeke and Forootan 2012).

2. ***IB researchers should systematically be pushed to disclose their actual command of the data they use, and their knowledge of the economic actors supposedly represented by these data or affected by these data.***

This challenge is becoming increasingly important in an era of access to big data and higher sophistication in the information and communications technology sphere. More data and better technology do not necessarily lead IB researchers to have a better command of their data. A few years ago, one of the editors of this volume contacted the OECD because some data collected longitudinally and across countries on inward and outward FDI revealed potential inconsistencies. Upon investigating the discrepancies found, it appeared that these data were collected by national agencies, and that the data sources drawn from, as well as the methods to collect and collate the data, changed every year in some countries. The OECD itself had been making significant changes to its data aggregation methods over time (these were noted in footnotes under the data tables, in font 3 and therefore difficult to decipher). The recommendation of an OECD expert was therefore never to assess FDI trajectories over periods longer than 3–5 years. This expert stated: "comparisons within a country over more than 3–5 years will at best be like comparing apples and oranges; comparing across countries over time, will be like comparing apples and sports cars".

The above example is indicative of a major problem facing IB researchers. The main problem is not that national and international agencies change their methods and sources to collect data, sometimes in a non-transparent way. The problem is that many IB researchers are unaware (or choose to remain unaware) of this situation and thus overestimate the quality and consistency of their data. Especially in an era when scholars are pushed to analyse large databases and to deploy sophisticated statistical techniques, it would appear that many of them simply do not conduct any background investigation of their data. That is, they know little or nothing about the firms in their samples

and are ignorant of these firms' historical trajectories and the real-world meaning behind the evolution of the values of the parameters they study. In some cases, they do not even have a basic command of the language in which the data were collected or published. They just view it as their task to test hypotheses, but without any in-depth understanding of the subjects they are studying. After the empirical analyses have been performed, very few IB scholars confront the managers or owners of the firms they have analysed with the results and conclusions of their studies. As a result, entire bodies of completely speculative research come into existence and feed off each other, without due diligence.

The answer to the above is clearly for journal editors and reviewers to demand tangible proof that the authors of articles have a proper command of their data. It should be mandatory for editors to ask authors submitting manuscripts which due diligence measures were undertaken to ascertain the quality of the underlying data and the plausibility of the results found. In many cases, this may imply contacting data collection agencies as well as firms, both ex ante (at the time of formulating hypotheses and collecting data) and ex post (after the results of the empirical analysis have been computed.)

3. ***IB researchers should systematically consider the possibility of "combinatorial notions" being associated with particular outcomes, rather than assuming from the outset one key independent variable affects the focal outcome with the impact moderated by other variables.***

As one example, the fsQCA approach discussed earlier in this chapter examines combinations of factors linked to an outcome variable in "complex" situations or "rich contexts". Complex means that the various parameters considered work in concert to influence the outcome variable. Different combinations of these parameters can lead equifinally to the same outcome. What is perhaps more critical is that no individual parameter might in and by itself determine a particular outcome, in sharp contrast with conventional regression analysis where scholars typically identify a key independent variable.

As with any model, there might be alternative explanations. In fsQCA these are taken into account by considering all theory-anchored

alternative explanations prevailing in the literature. FsQCA does not use control variables, because it does not measure linear, causal relationships. One of the arguments for the use of fsQCA is precisely that it forces researchers to use a more experimental approach, that is, trying to control for causality ex ante through theory and research design, as opposed to ex post testing through the use of control variables. Thus, using fsQCA entails choosing carefully models informed by causal antecedents for a given phenomenon, as identified in the theory-based literature.

As an alternative to fsQCA, SEM allows for simultaneous equations with both multiple x's and y's (as well as moderators and mediators), whereas most OLS type studies do not allow this. The tendency to test regressions hierarchically with controls first, followed by independent variables, and then moderators, and finally the "full" model, which is interpreted only in terms of p-values in relation to hypothesis testing, may not be an optimal approach. Nielsen and Raswant (2018) discuss this issue with regard to controls and suggest that IB scholars should run models both with and without controls (i.e., a model with only x's and no controls and a full model with no controls), in order to tease out the actual effect(s) of the controls. Something similar could be done with particular independent variables that are thought (theoretically) to be interdependent rather than independent of each other. While this violates OLS assumptions, other estimation techniques can address this challenge.

4. ***Technology is a powerful aid in research, but IB scholars should strive for methodological parsimony.***

   As society transforms and is transformed by new technology, novel pathways materialize for IB researchers to collect and analyse data. New forms of data also become more readily available. As of 2019, the sheer number and sophistication of technological tools that can help scholars collect, analyse, and interpret data is daunting. And more of these tools are on their way, as this very chapter is being written.

   For qualitative and quantitative scholars alike, increasingly sophisticated and complex tools are becoming available and this trend is driven by innovation in technology. For instance, the spread of video and photographic

technology allows scholars to use images both as sources of data and as tools for data collection. In addition, the increasingly digital form of most data (either as audio or video files) provides new ways of accessing, developing, analysing, and interpreting data. With the Internet now available to some estimated 60 percent of the world's population, tables, charts, maps, and articles, in addition to audio and video files, can be easily shared across the globe. Social media platforms such as Facebook, LinkedIn, Twitter, and so on further link previously disparate people throughout the world. For IB researchers, digital tools and platforms offer tremendous opportunities to collect data—both primary and secondary—from diverse cross-country and cross-cultural settings.

Technological advances have also led to new ways of analysing data, with increasingly sophisticated techniques and tools available to both qualitative and quantitative researchers. For instance, various computer assisted qualitative data analysis (CAQDAS) software packages exist and with artificial intelligence (AI) on the rise, such programs are likely to be even more capable of sorting through enormous amounts of data in various formats (i.e., text, audio, visual) and drawing out coherent information for analytical purposes. By the same token, quantitative methods have experienced significant technological leaps forward with the coming of Big Data and a shift from analogue to digital storage and distributed processing (e.g., via cloud-based platforms). Statistical software capable of analysing such large and complex data is following suit, with many of the "traditional" packages now offering Big Data programs (e.g., R). As with qualitative software, AI and other innovative technologies may further enhance our abilities to access, process, and interpret increasingly larger and more complex cross-cultural datasets.

Finally, machine learning may lead to significant improvements in research methodology in IB as it holds the potential to assist both quantitative and qualitative research, and perhaps lead to more mixed-methods applications (see also the Commentary by Eisenhardt in Chap. 10 in this volume). Such advances in technology are likely to enable IB researchers to ask broader questions about IB phenomena that influence (and are influenced by) many if not all of us, and compare and contrast results across regions, countries, sub-cultures, and even individuals—over vast distances in both time and space. It may also lead to more narrow research questions that

seek to tease out micro-foundational issues pertaining to individual behaviour(s) within and across contexts.

However, such technological developments raise important issues about the way researchers collect, process, and publish data, and how they produce high-quality analyses. The diversity of software means that there is a need for standards for storing and exchanging data and analyses. Moreover, with more analytical (statistical or other) power comes the risk of drowning good research in technically sophisticated modelling exercises.

While the issue of responsible and ethical research is an important challenge in its own right, we argue that IB scholars should strive for methodological parsimony rather than technical sophistication when designing and carrying out their studies. It may be enticing to apply the newest tools or most complex methodologies in a study—particularly if junior scholars with such skills are involved—but an important caveat surrounds the trade-offs between "necessary" and "sufficient" methodological complexity.

IB researchers would be wise to remember that rigor in methodology does not equate to complexity any more than larger datasets (such as Big Data) can ensure more validity or reliability. To be sure, large datasets may increase power to detect certain phenomena but potentially at the risk of committing type I errors. Add to this the concerns about veracity stemming from noise in the data and scholars may be left with less than desirable outcomes.

Another example of how technology is potentially a double-edged sword is the increasing inclusion of graphical user interface (GUI) in many software packages (for instance, in most SEM software packages). While such graphical interfaces may aid the researcher, they do so at the risk of sometimes losing the underlying meaning behind the study (i.e., drawing a diagram with arrows between boxes and having the software write the underlying equations removes the researchers one more step from the data and its implications). By the same token, some software packages also allow for automatic removal of outliers or capitalization of chance by data driven modelling procedures such as modification indices.

IB researchers collecting, analysing, and interpreting data from nationally and culturally diverse settings should utilize technologically sophisti-

cated techniques when warranted (e.g., multilevel modelling of nested data). However, they must also avoid the trap of "showing off" newly developed methodologies in situations where these are not necessary. The old adage still holds true: *if it ain't broke, don't fix it.* Many often used, mainstream techniques, such as regression and ethnographic studies, still work well to address the majority of our IB research questions.

It is the responsibility of the researcher to select and apply the best-suited methodology within a given research setting. Replication is important, and results should never be attributable to a particular method. We strongly recommend that IB scholars consider parsimony over technical sophistication when making such choices. A short statement of justification of methodological choices, including selection criteria, is warranted; and in the early round submissions, it can be worthwhile to illustrate the utility of a particular (advanced) technique by comparing and reporting results with more/less parsimonious techniques.

Conclusions

"The values of honesty, trust, fairness, respect and responsibility in learning, teaching and research" (Bretag 2019)—that is what acting with integrity means. To act with integrity in our research requires that we use high-quality research methods that promote the "truth" and minimize error.

FFP (falsification, fabrication, and plagiarism) and QRPs (questionable research practices) are not consistent with research integrity and have serious negative consequences for the credibility of our scholarship. Researchers, we believe, make mistakes mostly because they do not really understand the nuances of using different research methods. There will always be some scholars who engage in research misconduct, and a far larger number who engage in QRPs, but we believe that the bulk of errors in how scholars use research methods is due to unfamiliarity with best practices.

This book is designed to help reduce unfamiliarity hazards by explaining and exploring several best practices in IB research methods. We hope that reading and working through the chapters in this book will enhance

research integrity in IB scholarship and serve as inspiration for interesting, high-quality IB research. We also hope that this collection may promote more discussion among IB scholars about the importance and utility of research methods in furthering our field. Only through innovation (in both theory and methodology) can international business scholarship grow and prosper.

Acknowledgements As authors and editors of this book, we "stand on the shoulders" of the scholars, teachers, and colleagues who have provided us with knowledge, skills, and support—not only while writing this book but throughout our academic careers. We trace our interest and expertise in research methods to the wonderful teachers, mentors, and co-authors who shared their knowledge and expertise with us. We therefore first owe a debt of gratitude to the many individuals who have helped to improve our understanding of research methods during our own careers.

We also want to thank everyone who helped us in writing this book. In particular, we thank Anne Hoekman, long-time Managing Editor of *JIBS*, who put together the *JIBS* dataset from which we drew our 11 best articles. We want to thank all the authors of the original *JIBS* articles, together with the authors who wrote the Commentaries and Further Reflections on each of the 11 original pieces. We thank our families also for listening to (and putting up with) us while we were working on this book. We also extend a special thanks to the founding committee members at the Research Methods Shared Interest Group (RM-SIG)—Agnieszka Chidlow, Stewart Miller, and Catherine Welch—for their valuable feedback on this chapter. Lastly, we thank Liz Barlow and Lucy Kidwell at Palgrave Springer for their helpful advice on this book project.

Notes

1. The nationality of authors was determined by the university affiliation at the time of publication. Thus, this number does not take into account the multicultural backgrounds of authors in the same country. Hence, this number is likely very conservative and the real increase potentially much higher.
2. The authors of each original article were invited to write a Further Reflections note on their original piece. Some authors chose to do this; others decided not to, inter alia, because they felt there was little new to report in substantive terms.

References

Aguinis, H., W.F. Cascio, and R.S. Ramani. 2017. Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies* 48 (6): 653–663.

Andersson, U., A. Cuervo-Cazurra, and B.B. Nielsen. 2014. From the editors: Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies* 45: 1063–1071.

Antonakis, J., and J. Dietz. 2011. Looking for validity or testing it? The perils of stepwise regression, extreme-scores analysis, heteroscedasticity, and measurement error. *Personality and Individual Differences* 50: 409–415.

Arregle, J.-L., L. Hebert, and P.W. Beamish. 2006. Mode of international entry: The advantages of multilevel methods. *Management International Review* 46: 597–618.

Banks, G.C., E.H. O'Boyle Jr., J.M. Pollack, C.D. White, J.H. Batchelor, C.E. Whelpley, K.A. Abston, A.A. Bennett, and C.L. Adkins. 2016a. Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management* 42 (1): 5–20.

Banks, G.P., S.G. Rogelberg, H.M. Woznyj, R.S. Landis, and D.E. Rupp. 2016b. Editorial: Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business Psychology* 31: 323–338.

Bedeian, A.G., S.G. Taylor, and A.N. Miller. 2010. Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education* 9 (4): 715–725.

Beugelsdijk, S., B. Ambos, and P.C. Nell. 2018. Conceptualizing and measuring distance in international business research: Recurring questions and best practice guidelines. *Journal of International Business Studies* 49 (9): 1113–1137.

Bouter, L.M., J. Tijdink, N. Axelsen, B.C. Martinson, and G. ter Riert. 2016. Ranking major and minor research misbehaviors: Results from a survey among participants of four World Conferences on Research Integrity. *Research Integrity and Peer Review* 1: 17.

Bretag, T. 2019. Academic integrity. *Oxford Research Encyclopedia, Business and Management*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190224851.013.147.

Brewer, P., and S. Venaik. 2014. The ecological fallacy in national culture research. *Organization Studies* 35 (7): 1063–1086.

Burgelman, R.A. 2011. Bridging history and reductionism: A key role for longitudinal qualitative research. *Journal of International Business Studies* 42 (5): 591–601.

Butler, N., H. Delaney, and S. Spoelstra. 2017. The gray zone: Questionable research practices in the business school. *Academy of Management Learning & Education* 16 (1): 94–108.

Chang, S.J., A. Van Witteloostuijn, and L. Eden. 2010. Common method variance in international business research. *Journal of International Business Studies* 41 (2): 178–184.

Crilly, D. 2011. Predicting stakeholder orientation in the multinational enterprise: A mid-range theory. *Journal of International Business Studies* 42 (5): 694–717.

Cuervo-Cazurra, A. 2012. Extending theory by analyzing developing country multinational companies: Solving the Goldilocks debate. *Global Strategy Journal* 2 (3): 153–167.

Cuervo-Cazurra, A., U. Andersson, M.Y. Brannen, B.B. Nielsen, and A.R. Reuber. 2016. Can I trust your findings? Ruling out alternative explanations in international business research. *Journal of International Business Studies* 47 (8): 881–897.

Davidov, E., B. Meuleman, J. Cieciuch, P. Schmidt, and J. Billiet. 2014. Measurement equivalence in cross-national research. *Annual Review of Sociology* 40: 55–75.

Davis, H.L., S.P. Douglas, and A.J. Silk. 1981. Measure unreliability: A hidden threat to cross-national marketing research? *Journal of Marketing* 45 (2): 98–109.

Eden, L. 2010. Letter from the Editor-in-Chief: Scientists behaving badly. *Journal of International Business Studies* 41 (4): 561–566.

Eden, L., K. Lund Dean, and P.A. Vaaler. 2018. *The ethical professor: A practical guide to research teaching and professional service*. New York: Routledge.

Edwards, M.A., and S. Roy. 2017. Academic research in the 21st century: Maintaining scientific inquiry in a climate of perverse incentives and hypercompetition. *Environmental Engineering Science* 34 (1): 51–61.

Estrin, S., K.E. Meyer, B.B. Nielsen, and S. Nielsen. 2016. Home country institutions and the internationalization of state owned enterprises: A cross-country analysis. *Journal of World Business* 51 (2): 294–307.

Estrin, S., B.B. Nielsen, and S. Nielsen. 2017. Emerging market multinational companies and internationalization: The role of home country urbanization. *Journal of International Management* 23 (3): 326–339.

Fang, F.C., R.G. Steen, and A. Casadevall. 2012. Misconduct accounts for the majority of retracted scientific publications. *PNAS* 109 (42): 17028–17033.

Fielding, A., and H. Goldstein. 2006. *Cross-classified and multiple membership structures in multilevel models: An introduction and review*. Research Report No. 791. University of Birmingham, Department of Education and Skills. ISBN, 1(84478797).

Goerzen, A., C.G. Asmussen, and B.B. Nielsen. 2013. Global cities and multinational enterprise location strategy. *Journal of International Business Studies* 44 (5): 427–450.

Hahn, E.D., and S.H. Ang. 2017. From the editors: New directions in the reporting of statistical results in the Journal of World Business. *Journal of World Business* 2 (52): 125–126.

Hall, J., and B.R. Martin. 2019. Towards a taxonomy of research misconduct: The case of business school research. *Research Policy* 48: 414–427.

He, X., L. Eden, and M. Hitt. 2016a. Shared governance: Institutional investors as a counterbalance to the state in state owned multinationals. *Journal of International Management* 22: 115–130.

———. 2016b. The renaissance of state owned multinationals. *Thunderbird International Business Review* 58 (2): 117–129.

Hofstede, G. 2001. *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations*. Thousand Oaks: Sage Publications.

House, R.J., P.J. Hanges, M. Javidan, P.W. Dorfman, and V. Gupta, eds. 2004. *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks: Sage Publications.

Hussigner, K., and M. Pellens. 2019. Guilt by association: How scientific misconduct harms prior collaborators. *Research Policy* 48: 516–530.

Jowell, R., C. Roberts, R. Fitzgerald, and G. Eva, eds. 2007. *Measuring attitudes cross-nationally: Lessons from the European Social Survey*. London: Sage Publications.

Lawrence, B.S. 1997. Perspective – The black box of organizational demography. *Organization Science* 8 (1): 1–22.

Lindner, T., J. Puck, and A. Verbeke. 2019. Myths about multicollinearity in International Business research: Identification, consequences, and remedies. *Journal of International Business Studies* (forthcoming).

Marschan-Piekkari, R., and C. Welch, eds. 2004. *Handbook of qualitative research methods for international business*. Cheltenham: Edward Elgar.

Meade, R.D., and R.W. Brislin. 1973. Controls in cross-cultural experimentation. *International Journal of Psychology* 8 (4): 231–238.

Meyer, K.E., A. Witteloostuijn, and S. Beugelsdijk. 2017. What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies* 48 (5): 535–551.

Michalek, A.M., A.D. Hutson, C.P. Wicher, and D.L. Trump. 2010. The costs and underappreciated consequences of research misconduct: A case study. *PLoS Medicine* 7(8): e1000318:1–3. www.plosmedicine.org

Mullen, M.R. 1995. Diagnosing measurement equivalence in cross-national research. *Journal of International Business Studies* 26 (3): 573–596.

Mullen, M.R., G.R. Milne, and P.M. Doney. 1995. An international marketing application of outlier analysis for structural equations: A methodological note. *Journal of International Marketing* 3 (1): 45–62.

Nielsen, B.B., and S. Nielsen. 2010. A multilevel approach to understanding the multinationality–performance relationship. In *The past, present and future of international business & management*, ed. Timothy Devinney, Torben Pedersen, and Laszlo Tihanyi, 527–557. Cheltenham: Emerald Group Publishing Limited.

———. 2011. The role of top management team international orientation in international strategic decision-making: The choice of foreign entry mode. *Journal of World Business* 46 (2): 185–193.

Nielsen, B.B., and A. Raswant. 2018. The selection, use, and reporting of control variables in international business research: A review and recommendations. *Journal of World Business* 53 (6): 958–968.

Nielsen, B.B., A. Chidlow, S. Miller, C. Welch, R. Aguzzoli, E. Gardner, M. Karafyllia, and D. Pegoraro. 2019. The evolution of methodology in international business research: The past, the present, and the future. *Journal of International Business Studies* (forthcoming).

Peterson, M.F., J.L. Arregle, and X. Martin. 2012. Multilevel models in international business research. *Journal of International Business Studies* 43 (5): 451–457.

Rabe-Hesketh, S., and A. Skrondal. 2004. *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Raton: Chapman and Hall/CRC.

Ramamurti, R. 2012. What is really different about emerging market multinationals? *Global Strategy Journal* 2 (1): 41–47.

Raudenbush, S.W., and A.S. Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*. Vol. 1. Thousand Oaks: Sage.

Reeb, D., M. Sakakibara, and I.P. Mahmood. 2012. Endogeneity in international business research. *Journal of International Business Studies* 43 (3): 211–218.

Research Methodology Shared Interest Group of the Academy of International Business. https://rmsig.aib.world/

Responsible Research in Business and Management. https://rrbm.network/position-paper/principles-of-responsible-science/. Accessed 23 Apr 2019.

Ryen, A., C. Truman, D.M. Mertens, and B. Humphries. 2000. Colonial methodology? Methodological challenges to cross-cultural projects collecting data by structured interviews. In *Research and inequality*, 220–235. London: UCL Press, Taylor and Francis Group.

Rynes, S.L., A.E. Colbert, and E.H. O'Boyle. 2018. When the "best available evidence" doesn't win: How doubts about science and scientists threaten the future of evidence-based management. *Journal of Management* 44 (8): 2995–3010.

Santangelo, G.D., and K.E. Meyer. 2011. Extending the internationalization process model: Increases and decreases of MNE commitment in emerging economies. *Journal of International Business Studies* 42 (7): 894–909. https://doi.org/10.1057/jibs.2011.25.

Schwartz, S.H. 1994. Are there universal aspects in the structure and contents of human values? *Journal of Social Issues* 50 (4): 19–45.

Sekaran, U. 1983. Methodological and theoretical issues and advancements in cross-cultural research. *Journal of International Business Studies* 14 (2): 61–73.

Snijders, T.A.B., and R.J. Bosker. 2011. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. 2nd ed. London: Sage Publishers.

Stern, A.M., A. Casadevall, R.G. Steen, and F.C. Fang. 2014. Financial costs and personal consequences of research misconduct resulting in retracted publications. *eLIFE*, elifesciences.org. https://doi.org/10.7554/eLife.02956.

Teagarden, M.B., M.A. Von Glinow, and K. Mellahi. 2018. Contextualizing international business research: Enhancing rigor and relevance. *Journal of World Business* 53 (3): 303–306.

Tihanyi, L., A.E. Ellstrand, C.M. Daily, and C.M. Dalton. 2000. Composition of the top management team and firm international diversification. *Journal of Management* 26 (6): 1157–1177.

Tourish, D., and R. Craig. 2018. Research misconduct in business and management studies: Causes, consequences, and possible remedies. *Journal of Management Inquiry*. https://doi.org/10.1177/1056492618792621.

Verbeke, A., and M.Z. Forootan. 2012. How good are multinationality–performance (M-P) empirical studies? *Global Strategy Journal* 2 (4): 332–344.

Welch, C., R. Piekkari, E. Plakoyiannaki, and E. Paavilainen-Mäntymäki. 2011. Theorising from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies* 42 (5): 740–762.

Yang, Z., X. Wang, and C. Su. 2006. A review of research methodologies in international business. *International Business Review* 15 (6): 601–617.

Part II

Reproducibility and Replicability

2

Science's Reproducibility and Replicability Crisis: International Business Is Not Immune

Herman Aguinis, Wayne F. Cascio, and Ravi S. Ramani

Introduction

International business (IB) and many other management and organization studies' disciplines are currently immersed in an important debate regarding the credibility and usefulness of the scholarly knowledge that is produced (Cuervo-Cazurra et al. 2016; Davis 2015; George 2014; Meyer et al. 2017). A critical issue in this debate is the lack of ability to reproduce and replicate results described in published articles (Bakker et al. 2012; Bergh et al. 2017; Cuervo-Cazurra et al. 2016; Ioannidis 2005; Open Science

H. Aguinis (✉)
George Washington University, Washington, DC, USA
e-mail: haguinis@gwu.edu

W. F. Cascio
University of Colorado Denver, Denver, CO, USA
e-mail: wayne.cascio@ucdenver.edu

R. S. Ramani
Purdue University Northwest, Hammond, IN, USA
e-mail: RRAMANI@PNW.EDU

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_2

Collaboration 2015). *Reproducibility* means that someone other than a published study's authors is able to obtain the same results using the authors' own data, whereas *replicability* means that someone other than a published study's authors is able to obtain substantially similar results by applying the same steps in a different context and with different data. Clearly, it is difficult to make the case that research results are credible and useful if they are irreproducible and not replicable. Unfortunately, there is a proliferation of evidence indicating that lack of reproducibility and replicability are quite pervasive (e.g., Banks et al. 2016b; Cortina et al. 2017a; Cuervo-Cazurra et al. 2016; Ioannidis 2005; Open Science Collaboration 2015; Schwab and Starbuck 2017). Accordingly, as noted by Verbeke, Von Glinow, and Luo, "… the IB discipline faces the challenges of remaining at par with the methodological standards in adjacent fields for validity, reliability, replicability and generalizability" (Verbeke et al. 2017: 6). In short, IB is not immune to science's reproducibility and replicability crisis.

We argue that concerns about lack of reproducibility and replicability are actually not entirely surprising because of current methodological practices that enhance *systematic capitalization on chance.* Systematic capitalization on chance occurs when a researcher searches for a maximally predictive statistical model based on a particular dataset, and it typically involves several trial-and-error steps that are rarely disclosed in published articles. Currently, there is tremendous pressure to publish in the so-called top journals because the number of such publications has an important impact on faculty performance evaluations and rewards, including promotion and tenure decisions (Aguinis et al. 2014; Butler et al. 2017; Nosek et al. 2012). Thus researchers are strongly motivated to produce manuscripts that are more likely to be accepted for publication. This means submitting manuscripts that report tests of hypotheses that are statistically significant and "more highly" significant, models that fit the data as well as possible, and effect sizes that are as large as possible (Meyer et al. 2017). To paraphrase Friedman and Sunder (1994: 85), many researchers "torture the data until they confess" that effects are statistically significant, large, and supportive of favoured hypotheses and models. Each of these outcomes – which together are more likely to produce the desired result of a successful publication – can be reached more easily by systematically capitalizing on chance.

Researchers today have more "degrees of freedom" regarding methodological choices than ever (Freese 2007). Many of these degrees of

freedom involve practices that enhance capitalization on chance and improve the probability of successful publication. For example, researchers may include or delete outliers from a manuscript depending on which course of action results in a larger effect-size estimate (Aguinis et al. 2013). As a second illustration, researchers may capitalize on chance by selecting a particular configuration of control variables after analysing the impact of several groups of control variables on results and selecting the final set based on which configuration results in better fit indices for a favoured model (Bernerth and Aguinis 2016).

We emphasize that our focus on systematic capitalization on chance is different from *unsystematic capitalization of chance*, which is due to random fluctuations in any given sample drawn from a population. Unsystematic capitalization on chance is a known phenomenon and part of all inferential statistical tests. Specifically, the goal of inferential statistics is to maximize the predictive power of a model based on the data available by minimizing errors in prediction using sample scores (Cascio and Aguinis 2005). For example, ordinary least squares (OLS) regression, which is one of the most frequently used data-analytic approaches in IB and other fields (e.g., Aguinis et al. 2009a; Boellis et al. 2016; Fitzsimmons et al. 2017; Fung et al. 2016), minimizes the sum of the squared differences between fitted values and observed values. Unsystematic capitalization on chance is addressed by conducting inferential tests of the parameter estimates that include their standard errors, thereby providing information about the precision in the estimation process (i.e., larger sample sizes are associated with greater precision and smaller standard errors). Most articles in IB research include information on sample size and standard errors, which allows consumers of research to independently evaluate the accuracy of the estimation process and the meaning of results for theory and practice, thereby accounting for unsystematic capitalization on chance.[1]

Next, we describe several common practices that enhance systematic capitalization on chance and illustrate these practices using articles published in *Journal of International Business Studies* (JIBS). Because each of the issues we discuss is so pervasive, we do not "name names." We do not believe it would be helpful or constructive to point fingers at particular authors. However, we mention variable names and the overall substantive context of each study so that the methodological issues we discuss are

directly and specifically relevant for an IB readership. Then, we offer best-practice recommendations on how to minimize capitalization on chance in future IB research. Similar to previously published JIBS guest editorials (e.g., Andersson et al. 2014; Chang et al. 2010; Meyer et al. 2017; Reeb et al. 2012), these recommendations serve as resources for researchers, including doctoral students and their training, as well as for journal editors and reviewers evaluating manuscript submissions.

Common Methodological Practices That Enhance Systematic Capitalization on Chance

In this section, we discuss five common methodological practices that enhance systematic capitalization on chance: (1) selection of variables to include in a model, (2) use of control variables, (3) handling of outliers, (4) reporting of *p* values, and (5) hypothesizing after results are known (HARKing). We describe each of these issues and elaborate on how they lead to lack of reproducibility and replicability.

Selection of Variables to Include in a Model

The selection of variables to include in a model encompasses both the choice of variables to include, as well as the specification of the nature of the relations among these variables. Rapid advances in computational methodologies have allowed researchers to analyse increasingly larger amounts of data without much additional effort or cost (Simmons et al. 2011). Within the field of IB in particular, researchers routinely deal with "Big Data," that is, large amounts of information stored in archival datasets (Harlow and Oswald 2016). Because these datasets were not collected directly in response to a particular research question, they contain many variables that can be restructured to produce "favourable" results (i.e., better fit estimates, larger effect-size estimates) (Chen and Wojcik 2016). For example, consider the case of firm performance, which is one of the most frequently measured constructs in IB. As Richard, Devinney, Yip, and Johnson (2009) noted, firm performance can be defined and

assessed in terms of objective measures (e.g., shareholder returns, Tobin's *q*), and subjective measures (e.g., reputation, comparative ranking of firms). The choice of which firm-performance measure is examined should be driven by theory, and there should be a clear justification for why a particular measure was used, given the aims of the study (Richard et al. 2009).

Three recent articles published in JIBS have used the following measures of firm performance: (Study 1) increased reputation, overall performance, increased number of new products and customers, and enhanced product quality; (Study 2) return on assets; and (Study 3) return on equity, market-to-book ratio of assets, sales efficiency, and corporate risk-taking. Of these three studies, two did not provide any explanation or rationale for why they used those specific measures of firm performance, and the third cited "prior research" without providing any references or arguments in support of this particular choice. A healthily skeptical readership cannot judge if the firm-performance measures used in these studies were chosen because they aligned with the theories the researchers were testing, or because these measures produced outcomes that supported the favoured hypotheses. Moreover, it is not possible to ascertain if, initially, several measures of firm performance were considered, but only those that produced the most favourable results were retained in the published article.

Systematic capitalization on chance in terms of which variables are included in a predictive model, and how this final set of variables is chosen, has a direct detrimental impact of future efforts to reproduce and replicate substantive results. Almost 25 years ago, MacCallum, Roznowski, and Necowitz (1992) reported that researchers were making post-hoc modifications to improve the fit of models by utilizing results provided by the data-analytical software. MacCallum et al. (1992: 491) noted that this process of re-specifying models based on the data was "inherently susceptible to capitalization on chance" because the modifications were driven not by substantive reasons, but by the peculiarities of the dataset itself. Despite calls for a more thoughtful approach to the use and reporting of these modifications (e.g., Bentler 2007; Hurley et al. 1997), recent reviews show that they are still widely used, but rarely reported (Banks et al. 2016a; Cortina et al. 2017a; Sijtsma 2016). For example, a recently

published article in JIBS reported "relaxing" 35 of 486 constraints, including those associated with measurement error terms, until the model reached an acceptable fit. The article does not include any information on which specific paths were changed or any theory or measurement rationale for each of these "improvements" other than the goal of achieving a superior model fit. Given the popularity of data-analytical approaches such as structural equation modeling in research reported in JIBS (e.g., Funk et al. 2010; Lisak et al. 2016), we suspect that there are many other instances where researchers systematically capitalize on chance by making such modifications until an optimally fitting model is found – without necessarily reporting which paths were added or deleted from the original model, and why.

Use of Control Variables

Statistical controls are variables considered to be extraneous (i.e., noncentral) to the hypotheses being tested but that could provide alternative explanations for results. Control variables are used very frequently in management and organization studies (Becker 2005; Carlson and Wu 2012; Spector and Brannick 2011). For example, control variables are used by entering them in a hierarchical manner when conducting multiple regression analyses, under the presumption that they eliminate contamination between the predictor and outcome variables (Bernerth and Aguinis 2016). However, the assumptions and theoretical rationale underlying the use of control variables, namely, that including them provides a "truer" test of relations and that the controls used are measured reliably, are seldom tested (Bernerth and Aguinis 2016). Researchers rarely make explicit the reasons why certain variables (and not others) were chosen as controls (Becker 2005; Spector and Brannick 2011). Finally, control variables reduce the statistical power of the test and the variance associated with the criterion that can potentially be explained by substantive variables (Breaugh 2008), thereby increasing the chance that the results obtained are an artefact of the choice of control variables used (Bernerth and Aguinis 2016). Control variables therefore increase systematic capitalization on chance as researchers test several models,

including and excluding controls piecemeal until they obtain a desired result (Banks et al. 2016a; Bernerth and Aguinis 2016; Simmons et al. 2011). As noted by Cuervo-Cazurra et al. (2016: 894), "without specific knowledge about which controls were included, how they were measured and where they come from, replication is impossible".

Systematic capitalization on chance regarding the use of control variables seems pervasive in IB research. For example, four recent studies published in JIBS included the following sets of control variables: (Study 1) retained earnings scaled by the book value of assets, the ratio of shareholders' equity to the book value of assets, the natural logarithm of the ratio of current year sales revenue to prior year sales, and an indicator variable denoting the incidence of share repurchases; (Study 2) the natural log of a firm's book value of tangible assets per employee and the log number of employees; (Study 3) gender, age, job rank, exposure to female managers, and organizational sector; and (Study 4) organizational tenure, tenure with supervisor, group size, and country affiliation. Of these four studies, two did not provide any explanation or rationale for the authors' choice to include those specific control variables or information on any control variables that were initially included but later excluded. The authors of the other two studies justified their choices by citing "past research" examining the impact of the same control variables. But, readers have no way of knowing whether the control variables had a conceptual justification, or whether they were added in a post-hoc manner after much trial and error involving several potential controls, and the final set was chosen because it improved model fit or provided better results in support of the favoured hypotheses.

Handling of Outliers

Outliers are "data points that deviate markedly from others" (Aguinis et al. 2013: 270), and are commonly found in management and organization studies (Hunter and Schmidt 2015; Rousseeuw and Leroy 2003). Outliers are a challenge because they can substantially affect results obtained when testing hypotheses (Bobko 2001; Orr et al. 1991). Because of their outsized influence, the management of outliers presents an

opportunity for researchers to systematically capitalize on chance when analysing data, often in the direction of supporting their hypotheses (Cortina 2002). However, many researchers routinely fail to disclose whether they tested for outliers within their datasets, whether any outliers were identified, the type of outliers found, and the rationale behind choosing to include or exclude outliers from analyses (Aguinis et al. 2013).

Recently published articles in JIBS suggest the presence of systematic capitalization on chance regarding the management of outliers. For example, reported practices include winsorizing firm-level variables at the 5% level to account for outliers, trimming the sample by excluding observations at the top and bottom one percentile of variables, and removing an outlier based on studentised residuals and Cook's D.[2] In none of these cases did the authors define the type of outlier they were addressing. Specifically, error outliers (i.e., data points that lie at a distance from other data points), interesting outliers (i.e., non-error data points that lie at a distance from other data points and may contain valuable or unexpected knowledge), or influential outliers (i.e., non-error data points that lie at a distance from other data points, are not error or interesting outliers, and also affect substantive conclusions). In addition, in none of these published articles did the authors take appropriate steps such as correcting the data for error outliers and reporting the results with and without outliers (Aguinis et al. 2013). Therefore by not providing clear and detailed reporting of the manner in which they addressed the issue of outliers, it is virtually impossible to reproduce and replicate substantive conclusions.

Reporting of *p* Values

Another issue that involves systematic capitalization on chance refers to the reporting of *p* values associated with tests of significance. Despite its many flaws, null hypothesis significance testing (NHST) continues to be the choice of researchers in management and organization studies (Bettis et al. 2016; Meyer et al. 2017). In NHST, the tenability of a null hypothesis (i.e., no effect or relation) is primarily judged based on the observed *p* value associated with the test of the hypothesis, and values smaller than

0.05 are often judged as providing sufficient evidence to reject it (Bettis et al. 2016; Goldfarb and King 2016). Of the many problems associated with this interpretation of *p* values, the most pernicious is that it motivates researchers to engage in a practice called "*p*-hacking" and to report "crippled" *p* values (see below) (Aguinis et al. 2010; Banks et al. 2016b). For example, consider a researcher who interprets $p = 0.0499$ as sufficient evidence for rejecting the null hypothesis, and $p = 0.0510$ as evidence that the null hypothesis should be retained, and believes that journals are more likely to look favourably on rejected null hypotheses. This researcher will be highly motivated to "*p*-hack," that is, find some way, such as using control variables or eliminating outliers, to reduce the *p* value below the 0.05 threshold (Aguinis et al. 2010, Goldfarb and King 2016; Starbuck 2016; Waldman and Lilienfeld 2016). Similarly, this researcher will be motivated to report *p* values using cutoffs (e.g., $p < 0.05$), rather that report the actual *p* value (0.0510). Using this cutoff not only "cripples" the amount of information conveyed by the statistic (Aguinis et al. 2009b), but also allows the researcher to claim that his or her hypothesis was supported (Aguinis et al. 2010).

Many of the aforementioned practices regarding the reporting of *p* values are commonly found in articles published in JIBS. For example, recent studies in JIBS reported *p* values by using cutoffs instead of reporting actual *p* values, using multiple *p* value cutoffs within the same article, and using the term "marginally significant" to indicate $p < 0.10$. In classical hypothesis testing, conventional Type 1 error probabilities are $p < 0.05$ or 0.01. There are situations where a higher Type 1 error probability, such as $p < 0.10$, might be justified (Cascio and Zedeck 1983), but it is the responsibility of the researcher to provide such justification explicitly (Aguinis et al. 2010). In classical hypothesis testing, results either are or are not significant; there is no such thing as "marginally significant" results. The examples regarding the use of control variables and outliers provided above, along with evidence from other fields, such as strategic management (Bettis et al. 2016; Goldfarb and King 2016) and psychology (Bakker and Wicherts 2011; Nuijten et al. 2015) suggest the existence of published articles in which researchers exercised their "degrees of freedom" to systematically manipulate the data to obtain a significant (i.e., $p < 0.05$) result. Engaging in these practices increases systematic

capitalization on chance and diminishes the probability that results will be reproducible and replicable.

Hypothesizing After Results Are Known (HARKing)

Hypothesizing after results are known (HARKing) occurs when researchers retroactively include or exclude hypotheses from their study after analysing the data, that is, post-hoc hypotheses presented as a-priori hypotheses, *without acknowledging having done so* (Kerr 1998). A key issue regarding HARKing is lack of transparency. Specifically, epistemological approaches other than the pervasive positivistic model, which has become dominant in management and related fields since before World War II (Cortina et al. 2017b), are indeed useful and even necessary. For example, inductive and abductive approaches can lead to important theory advancements and discoveries (Bamberger and Ang 2016; Fisher and Aguinis 2017; Hollenbeck and Wright 2016). So, we are not advocating a rigid adherence to a positivistic approach, but rather, methodological plurality that is fully transparent so that results can be reproduced and replicated.

While primary-level and meta-analysis estimates based on self-reports indicate that 30–40% of researchers engage in HARKing, the number is likely higher because only a minority of researchers are likely to admit openly that they engaged in this practice (Banks et al. 2016a; Bedeian et al. 2010; Fanelli 2009). Consider the study by John, Loewenstein, and Prelec (2012), who surveyed 2155 academic psychologists regarding nine questionable research practices, including "reporting an unexpected finding as having been predicted from the start." John et al. (2012) asked these researchers (a) whether they had engaged in those practices (self-admission rate), (b) the percentage of other psychologists who had engaged in those practices (prevalence estimate), and (c) among those psychologists who had, the percentage that would admit to having done so (admission estimate). For this particular question addressing HARKing, the self-admission rate was about 30%, but the prevalence rate was about 50%, and the admission estimate was about 90%. More recently, O'Boyle, Banks, and Gonzalez-Mule (2017) examined doctoral dissertations and

the subsequent academic journal articles that they spawned. Their results revealed that the ratio of supported versus non-supported hypotheses was roughly 2 to 1. That is, somewhere between dissertation defense and published journal article, authors chose, altered, or introduced hypotheses after examining their data, likely to enhance the probability of publication (Bedeian et al. 2010; Edwards and Berry 2010; Starbuck 2016).

Even more worrisome is that many instances of HARKing are driven and even encouraged by reviewers and editors as part of the peer-review process (Banks et al. 2016a; Bedeian et al. 2010). In fact, Bosco, Aguinis, Field, Pierce, and Dalton (2016) conducted a survey of authors who had published in *Journal of Applied Psychology* and *Personnel Psychology* and found that 21% reported that at least one hypothesis change had occurred as a result of the review process. Because HARKing involves researchers fabricating or altering hypotheses based on the specific peculiarities of their datasets and not openly and honestly reporting so, it represents a particularly blatant instance of systematic capitalization on chance.

To illustrate the aforementioned discussion, we reviewed all articles published in JIBS in 2016 that proposed and quantitatively tested hypotheses. Let us be clear: our intentions are not to disparage any of the researchers or studies we examined, but simply to highlight trends. Across 30 studies published in JIBS in 2016 that met our criteria, researchers proposed 137 hypotheses, of which 115 (84%) received complete or partial support, and only 22 (16%) were not supported. Based on these results, it seems that researchers are almost five-times more likely to find support for their favoured hypotheses than they are to reject them. While not definitive, these results, combined with known self-reports of researchers admitting to HARKing, are a "smoking gun" (Bosco et al. 2016) that hints at the existence of HARKing in IB research.

Strategies to Minimize Capitalization on Chance

Meta-analysis seems to be a possible solution to understand whether a particular body of work has been subjected to capitalization on chance because it allows researchers to account for variables that create fluctua-

tions in the observed estimates of effect sizes (Hunter and Schmidt 2015). Because meta-analysis can correct for the effects of methodological and statistical artefacts, such as sampling error and measurement error, it has become a popular methodological approach in IB research (e.g., Fischer and Mansell 2009; Stahl et al. 2010; van Essen et al. 2012). However, meta-analysis only corrects for unsystematic capitalization on chance and not for systematic capitalization on chance. As noted by Eysenck almost 40 years ago: "garbage in, garbage out is a well-known axiom of computer specialists; it applies here [for meta-analysis] with equal force" (Eysenck 1978: 517). In other words, if effect-size estimates in primary-level studies are upwardly biased due to systematic capitalization on chance, accumulating all of those estimates will lead to a meta-analytic summary effect that will be similarly biased. Thus even if the estimated parameters are used to create distributions (i.e., funnel plots) (Dalton et al. 2012; Macaskill et al. 2001), systematic capitalization on chance biases the entire distribution. In short, meta-analysis is not a solution to address systematic capitalization on chance and its biasing effects on results and substantive conclusions.

Cross-validation is another strategy that could potentially be used to minimize the effects of capitalization on chance, but it also addresses its unsystematic and not its systematic variety. Specifically, ρ_c, an estimate of cross-validity in the population, refers to whether parameter estimates (usually regression coefficients) derived from one sample can predict outcomes to the same degree in the population as a whole or in other samples drawn from the same population. If cross-validity is low, the use of assessment tools and prediction systems derived from one sample may not be appropriate in other samples from the same population. Cascio and Aguinis (2005) provided a detailed discussion of various approaches to cross-validation and recommended estimating the cross-validity in the population (i.e., ρ_c) by adjusting the sample-based multiple correlation coefficient (R) by a function of sample size (N) and the number of predictors (k). It is important to note that what most computer outputs label "adjusted R^2" is only an intermediate step in computing the cross-validity in the population. Adjusted R^2 does not address the issue of prediction optimization based on the capitalization on chance factors in the original sample and, therefore, underestimates the shrinkage (i.e., amount by

which observed values were overestimated). Based on a careful review of the relevant literature, Cascio and Aguinis (2005) suggested appropriate formulas for estimating cross-validity in the population. Next, we offer suggestions for how to minimize systematic capitalization on chance specifically regarding each of the first five issues we mentioned earlier.

Issue #1 is the selection of variables in models. To improve reproducibility and replicability, researchers must clearly report the rationale behind the decision rules used in determining the sample-size and data-collection procedures, and report all the variables that they have considered (Simmons et al. 2011). If a construct can be assessed using several measures available (e.g., firm performance), researchers should justify their choice in light of theoretical considerations and the aims of their study (Richard et al. 2009). When making modifications to models, researchers should consider sample size, as modifications made to models drawing on small samples are likely to yield larger and more idiosyncratic results (MacCallum et al. 1992). Because each modification made to a model increases the fit of the model to the data in hand and decreases replicability (MacCallum et al. 1992), researchers should explicitly report all modifications made to their models, the theoretical rationale for the modifications, and the fit statistics for each model tested (Credé and Harms 2015; MacCallum et al. 1992).

Issues #2 and #3 relate to the use of control variables and the handling of outliers. Choosing which variables to use as controls or which data points to include or exclude from the analyses offers researchers an opportunity to systematically capitalize on chance. To minimize this, researchers should provide a theoretical justification for the choice of each control variable, along with evidence of prior empirical work showing a relationship between the proposed control and the focal variable. They should explain why the control variable is integral to the model they propose to test, and offer evidence regarding the reliability of the control variable (Bernerth and Aguinis 2016). When reporting results, researchers should provide descriptive statistics for all control variables, as well as reporting results with and without control variables (Aguinis and Vandenberg 2014; Becker 2005; Bernerth and Aguinis 2016). Regarding outliers, researchers should provide evidence showing that they tested for outliers in their datasets. They should specify the rules used to identify and classify outliers as error, interesting, or influential, and disclose whether

influential outliers affect model fit or prediction. Finally, they should test their models using robust approaches (e.g., absolute deviation) and report results with and without outliers (Aguinis et al. 2013).

Issue #4 is the reporting of *p* values. Relying on arbitrary *p* values (such as 0.05) to guide decisions motivates researchers to engage in "*p*-hacking," report "crippled" results, and conflate statistical and practical significance (Aguinis et al. 2010). To counter these deleterious effects, researchers should formally state the α level used to evaluate their hypotheses given the relative seriousness of making a Type I (probability of wrongly rejecting the null hypothesis) versus Type II (probability of mistakenly retaining the null hypothesis) error; justify the use of multiple cutoffs within the same paper; report complete *p* values to the second decimal place; not use terms such as "marginally significant" or "very significant" when referring to *p* values; and discuss the practical significance of their results in terms of the context of their study (Aguinis et al. 2010).

Lastly, we examined how researchers might systematically capitalize on chance through HARKing by creating and reporting hypotheses after analysing their data, either of their own volition, or as directed to by reviewers, and not describing hypotheses as being post-hoc in an open and honest manner. To counter this practice, researchers should conduct more studies using strong inference testing and report results of post-hoc hypotheses in a separate section from a-priori hypotheses (Banks et al. 2016a; Bosco et al. 2016; Hollenbeck and Wright 2016). In addition, influential and highly visible journals like JIBS can play a prominent role in countering this practice by encouraging more replication studies, promoting inductive and abductive research (Fisher and Aguinis 2017), and using study registries where authors post the details of their proposed research before collecting and analysing the data (Aguinis and Vandenberg 2014; Bosco et al. 2016; Kepes and McDaniel 2013).

Conclusions

A manuscript is more likely to be accepted for publication if results are statistically significant, effect-size estimates are large, and hypotheses and models are supported. So, consciously or not, it is in the best interests of

researchers to achieve these outcomes, and this is facilitated by engaging in methodological practices that systematically capitalize on chance, which, in turn, lead to lack of reproducibility and replicability. Irreproducible and non-replicable research results threaten the credibility, usefulness, and very foundation of all scientific fields; IB is certainly not immune.

Our intention in this editorial is not to point fingers at authors, journal editors, or reviewers. Rather, we believe that there are systemic issues that we must tackle collectively because they are the result of multiple causes operating at different levels of analysis. They include, among other factors, author motivation, methodological training (or lack thereof) of authors and reviewers, the rapid progress of methodological advancements, the availability of large, archival datasets, the low cost of computing tools, increased competition for journal space, pressures on universities to produce increasingly high levels of research output, and university promotion and tenure systems that encourage publishing as many articles as possible in the so-called top journals.

We addressed five admittedly selective issues that are particularly prone to being affected by systematic capitalization on chance. Some of the issues we discussed are not new and have already been noted in the methodological literature and also in the substantive literature in IB (e.g., Cascio 2012). We also offered suggestions on how to minimize the detrimental effects of capitalization on chance. But, realistically, even if researchers are aware of how to do things right, the issue of context (i.e., reward systems, manuscript-review processes) will remain as powerful hurdles. Thus we believe that a critical and necessary step is to enforce good methodological practices through the review process and also journal policies – because these are actions within the purview of journals. For example, Verbeke et al. offered guidelines for reviewers, including being "promoters of good methods" (Verbeke et al. 2017: 6) and the *Journal of Management* has recently included the following item on its reviewer-evaluation form: "To ensure that all papers have at least one reviewer with deep knowledge of the methods used, given your expertise in the statistical methods used in this paper, please indicate your comfort/confidence in your ability to rigorously evaluate the results reported: (Very uncomfortable, some discomfort, comfortable, confident, very confident, not applicable)" (Wright 2016).

As an actionable implication of our discussion, we offer the following modest proposal. Our recommendation is to include additional items on the manuscript-submission form such that authors acknowledge, for example, whether hypotheses were created retroactively after examining the results. Similar items can be included on the manuscript-submission form regarding the selection of variables in a model, handling of control variables and outliers, and other methodological choices and judgment calls that capitalize on chance systematically. Clearly, not all methodological details can be included in a manuscript itself due to page limitations, and this is why some journals have chosen to reduce the font size of the Method section (Cortina et al. 2017b). However, given that many journals allow authors to submit a supplemental file to be posted online, together with any published article, page limitations as a reason for not including sufficient detail about methodological procedures are no longer a valid constraint.

In closing, we believe that the motivation *not* to engage in systematic capitalization on chance needs to be greater than the motivation to engage in such practices. Hopefully, our article will provide a small step in this direction. One thing is clear, however: Lack of reproducibility and replicability, retractions, and negative effects on the credibility and usefulness of our research are unlikely to improve if we do not take proactive and tangible actions to implement a change in course.

Acknowledgements We thank Alain Verbeke and two *Journal of International Business Studies* anonymous reviewers for their highly constructive feedback that allowed us to improve our manuscript.

Notes

1. As noted by an anonymous reviewer, multilevel modeling is as susceptible to capitalization on chance as other methods, including OLS regression. Although the existence of a dependent data structure allows multilevel modeling to produce more accurate standard errors compared to OLS regression (Aguinis and Culpepper 2015), this is an improvement regarding unsystematic but not systematic capitalization on chance.

2. These are different ways to "manage outliers." Winsorization involves transforming extreme values to a specified percentile of the data (e.g., a 90th percentile Winsorization would transform all the data below the 5th percentile to the 5th percentile, and all the data above the 95th percentile would be set at the 95th percentile). Studentised residuals are computed by dividing a residual by an estimate of its standard deviation, and Cook's D measures the effect of deleting a given observation.

References

Aguinis, H., and S.A. Culpepper. 2015. An expanded decision making procedure for examining cross-level interaction effects with multilevel modeling. *Organizational Research Methods* 18 (2): 155–176.

Aguinis, H., and R.J. Vandenberg. 2014. An ounce of prevention is worth a pound of cure: Improving research quality before data collection. *Annual Review of Organizational Psychology and Organizational Behavior* 1 (1): 569–595.

Aguinis, H., C.A. Pierce, F.A. Bosco, and I.S. Muslin. 2009a. First decade of *Organizational Research Methods:* Trends in design, measurement, and data-analysis topics. *Organizational Research Methods* 12 (1): 69–112.

Aguinis, H., C.A. Pierce, and S.A. Culpepper. 2009b. Scale coarseness as a methodological artifact: Correcting correlation coefficients attenuated from using coarse scales. *Organizational Research Methods* 12 (4): 623–652.

Aguinis, H., S. Werner, J.L. Abbott, C. Angert, J.H. Park, and D. Kohlhausen. 2010. Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods* 13 (3): 515–539.

Aguinis, H., R.K. Gottfredson, and H. Joo. 2013. Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods* 16 (2): 270–301.

Aguinis, H., D.L. Shapiro, E. Antonacopoulou, and T.G. Cummings. 2014. Scholarly impact: A pluralist conceptualization. *Academy of Management Learning & Education* 13 (4): 623–639.

Andersson, U., A. Cuervo-Cazurra, and B.B. Nielsen. 2014. From the editors: Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies* 45 (9): 1063–1071.

Bakker, M., and J.M. Wicherts. 2011. The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods* 43 (3): 666–678.

Bakker, M., A. van Dijk, and J.M. Wicherts. 2012. The rules of the game called psychological science. *Perspectives on Psychological Science* 7 (6): 543–554.

Bamberger, P., and S. Ang. 2016. The quantitative discovery: What is it and how to get it published. *Academy of Management Discoveries* 2 (1): 1–6.

Banks, G.C., E.H. O'Boyle Jr., et al. 2016a. Questions about questionable research practices in the field of management: A guest commentary. *Journal of Management* 42 (1): 5–20.

Banks, G.C., S.G. Rogelberg, H.M. Woznyj, R.S. Landis, and D.E. Rupp. 2016b. Evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology* 31 (3): 323–338.

Becker, T.E. 2005. Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods* 8 (3): 274–289.

Bedeian, A.G., S.G. Taylor, and A.N. Miller. 2010. Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education* 9 (4): 715–725.

Bentler, P.M. 2007. On tests and indices for evaluating structural models. *Personality and Individual Differences* 42 (5): 825–829.

Bergh, D.D., B. Sharp, and M. Li. 2017. Tests for identifying "red flags" in empirical findings: Demonstration and recommendations for authors, reviewers and editors. *Academy of Management Learning & Education* 16 (1): 110–124.

Bernerth, J., and H. Aguinis. 2016. A critical review and best-practice recommendations for control variable usage. *Personnel Psychology* 69 (1): 229–283.

Bettis, R.A., S. Ethiraj, A. Gambardella, C. Helfat, and W. Mitchell. 2016. Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal* 37 (2): 257–261.

Bobko, P. 2001. *Correlation and regression*. 2nd ed. Thousand Oaks: Sage.

Boellis, A., S. Mariotti, A. Minichilli, and L. Piscitello. 2016. Family involvement and firms' establishment mode choice in foreign markets. *Journal of International Business Studies* 47 (8): 929–950.

Bosco, F.A., H. Aguinis, J.G. Field, C.A. Pierce, and D.R. Dalton. 2016. HARKing's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology* 69 (3): 709–750.

Breaugh, J.A. 2008. Important considerations in using statistical procedures to control for nuisance variables in non-experimental studies. *Human Resource Management Review* 18 (4): 282–293.

Butler, N., H. Delaney, and S. Spoelstra. 2017. The grey zone: Questionable research practices in the business school. *Academy of Management Learning & Education* 16 (1): 94–109.

Carlson, K.D., and J. Wu. 2012. The illusion of statistical control: Control variable practice in management research. *Organizational Research Methods* 15 (3): 413–435.

Cascio, W.F. 2012. Methodological issues in international HR management research. *International Journal of Human Resource Management* 23 (12): 2532–2545.

Cascio, W.F., and H. Aguinis. 2005. Test development and use: New twists on old questions. *Human Resource Management* 44 (3): 219–235.

Cascio, W.F., and S. Zedeck. 1983. Open a new window in rational research planning: Adjust alpha to maximize statistical power. *Personnel Psychology* 36 (3): 517–526.

Chang, S.J., A. van Wittleloostuijn, and L. Eden. 2010. From the editors: Common method variance in international business research. *Journal of International Business Studies* 41 (2): 178–184.

Chen, E.E., and S.P. Wojcik. 2016. A practical guide to big data research in psychology. *Psychological Methods* 21 (4): 458–474.

Cortina, J.M. 2002. Big things have small beginnings: An assortment of "minor" methodological misunderstandings. *Journal of Management* 28 (3): 339–362.

Cortina, J.M., J.P. Green, K.R. Keeler, and R.J. Vandenberg. 2017a. Degrees of freedom in SEM: Are we testing the models that we claim to test? *Organizatonal Research Methods* 20 (3): 350–378.

Cortina, J.M., H. Aguinis, and R.P. DeShon. 2017b. Twilight of dawn or of evening? A century of research methods in the *Journal of Applied Psychology*. *Journal of Applied Psychology* 102 (3): 274–290.

Credé, M., and P.D. Harms. 2015. 25 years of higher-order confirmatory factor analysis in the organizational sciences: A critical review and development of reporting recommendations. *Journal of Organizational Behavior* 36 (6): 845–872.

Cuervo-Cazurra, A., U. Andersson, M.Y. Brannen, B. Nielsen, and A.R. Reuber. 2016. From the editors: Can I trust your findings? Ruling out alternative explanations in international business research. *Journal of International Business Studies* 47 (8): 881–997.

Dalton, D.R., H. Aguinis, C.A. Dalton, F.A. Bosco, and C.A. Pierce. 2012. Revisiting the file drawer problem in metaanalysis: An empirical assessment of published and non-published correlation matrices. *Personnel Psychology* 65 (2): 221–249.

Davis, G.F. 2015. What is organizational research for? *Administrative Science Quarterly* 60 (2): 179–188.

Edwards, J.R., and J.W. Berry. 2010. The presence of something or the absence of nothing: Increasing theoretical precision in management research. *Organizational Research Methods* 13 (4): 668–689.

Eysenck, H.J. 1978. An exercise in mega-silliness. *American Psychologist* 33 (5): 517.

Fanelli, D. 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 4: e5738. https://doi.org/10.1371/journal.pone.0005738.

Fischer, R., and A. Mansell. 2009. Commitment across cultures: A meta-analytical approach. *Journal of International Business Studies* 40 (8): 1339–1358.

Fisher, G., and H. Aguinis. 2017. Using theory elaboration to make theoretical advancements. *Organizational Research Methods* 20 (3): 438–464.

Fitzsimmons, S., Y. Liao, and D. Thomas. 2017. From crossing cultures to straddling them: An empirical examination of outcomes for multicultural employees. *Journal of International Business Studies* 48 (1): 63–89.

Freese, J. 2007. Replication standards for quantitative social science: Why not sociology. *Sociological Methods & Research* 36 (2): 153–172.

Friedman, D., and S. Sunder. 1994. *Experimental methods: A primer for economists*. New York: Cambridge University Press.

Fung, S.K., G. Zhou, and X.J. Zhu. 2016. Monitor objectivity with important clients: Evidence from auditor opinions around the world. *Journal of International Business Studies* 47 (3): 263–294.

Funk, C.A., J.D. Arthurs, L.J. Treviño, and J. Joireman. 2010. Consumer animosity in the global value chain: The effect of international production shifts on willingness to purchase hybrid products. *Journal of International Business Studies* 41 (4): 639–651.

George, G. 2014. Rethinking management scholarship. *Academy of Management Journal* 57 (1): 1–6.

Goldfarb, B., and A.A. King. 2016. Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal* 37 (1): 167–176.

Harlow, L.L., and F.L. Oswald. 2016. Big data in psychology: Introduction to the special issue. *Psychological Methods* 21 (4): 447–457.

Hollenbeck, J.H., and P.M. Wright. 2016. Harking, sharking, and tharking: Making the case for post hoc analysis of scientific data. *Journal of Management* 43 (1): 5–18.

Hunter, J.E., and F.L. Schmidt. 2015. *Methods of meta-analysis: Correcting error and bias in research findings*. 3rd ed. Thousand Oaks: Sage.

Hurley, A.E., et al. 1997. Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior* 18 (6): 667–683.

Ioannidis, J.P.A. 2005. Why most published research findings are false. *PLoS Medicine* 2 (8): e124. https://doi.org/10.1371/journal.pmed.0020124.

John, L.K., G. Loewenstein, and D. Prelec. 2012. Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science* 23 (5): 524–532.

Kepes, S., and M.A. McDaniel. 2013. How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology* 6 (3): 252–268.

Kerr, N.L. 1998. HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review* 2 (3): 196–217.

Lisak, A., M. Erez, Y. Sui, and C. Lee. 2016. The positive role of global leaders in enhancing multicultural team innovation. *Journal of International Business Studies* 47 (6): 655–673.

Macaskill, P., S. Walter, and L. Irwig. 2001. A comparison of methods to detect publication bias in meta-analysis. *Statistics in Medicine* 20 (4): 641–654.

MacCallum, R.C., M. Roznowski, and L.B. Necowitz. 1992. Model modification in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin* 111 (3): 490–504.

Meyer, K.E., A. van Witteloostuijn, and S. Beugelsdijk. 2017. What's in a *p*? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies* 48 (5): 535–551.

Nosek, B.A., J.R. Spies, and M. Motyl. 2012. Scientific utopia II: Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science* 7 (6): 615–631.

Nuijten, M.B., C.H. Hartgerink, M.A. Assen, S. Epskamp, and J.M. Wicherts. 2015. The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods* 48 (4): 1–22.

O'Boyle, E.H., G.C. Banks, and E. Gonzalez-Mule. 2017. The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management* 43 (2): 376–399.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* 349 (6251): aac4716. https://doi.org/10.1126/science.aac4716.

Orr, J.M., P.R. Sackett, and C.L.Z. DuBois. 1991. Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology* 44 (3): 473–486.

Reeb, D., M. Sakakibara, and I.P. Mahmood. 2012. Endogeneity in international business research. *Journal of International Business Studies* 43 (3): 211–218.

Richard, P.J., T.M. Devinney, G.S. Yip, and G. Johnson. 2009. Measuring organizational performance: Towards methodological best practice. *Journal of Management* 35 (3): 718–804.

Rousseeuw, P.J., and A.M. Leroy. 2003. *Robust regression and outlier detection*. Hoboken: Wiley.

Schwab, A., and W.H. Starbuck. 2017. A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning & Education* 16 (1): 125–141.

Sijtsma, K. 2016. Playing with data – Or how to discourage questionable research practices and stimulate researchers to do things right. *Psychometrika* 81 (1): 1–15.

Simmons, J.P., L.D. Nelson, and U. Simonsohn. 2011. Falsepositive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22 (11): 1359–1366.

Spector, P.E., and M.T. Brannick. 2011. Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods* 14 (2): 287–305.

Stahl, G.K., M.L. Maznevski, A. Voigt, and K. Jonsen. 2010. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies* 41 (4): 690–709.

Starbuck, W.H. 2016. 60th anniversary essay: How journals could improve research practices in social science. *Administrative Science Quarterly* 61 (2): 165–183.

van Essen, M., P.P. Heugens, J. Otten, and J. van Oosterhout. 2012. An institution-based view of executive compensation: A multilevel meta-analytic test. *Journal of International Business Studies* 43 (4): 396–423.

Verbeke, A., M.Y. Von Glinow, and Y. Luo. 2017. Becoming a great reviewer: Four actionable guidelines. *Journal of International Business Studies* 48 (1): 1–9.

Waldman, I.D., and S.O. Lilienfeld. 2016. Thinking about data, research methods, and statistical analyses: Commentary on Sijtsma's (2014) "playing with data". *Psychometrika* 81 (1): 16–26.

Wright, P.M. 2016. Ensuring research integrity: An editor's perspective. *Journal of Management* 42 (5): 1037–1043.

3

Science's Reproducibility and Replicability Crisis: A Commentary

Andrew Delios

The Reproducibility and Replicability Crisis: Five Elements

After a long undercurrent of discussion in the management and international business (IB) fields, our scholarly community has started to develop a consensus on the importance of the reproducibility and replicability of our research (Cuervo-Cazurra et al. 2016; Meyer et al. 2017).

Long chided as the off-spring of unimaginative researchers, replication studies were not looked upon favourably by the scholarly community at large. Slowly, a few replication studies began to be published. As the results of these initial studies demonstrated the challenges to successful

Aguinis, H., W. F. Cascio, and R. S. Ramani. 2017. Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies* 48 (6): 653–663.

A. Delios (✉)
National University of Singapore, Singapore, Singapore
e-mail: bizakd@nus.edu.sg

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_3

replication and reproducibility, scholars began to devote more effort to not only undertaking replications and reproductions of existing work; they began to identify reasons why studies could be difficult to replicate and reproduce.

It is at this juncture in the process where Aguinis, Cascio, and Rarnani (2017) help to progress our understanding of the research process-related issues that contribute to the so-called replication crisis (Ioannidis 2012; Pashler and Harris 2012). Aguinis and his colleagues (2017) identify five critical elements in the implementation of empirical designs that can lead to sample-specific results: variable definition, model specification, development of the sample, identification of relationships, and post-analysis hypothesis deduction.

Each of the five elements is a critical step in the implementation of empirical research. Anyone who has engaged in empirical research fully understands that there are multiple choices that can be made in the implementation of each element. Most choices are technically correct; yet these same choices can potentially yield different results (Simmons et al. 2011).

The main concern of Aguinis et al. (2017) is that when making these choices, empirical researchers are not seeking to tell the empirical story in the way most consistently revealed in the data. Instead, they are telling an empirical story that is most consistent with two objectives: matching the conceptual ideas being advocated in the study and matching the researchers' objective of publishing their work.

Seeking to meet either or both of these objectives does not suggest malfeasance on the part of scholars, but what it does highlight is that the standards we employ in our empirical IB research are guidelines, not rules. We have few hard-and-fast principles that guide critical choices for variable definition, model specification, sample development, and post-analysis hypothesis deduction.

Critically, I leave out the fourth element in this chain, which is the reporting of *p*-values. The one hard-and-fast rule we teach our students when they are first being indoctrinated into the world of empirical research is the magic of a *t* of 1.96 and a *p* of 0.05. Aguinis et al. (2017) argue for removing the requirement to report *p*-values, under the idea that this reporting rule could potentially distort the ways in which

models are developed through the increasingly well-known process of *p*-hacking. *p*-hacked results are those in which empirical researchers have made modelling choices that enable a result to be reported as significant beyond some critical threshold value.

Although *p*-hacking can be a problem, it does not necessarily result in as much information truncation as suggested in Aguinis et al. (2017), provided that coefficient estimates are accompanied by standard errors or *t*-statistics. Doing so permits a reader to make independent inferences about hypothesis tests, without being thrust into the black-and-white world of *p*-value based significance testing.

Moving away from this one hard-and-fast rule brings us back to the point that to a large extent, empirical research is embedded with a myriad of decisions, where it is challenging for a researcher to be clear which is the best decision to make. The consequence is that we often see reporting of robustness tests to show the implications of making different decisions.

The Role of Meta-Analysis in the Crisis

With the ability to digitally communicate the results of any published paper, empirical researchers can now include multiple appendices that show the implications of various specifications, variable definitions, and sample selection-related choices. The additional empirical work also comes with the commensurate challenge that reading and digesting published work becomes a more involved and time-consuming task.

Even with the addition of detailed appendices reporting multiple empirical tests, Aguinis et al. (2017) remain concerned that researchers, when making their choices, may still choose the empirical story most consistent with the conceptual ideas advocated in the study, and the one most likely to achieve the researchers' desire for publication of their work. To move beyond this issue, Aguinis et al. (2017) promote meta-analysis as a kind of empirical catch-all that can wash away the veneer of all the choices made in the modelling process to reveal the so-called true relationships that exist in a given field of research.

Meta-analysis is meant to integrate findings. It takes results from multiple different samples often obtained in different settings. This is done

not only to increase statistical power but also to decrease the reliance of inference from any one, potentially sample-specific, finding. Meta-analysis is particularly useful in fields where experiments or small sample studies are common, and where the data gathered are of a primary nature.

We have seen meta-analyses emerge on key relationships in the IB field (Buckley et al. 2014). The challenge with using meta-analysis to uncover where research has capitalized on chance, in the words of Aguinis et al. (2017), is that we have to be sure that the constituent studies to a meta-analysis are consistent with the theory of this methodology.

However, if we look at much of the research in international business, it is undertaken using large archival datasets. These datasets are few in number and accessible to all empirical researchers, provided funding is available. Secondary data of this nature does not suit the original design and intent of meta-analysis, especially where the meta-analysis is reliant on samples that use portions of the same source data. As such, meta-analysis can only partially address some of the core research design issues that contribute to the reproducibility and replication crisis. If we are to truly address the crisis, we need a more substantial re-thinking of the design of our empirical methods.

New Solutions to the Crisis

The ideas advocated in Aguinis et al. (2017) represent opportunities to improve our empirical methods, but their recommendations are made in the context of amendments to existing research practices. The authors miss the opportunity to question more fundamentally whether IB researchers should take a larger step backwards and consider whether existing research protocols should continue to be standard operating procedures or replaced (see Table 3.1).

Take as an example their fifth problem, post-analysis hypothesis deduction, or in the vernacular, HARKing. Aguinis et al. (2017) provide a list of potential solutions such as using strong inference testing and identifying hypotheses as post-hoc testing. The one idea in Aguinis et al. (2017) that merits more discussion is study registry (Aguinis and Vandenberg 2014). Pre-registering hypotheses and research designs is becoming more

Table 3.1 Extensions on treatments to reduce risks of capitalizing on chance

Areas of empirical design	Aguinis et al. (2017) recommendations	Alternative approaches
Variable definition	Align with theory Risk of capitalizing intentionally on chance greater with big data	Create clear standards Work as a field to define these standards
Model specification	Be clear on logic used to include control variables All variables in a model require justification	Clearly link empirical design of new study to prior study or studies Establish clearly how new study builds on previous study Deviations from standard specifications, if tolerated, need to be clearly identified
Development of the sample	Identify and treat outliers transparently	Define characteristics of data Discuss alternatives for dealing with missing data and extreme values Examine leverage not simply presence of outliers Report sensitivity of models to various choices regarding inclusion or exclusion of cases depending on treatment of missing data and outliers
Identification of relationships	More complete reporting of *p*-values Use numbers not qualitative terms Avoid several standards in one paper	Report t-statistics or standard errors Give clear graphical depictions of results including error bands Present enough information so that readers can make their own interpretations of a relationship from the reported statistical analysis
Post-analysis hypothesis deduction	Strong inference testing Deductive, abductive, and inductive research Study registries	Pre-register study design and hypotheses Crowd-sourcing Open Science Badge

common in various fields. Once a study is registered, a researcher can also engage in crowd-sourcing results, hence conducting replications at the same time as conducting the original hypothesis test (Silberzahn and Uhlmann 2015; Silberzahn et al. 2017). These two steps when taken together can more effectively address concerns about HARKing.

These two steps—pre-registration and crowd-sourcing—represent more fundamental changes to the research process than many of the solutions suggested in Aguinis et al. (2017). Whether one wants to make such changes to the research process depends on how deeply one is concerned with the present-day state of empirical methods in IB research.

A related amendment to the research process that I would like to suggest connects to the Open Science movement, where researchers are encouraged to share data and research materials, and pre-register hypotheses (Center for Open Science 2019). This related idea is that empirical researchers could construct more accurate and detailed logs for the process of moving from raw data to the finalized model. A live log (let us call it an "online research log" or "r-log") would identify, in real time, all variable transformations, all data manipulations, all the steps taken to merge and clean data, complete with saved codes. An r-log could provide valuable and non-retractable information on how the empirical researchers have arrived at their final model.

The reasoning for recording and publishing detailed logs is embedded in the first few paragraphs of this commentary. Empirical research involves a large number of decisions to move from raw data to results. The amount of information reported in a detailed methods section of a published paper is never enough to show even a small subset of all the decisions that were made over the months and sometimes years of working with a data set. Aside from an r-log permitting researchers to retrace their steps, to the extent we can have better records of the decisions, we can better understand how and why results can be reproduced or replicated.

Conclusion

Finally, at the end of the day, we need to have some acceptance of where empirical research sits in the work we do as IB scholars. The extent to which we need to be concerned with decision-making and how detailed and transparent we choose to make the reporting process both connect to the point about how strongly we believe we are able to accurately reflect the real world in our empirical models.

If we do not have common standards emerging for variable definitions or for model specifications, we have little chance to develop a coherent body of empirical evidence for any area of IB research. The advances we receive in research funding, the advances we make in research methods, and the advances we have made in developing a broad base of capable and engaged IB scholars together suggest that we have a greater chance to seek harmony and agreement on standards in measurement and specification. Until now, we have been flexible in our standards under the idea that IB research would be practically infeasible if we enforced standards that functioned as hard-and-fast rules for publication.

If we do not develop such standards, then the so-called replication crisis is not a crisis; it is a reality and a logical off-shoot of the accepted research standards we have in the field of IB research. Aguinis et al. (2017) usefully identify exactly why that is the case. Hence, our decision as a community of scholars is not whether we should engage in replication or reproducibility studies. Instead, our decision is whether we want to make the investments necessary to re-think the core of our methods and to address the long-standing systemic challenges to conducting good, repeatable empirical research in international business.

References

Aguinis, H., and R.J. Vandenberg. 2014. An ounce of prevention is worth a pound of cure: Improving research quality before data collection. *Annual Review of Organizational Psychology and Organizational Behavior* 1 (1): 569–595.

Aguinis, H., W.F. Cascio, and R.S. Rarnani. 2017. Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies* 48 (6): 653–663.

Buckley, P.J., T.M. Devinney, and R.W. Tang. 2014. Meta-analytic research in international business and international management. In *The multinational enterprise and the emergence of the global factory*, ed. P.J. Buckley, 100–134. Basingstoke: Palgrave Macmillan.

Center for Open Science. 2019. Open Science Badges enhance openness, a core value of scientific practice. https://cos.io/our-services/open-science-badges/. Accessed 17 Mar 2019.

Cuervo-Cazurra, A., U. Andersson, M.Y. Brannen, B. Nielsen, and A.R. Reuber. 2016. From the editors: Can I trust your findings? Ruling out alternative explanations in international business research. *Journal of International Business Studies* 47 (8): 881–997.

Ioannidis, J.P. 2012. Why science is not necessarily self-correcting. *Perspectives on Psychological Science* 7 (6): 645–654.

Meyer, K.E., A. van Witteloostuijn, and S. Beugelsdijk. 2017. What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies* 48 (5): 535–551.

Pashler, H., and C.R. Harris. 2012. Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science* 7 (6): 531–536.

Silberzahn, R., and E.L. Uhlmann. 2015. Crowdsourced research: Many hands make tight work. *Nature* 526 (7572): 186–191.

Silberzahn, R., E.L. Uhlmann, D.P. Martin, P. Anselmi, F. Aust, E.C. Awtrey, et al. 2017. Many analysts, one dataset: Making transparent how variations in analytical choices affect results. https://doi.org/10.1177/2515245917747646. Accessed 17 Mar 2019.

Simmons, J., L. Nelson, and U. Simonsohn. 2011. False-positive psychology: Undisclosed flexibility in data collection and analysis allow presenting anything as significant. *Psychological Science* 22 (11): 1359–1366.

Part III

Hypothesis-Testing Research

4

What's in a *p*? Reassessing Best Practices for Conducting and Reporting Hypothesis-Testing Research

Klaus E. Meyer, Arjen van Witteloostuijn, and Sjoerd Beugelsdijk

Introduction

> The value for which $p = 0.05$, or 1 in 20, is 1.96 or nearly 2; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not. (Fisher 1925: 45)

K. E. Meyer (✉)
Western University, London, ON, Canada
e-mail: kmeyer@ivey.uwo.ca; kmeyer@ceibs.edu

A. van Witteloostuijn
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
e-mail: a.van.witteloostuijn@vu.nl

S. Beugelsdijk
University of Groningen, Groningen, The Netherlands
e-mail: s.beugelsdijk@rug.nl

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_4

If one were to believe all results published in business journals, scholars would be able to predict the future (almost) perfectly. In the vast majority of the papers published, almost all theoretically derived hypotheses are empirically supported. For example, of the 711 hypotheses tested in articles published in the *Journal of International Business Studies (JIBS), Strategic Management Journal (SMJ)*, and *Organization Science* in the 2016 volumes, about 89% find empirical support for the theoretical predictions. In a similar exercise in 1959, Sterling reported a statistical significance percentage of 97% (Sterling 1959). The above interpretation of scholars as futurologists assumes that published research is representative of the population of all completed studies on a specific topic. There is plenty of evidence that this is not the case (Sterling 1959; Ioannidis 2005). What is known as the "file-drawer problem" is very common: scientific studies with negative or nil-results often remain unpublished (Rosenthal 1979; Rothstein et al. 2005).

Moreover, accumulating evidence suggests that authors actively engage in pushing significance levels just below the magic threshold of $p = 0.05$, a phenomenon referred to as '*p*-hacking' or 'search for asterisks' (Bettis 2012; Brodeur et al. 2016). Similarly, some authors appear to engage in HARKing, which stands for Hypothesizing After the Results are Known (Bosco et al. 2016; Kerr 1998). The problem of both practices is that the reported significance levels are misleading because readers are given no information how many nulls and negatives ended up in the research dustbin along the way. Editorial boards, reviewers, and authors are increasingly aware of the challenge to minimize 'the search for asterisks,' In this editorial, we document recent initiatives and suggest ten concrete guidelines in order to minimize the risk of reporting false positives (i.e., type I errors), and more generally improve the quality of hypothesis-testing research and statistical reporting in our field.

Our editorial responds to the recent surge of pleas to change extant research practices, across a wide variety of disciplines, including business studies. For instance, Bettis et al. (2016a, b) in strategic management, Barley (2016) in administrative sciences, Aguinis et al. (2010) in organizational studies, and van Witteloostuijn (2016) in international

business raise major concerns about the current state of affairs. These recent pleas, triggered by hot debates in disciplines such as medicine (Crosswell et al. 2009; Ioannidis 2005; Lexchin et al. 2003) and psychology (Gigerenzer 2004; John et al. 2012; Simmons et al. 2011), fit into a long tradition of work highlighting the need for the scholarly community to continuously improve its research practices (e.g., Sterling 1959; Rosenthal 1979). In this editorial, we focus in particular on calls for more transparency regarding the process of empirical research, and hence more accurate reporting and comprehensive interpretation of empirical results. Our aim is to derive from the ongoing discussions, a set of concrete and actionable treatments, which we translate into guidelines and best practices for *JIBS* authors.

Our starting point is the observation that current practices stimulate the publication of false positives. This argument is anything but new and the reasons for this problem have been extensively analysed, a particularly forceful voice being Ioannidis's (2005, 2012). The root of the problem is the publication bias, caused by journals seeking theoretical novelty with empirical confirmation, in combination with counterproductive university-level career incentives focused on publications in a limited number of journals (for a recent summary, see van Witteloostuijn 2016). However, a recent upsurge of scandals such as Stapel's data-fabricating misconduct in social psychology (*New York Times* 2011) triggered a powerful movement toward changing the ways in which the scientific community has institutionalized practices that stimulate rather than discourage such behaviour. Examples, among many, are orchestrated replication projects (e.g., *Open Science Collaboration* 2015) and journal repository requirements (e.g., the *American Economic Review*). In the business studies domain, the recent change of statistical reporting guidelines by the *Strategic Management Journal* (Bettis et al. 2016a, b), swiftly followed by *Organization Science* and other journals (see, e.g., Lewin et al. 2016), is a clear signal that research practices are currently being revised and updated.

As the leading journal in our field, *JIBS* is committed to engage in this debate, being part of this wider movement continuously (re)assessing the criteria for what counts as rigorous empirical research. We hope that our suggestions will help further improving the work published in (interna-

tional) business, as well as in triggering an ongoing reflection on what best research practices entail. To do so, we propose ten guidelines that are concrete and actionable. These guidelines serve as suggestions (not as fixed rules), providing direction for authors submitting papers employing quantitative hypothesis-testing methods. These guidelines should not result in a uniform straightjacket, but help advance research practices and stimulate the search for solutions to shortcomings in contemporary practice. Research best practices are not set in stone, but experience shows that a set of benchmarks for both researchers and reviewers can be very helpful to push the quality bar of research upward.

Challenges to Current Practice

The Focus on *p*-Values and False Positives

The null hypothesis significance testing practice was introduced by Fisher (1925) to distinguish between interesting relationships and noise. Null hypothesis significance testing has quickly become the norm in social sciences, including business studies. Before computers and software packages such as SPSS and STATA became widely available, the *p*-values associated with specific test statistics related to a particular relationship were looked up in a statistical table. As *p*-values were given for a limited set of cutoff values (particularly $p = 0.10$, $p = 0.05$, and $p = 0.01$), a practice emerged to report *p*-values with respect to these benchmarks (e.g., $p = 0.05$), and to indicate the significant estimates with $*$, $**$ or $***$. Fisher (1925) suggested, somewhat arbitrarily, using $p = 0.05$ as the most appropriate cutoff level. With increased computing power, however, scholars became able to calculate exact *p*-values for even the most advanced statistical models. But due to path dependency, the old asterisks habit remained in place.

Despite the importance and influence of Fisher's work, and the intuitive attractiveness of using a simple cutoff value, the focus on *p*-values is not without its negative external effects. Particularly, the focus on *p*-values leads to publication bias. It has always been the case that journals have an interest in publishing interesting results – i.e., significant estimates – and

not noise (to paraphrase Fisher), but the introduction of the publish-or-perish culture appears to have increased the publication bias. It has been argued that this development is due to a counterproductive academic reward structure, arising from the combination of top-tier journals' preference for 'statistically significant results' and a highly competitive tenure-track system in many universities that relies disproportionately on top-tier journal publications (Bedeian et al. 2010; Pashler and Wagenmakers 2012).

This reward structure encourages practices inconsistent with statistical best practice (Wasserstein and Lazar 2016), specifically *ex post* writing of hypotheses supposedly *ex ante* tested, also referred to as HARKing (Kerr 1998), and of manipulating of empirical results to achieve threshold values, varyingly referred to as *p*-hacking (Head et al. 2015; Simmons et al. 2011) star wars (Brodeur et al. 2016), and searching for asterisks (Bettis 2012). The heavy focus on significant effects opens the door to a variety of questionable (and occasionally plain bad) practices, some of which we discuss below in greater detail. Most fundamentally, such approaches are inconsistent with Popper's (1959) falsification criterion, which is the philosophical foundation for conducting hypothesis tests in the first place (van Witteloostuijn 2016). As a consequence, the reliability and validity of cumulative work are not as high as they could be without biases in the publication process.

The publication bias arises from two practices. First, papers reporting significant relationships are more likely to be *selected* for publication in journals, leading to a bias towards tests rejecting the null hypothesis. Second, authors 'fine-tune' their regression analysis to turn marginally nonsignificant relations (those just above $p = 0.01$, $p = 0.05$ or $p = 0.10$) to significant relations (i.e., just below these thresholds), which causes an *inflation* of significance levels in (published and unpublished) empirical tests.[1] As said, these biases in article selection and significance inflation are anything but new (Sterling 1959), and evidence for such unbalances has been firmly established in sciences (Head et al. 2015).

The selection bias has received a great deal of attention in medical research, obviously because of the immediate medical and societal implications of prescribing medications based on possibly flawed results. In the medical field, the problem of selection bias is exacerbated by the intricate relationships between pharmaceutical companies and research (Lexchin

et al. 2003). However, the problem of publication bias for social sciences is not to be underestimated either: exactly because in social sciences, reality and truth are partially socially constructed (or at least socially interpreted), public policies, managerial practices, and other practical implications derived from social science research can have great impact on many.

Selection bias can be found in economics (Brodeur et al. 2016), political science (Gerber et al. 2001), and psychology (Ferguson and Heene 2012) too. With the advent of meta-analytical techniques, it has become more and more common to explore the sensitivity of the results for a selection bias. Typically, it is found that – all else equal – the probability of finding significant regression coefficients in published articles is much higher than in working papers addressing the same topic (Gorg and Strobl 2001; Rothstein et al. 2005).[2] The variation in the severity of the selection bias across domains has been related to characteristics such as the size of the discipline and the degree of methodological consensus, as well as to the extent to which there is competition between theoretical predictions (Brodeur et al. 2016). Interestingly, the selection bias is lower when theory is contested (Doucouliagos and Stanley 2013), suggesting that academic debate remains critical even, or perhaps especially so, in the face of so-called stylized facts (i.e., established findings). Also, papers published by tenured and older researchers seem to "suffer" less from *p*-hacking (Brodeur et al. 2016), probably because for them career concerns are less of an issue.

In econometrics, the discussion of inflation bias goes back to the debate on pretesting in the 1970s and 1980s. It was generally acknowledged that as a result of running multiple tests, and leaving out insignificant variables, the final model typically includes focal variables with *p*-values that are inflated. One econometric strategy developed by Leamer in response to the discussion on the inclusion of control variables in the early 1980s has been to perform a so-called extreme-bounds analysis (Leamer 1985). The basic idea of this analysis is to analyse the consequences of changing the set of control variables for the estimated effect of x_i on a specific dependent variable. Instead of selecting a fixed set of control variables (that happen to give the lowest *p*-values), extreme-bounds analysis implies a series of regressions in which the coefficient of

the variable of interest is estimated by changing the set of control variables (for an application, see Beugelsdijk et al. 2004). Although this is an interesting method that received follow-up especially in economics (Sala-i-Martin 1997; Angrist and Pischke 2010), but not in business studies, extreme-bounds analysis is a rather mechanical way to explore just one dimension of robustness: sensitivity of the coefficient of the variable of interest to (selective) inclusion of control variables. It remains vulnerable to "meta-level" *p*-hacking and inflated *p*-values because of its vulnerability to the selection of the set of control variables in the first place.

Moreover, *p*-hacking occurs not only by selecting control variables depending on results obtained, but takes many different forms and shapes (Bosco et al. 2016; Head et al. 2015; John et al. 2012). For example, the decision whether to drop or include influential observations may be biased if made after the initial analysis. Some even suggested that overoptimism among academic researchers is one of the reasons why too many false positives are reported (Mullane and Williams 2013). As editors, we also observed reviewers asking for changes that promote significance (and confirmation of hypotheses), a practice running counter to establishing the validity of empirical results.

The practice of *p*-hacking is problematic because it not only affects individual careers, but also erodes the reliability of scientific studies. Bettis et al. (2016a, b) illustrate the challenge as follows. Suppose three junior scholars test the same hypothesis. Scholars A and B find no significant results; they quickly move to other topics because 'not statistically significant' will not be published in top management journals. Scholar C finds a result significant at $p < 0.05$ level, which gets published in a high-impact outlet on the basis of which s/he receives tenure. The published result is treated as scientifically proven, and not challenged. Yet the actual evidence is that two out of three studies did not find a significant effect – and no one knows how many regressions scholar C ran in addition to the one with the significant effect. This problem is not unique for nonexperimental field work; experimental study designs are not immune to *p*-hacking either, as researchers may well stop their experiments once analysis yields a significant *p*-value.

The practice of *p*-hacking may have been a matter of pluralistic ignorance in the past (many may oppose these practices, but assume that oth-

ers support them, leading to collective inaction and thereby sustained support). However, the increased publicity regarding these practices, which in some extreme situations of fraud have even led to legal cases (Bhattacharjee 2013), call for action. Scholars under pressure of 'publish or perish' face a slippery slope, moving from the subjectivity of 'sloppy science', to incomplete reporting that inhibits replication, to deliberate exclusion of key variables and/or observations, to manipulation of data, and to outright fabrication of data.

Globalization and the Internet facilitate the tracking of suspicious articles *(The Economist*, June 14, 2014), and amplify negative reputation effects after serious statistical fraud, for both authors and journals. Starting in 2010, the blog *RetractionWatch.com* discusses and reports on retractions of scientific papers. Excluding repeat offenders and adjusting for the growth of published medical and nonmedical literature, the number of articles retracted increased by a factor 11 between 2001 and 2010 (Grieneisen and Zhang 2012). One interpretation of this increase is that scientific inquiry is in crisis. Our interpretation is that open access, convergence of knowledge on statistics (partly thanks to the Internet), and increased awareness of publication ethics in tandem increase the pressure to adhere to proper statistical standards, enhance transparency and thereby boost the post-publication detection of poor practices.

The Biases and Misinterpretations of *p*-Logic Practices

The biases that cause inflated *p*-values (i.e., *p*-values that are lower than they "truly" are)[3] are problematic because the final result is research reporting too many false positives, which, in turn, lead to misguided advice to practice (e.g., Aguinis et al. 2010). As a simple yet powerful illustration, we took the last two years (2015–2016) of *JIBS, Organization Science*, and *SMJ*, and collected information on the *p*-values of all variables of interest in the estimated regression models. We followed the approach of Brodeur et al. (2016) and collected for all tests of a variable of interest in a hypothesis-testing paper information on the coefficient, reported *p*-values, and standard errors of the coefficient (or *t*-value when reported). The vast majority of the articles present the coefficient and the standard error; only few report *t*-values. We omit control variables.

For the three journals combined, this amounts to 313 articles and 5579 null hypothesis tests. This includes robustness tests (but excludes the ones published in online appendices). We do not round coefficients and standard errors, but use the full data as provided in the articles considered. Out of the 5579 hypothesis tests extracted from the three journals, 3897 are rejected at the $p < 0.10$ level, 3461 at the $p < 0.05$ level, and 2356 at $p < 0.01$ level. To obtain a homogenous sample, we transform the *p*-values into the equivalent *z*-statistics. A *p*-value of 0.05 becomes a *z*-statistic of 1.96. Following Brodeur et al. (2016), we simply construct the ratio of the reported coefficient and the standard error, assuming a standard normal distribution.[4]

The findings are visualized in Fig. 4.1. It shows the raw distribution of *z*-scores (*p*-values) in a histogram as well as the kernel density plot,

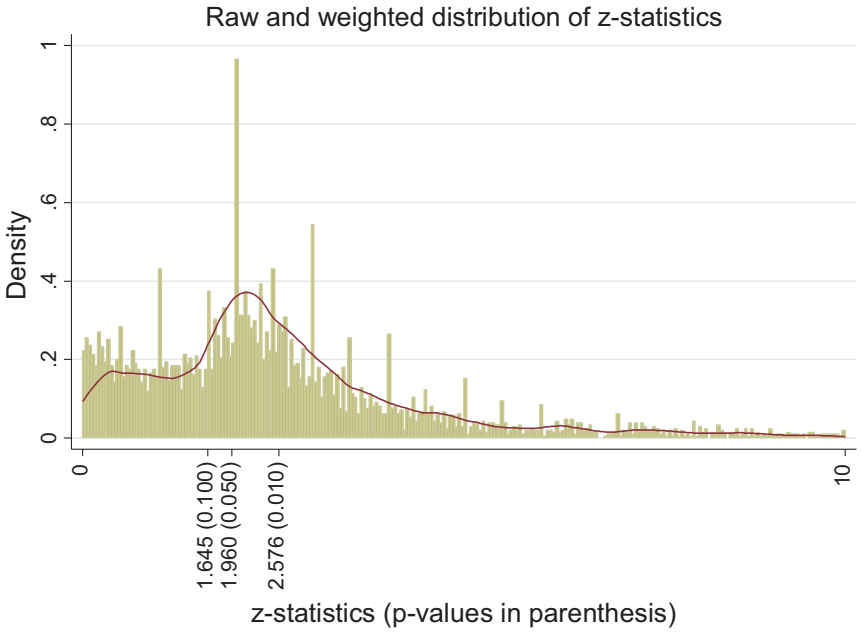

Fig. 4.1 Camel-shaped distribution of *p*-values in *JIBS, OrgScience* and *SMJ* (2015 and 2016). (Note: The graph shows the histogram as well as the kernel density plot of the weighted distribution of *z*-scores in all hypotheses testing articles published in *JIBS, Organization Science*, and *SMJ* in 2015 and 2016)

weighted for the number of hypotheses tested in an article. A kernel density plot is a nonparametric technique to visualize the underlying distribution of a continuous variable, in this case the distribution of *p*-values. It is nonparametric because it does not assume any underlying distribution such as a normal one. Intuitively, a kernel density plot can be seen as a sum of bumps. In plotting the kernel density plot, we weigh by number of hypotheses tests per article, because we want to give each article equal weight in the overall distribution. Some papers may have many hypotheses (e.g., Choi and Contractor 2016), whereas others may only have one or two (e.g., Husted et al. 2016). Separate graphs for *JIBS, SMJ* and *Organization Science* produce similar distributions and density plots (available upon request from the authors). Including or excluding robustness tests does not affect overall findings either.[5]

The shape of the figure is striking. The distribution of *p*-values in these three top management journals is not normally distributed, but has a camel-shaped distribution with a local maximum just above 1.96 (*p*-value is under 0.05), and a valley just left of 1.96. The combination of a spike just above the *p*-value of 0.05 and the valley just below in the distribution of *p*-values close to the critical value of 0.05 (critical from a reporting point of view) corresponds with similar findings in economics and psychology. Brodeur et al. (2016) also find such a camel-shaped distribution of *p*-values for top economics journals like the *American Economic Review, Quarterly Journal of Economics*, and *Journal of Political Economy.* Masicampo and Lalande (2012) report a significantly higher incidence of *p*-values just below $p = 0.05$ for the *Journal of Experimental Psychology*, *Journal of Personality and Social Psychology*, and *Psychological Science*. Hence such a skewed distribution of *p*-values in (these) business journals is no exception to the distribution of *p*-values in other disciplines. The finding is not the result of a selection bias (only significant results are published), because a similar exercise comparing conference papers and published papers in strategy research shows "an abundance of false or inflated findings" also for conference papers, suggesting that "even in early stage work, authors seem to filter results to find those that are statistically significant" (Goldfarb and King 2016: 169). Combined, this evidence is strongly suggestive of a possible inflation bias resulting from *p*-hacking.

Another reason why the focus on *p*-values leads to too many false positives is cumulative incidence of false positives. If a study was conducted in a strictly sequential manner, where first the hypotheses are developed and then a single test was conducted, then the *p*-value would indicate what we stated above. However, in practice, scholars often conduct many tests, and develop their theory *ex post* but present it as if the theory had been developed first. In part, this is driven by the review process as authors anticipate less favourable reactions of reviewers to nonsignificant results (Orlitzky 2012; Pfeffer 2007). Even within the review process, hypotheses may be added or dropped, often on request of reviewers (Bedeian et al. 2010; Pfeffer 2007). However, if a 'best' result is selected from many regressions, then the *p*-value overstates the degree of support for the theoretical argument. In the extreme case, if in truth there is no effect but a numbers cruncher runs 20 different regressions, then on average one of these 20 regressions (i.e., 5%) should be significant at the $p < 0.05$ level (Bobko 2001).

To the best of our knowledge, there is no empirical work in business studies on the cumulative incidence of false positives. However, in medical research, it appears to be very serious. In randomized trials among 150,000 men and women receiving multiple cancer screenings, the risk of a false positive test is 50% higher after the 14th test (Crosswell et al. 2009). The practical consequence of a false positive may be more severe in a medical screen than in a test of management theory, but unbiased evidence is an essential precondition for business scholars to be relevant to practicing managers and thus to make a positive real-world impact (Aguinis et al. 2010).

Paradoxically, the focus on *p*-values does not coincide with a thorough understanding of the meaning of *p*-values: *p*-values are often misinterpreted (Aguinis et al. 2010). The *p*-value generated by regression analyses is "the probability under a specified statistical model that a statistic summary of the data (e.g. the sample mean difference between two compared groups) would be equal to or more extreme than its observed value" (Wasserstein and Lazar 2016: 131). Or, in the word of classic textbook authors, the *p*-value is the probability that the sample value would be at least as large as the value observed if the null hypothesis was true (Wonnacott and Wonnacott 1990: 294). The regression result does not

prove or disprove a hypothesis, it does not provide evidence regarding the reliability of the research (Branch 2014: 257), and it does not make statements about a population other than the sample.

Moreover, the *p*-value does not tell us anything about the strength of a particular association: lower *p*-values do not make relations more substantively significant, although a finding at $p < 0.01$ is often interpreted as a stronger result than one at $p < 0.05$. For example, a regression analysis of Z on X and Y may lead to a *p*-value of 0.051 for variable X and a *p*-value of 0.049 for variable Y, yet the effect size of X can be significantly larger. As noted long ago, this aspect is often overlooked (in economics, see McCloskey 1985, and McCloskey and Ziliak 1996; in psychology, see Kirk 1996). In the above example, given the selection bias, the finding on X may even never be published. However, a *p*-value of 0.05 is just a rule of thumb suggested by Fisher in 1925 in times without computers and statistical software packages, but was never meant to be interpreted as an absolute yes-or-no threshold.

In other words, *p*-values of 0.06 versus 0.04 are (almost) equally interesting. This is especially relevant for intellectually controversial and thought-provoking pieces, where we do not want a manuscript to get rejected on the basis of a *p*-value of 0.06. Moreover, statistical significance does not say anything about effect size. Although it has become more common to include effect size discussions, not all published articles discuss effect size, and many original submissions received by *JIBS* do not (yet) include an explicit discussion of the effect size. A count for all hypotheses-testing papers in the 2016 volumes of *JIBS* (54%), *Organization Science* (40%) and *SMJ* (56%) suggests substantial variation in the practice of discussing effect size for the variables of interest.

Towards Better Practice

Alternative Study Designs

Scholars may be able to enhance the rigor of their empirical evidence through their study design. First, they may conduct multiple studies to test the same hypothesis, thus providing not only evidence of validity

under different conditions, but also reducing the opportunities for HARKing. In academic disciplines investigating behaviours of individuals, such as organizational psychology, organizational behaviour, and human resource management, it is established good practice to include multiple studies to test a new hypothesis (see, e.g., the *Journal of Applied Psychology).* In international business, where the validity of theory across geographic contexts is a key theme, offering evidence from two or more countries would often be a valuable contribution (Meyer 2006). However, for many of the research questions of interest for international business scholars, this is not realistically feasible, especially if the unit of analysis is firms or countries rather than individual people.

Second, experimental study designs offer interesting opportunities to advance international business knowledge that have yet to be fully exploited in the field. Specifically, experimental study designs allow varying specific variables of interests while keeping everything else constant, which is usually not feasible using field data. However, the empirical evidence of experimental studies also has been challenged due to sample selection biases (Henrich et al. 2010a, b) and endogeneity issues (Antonakis et al. 2010). Recent *JIBS* contributions by Buckley, Devinney, and Louviere (2007) and by Zellmer-Bruhn, Caligiuri, and Thomas (2016) outline opportunities to apply experimental designs in the field of international business, and offer methodological guidance (cf. van Witteloostuijn 2015).

As *JIBS* is interested in both rigor and relevance, we as editors are acutely aware that these methodological research design alternatives hold great potential, but are not always suitable to address many of the research questions of interest to the international business research community. Therefore, the challenge remains how we can improve the reliability of research findings based on testing hypotheses using regression analysis with single-sample field data.

Enhancing Reporting Practices

In a nutshell, *JIBS* expects that authors do the best feasible analysis with the available data in their line of research, do not engage in any research malpractices, report statistical results based on a full analysis of *p*-values, and provide maximum transparency to enable other scholars to build on

their work (including reproduction and replication; cf. Bettis et al. 2016b; Hubbard et al. 1998). In the context of the current editorial, this translates into ten suggestions for how research and reporting practices can be enhanced.

Rigorous scholarship requires discussing the evidence for and against a hypothesis based on the full evidence, not limited to a single *p*-value of a specific test. The *American Statistical Association* (*ASA*) has recently debated this concern and issued guidelines (Wasserstein and Lazar 2016). In our view, these guidelines represent current best practice, and *JIBS* editors and reviewers can refer to these guidelines when assessing papers submitted for publication. Authors should in particular avoid over-interpreting the strength of evidence for or against a hypothesis based on levels of significance. Rather, in line with guidelines by *SMJ* (Bettis et al. 2016a, b) and others, actual *p*-values, confidence intervals, and effect sizes should be fully reported and discussed (see also Bosco et al. 2015; Hunter and Schmidt 2015). Thus, the results of hypotheses tests should normally include the following:

Guideline 1: At a basic level, all regression analyses should include, for each coefficient, standard errors (as well as mention the confidence intervals for the variable of interest) and, for each regression model, the number of observations as well as the R^2 statistics or equivalent

Guideline 2: Authors should refer to the actual *p*-value rather than the threshold *p*-value when assessing the evidence for and against their hypothesis

Guideline 3: Authors should *not* report asterisks to signal *p*-value thresholds

For guideline 1, in straightforward OLS models, the standard error of the coefficient can be calculated on the basis of the estimated coefficient and the *p*-value, but for more complicated models this is not so straightforward. We therefore expect authors to report the estimated coefficient, its standard error and *exact p*-value where relevant. Guideline 2 is in line with the call for comprehensive assessment without undue focus on the traditional threshold rules of thumb. The actual *p*-values may be included

in the results table, but in many instances it may suffice to report them within the results section in the main text. In addition, Hunter and Schmidt (2015) suggest discussing not just the estimated coefficient, but also the confidence interval associated with the point estimate.

The discussion should include a reflection regarding levels of significance, given the evidence in similar studies. Sample size is critical here. For example, studies with few observations (for instance, when countries are the unit of analysis) obtain lower levels of significance than studies of independent individual decisions generated using Big Data methodologies. Guideline 3 responds to the observation that journals without "stars" have a lower probability of *p*-hacking (Brodeur et al. 2016).

Evaluating the Evidence

Good scientific practice requires that authors assess hypotheses based on a comprehensive assessment using all available evidence, rather than a singular focus on a single test statistic in a specific regression analysis. When interpreting the results, it is good practice to offer reflections and supplementary analyses that enable readers to comprehensively assess the empirical evidence. Specifically, we recommend authors to follow the following guidelines from 4 to 6 when writing their methods and results sections:

Guideline 4: Reflections on effect sizes are included, reporting and discussing whether the effects (the coefficients and, if appropriate, marginal effects) are substantive in terms of the research question at hand

Guideline 5: Outlier observations are discussed carefully, especially when they have been eliminated from the sample (e.g., through technical practices such as 'winzorizing')

Guideline 6: Null and negative findings are equally interesting as are positives, and hence are honestly reported, including a discussion of what this implies for theory

In other disciplines, such as psychology, the discussion of effect sizes has already become standard (see, e.g., Zedeck 2003: 4) and is required,

for example, at the *Journal of Applied Psychology.* Ideally, effect size is a standardized, scale-free measure of the relative size of the effect. Although not without criticism, Cohen's d is an example of such a measure (Cohen 1969).

Reflections regarding effect sizes are especially important when dealing with large datasets where it is easy to obtain statistical significance even for small effects, an increasing challenge in the era of Big Data, reflecting the fact that significance is a function of sample size, as well as the alpha and effect size. As effect-size reporting is not so straightforward, we provide further suggestions and more fine-grained guidelines. The appropriate methods to generate effect sizes vary across empirical methods, and may require additional analyses. We briefly discuss several methods that are common in IB.

First, for OLS and GLS types of models, effect sizes should be calculated and reported in the usual way using the standard error of the estimated coefficient. Standardized coefficients help for interpretative reasons. Moreover, explicit comparisons can make the interpretation much more informative. For example, authors may use wording such as "Ceteris paribus, a one standard deviation increase of cultural distance (which is comparable with a change in distance from, say, US–UK to, e.g., US–Italy) reduces the longevity of joint ventures with two to four years. For comparability, the effect of a similar increase of one standard deviation of geographic distance results in a reduction of joint longevity by eight years."

Second, for logit and hazard models, we expect a discussion of effect sizes that are readily interpretable (cf. Hoetker 2007; Zelner 2009). One way is to provide odds ratios, but most readers find these hard to interpret without additional explanation. We therefore suggest authors to provide at a minimum a clear intuitive explanation of their findings. The following example can illustrate this. Assume one is interested in exploring the relation between the institutional environment in which an R&D subsidiary operates, and the probability that this subsidiary generates new product innovations (measured as a 0–1). After having established a statistically significant relation between institutional setting and subsidiary innovative performance, the effect size discussion should explore the probability that a subsidiary generates a product innovation. For exam-

ple: "The probability of a subsidiary reporting an innovation is 5% higher when they are located in a country with a favorable institutional setting (e.g., country X) compared to the likelihood of a subsidiary developing product innovations in a country with a less favorable institutional regime (e.g., country Y) for which we find a probability of 1% – all else equal."

Third, for multilevel analysis (Peterson et al. 2012), the common practice is to calculate intra-class correlations, which provide scores for the explained variance at each level of analysis. In business studies, the lowest level of aggregation typically concerns a team-or subsidiary-level-dependent variable, firm and industry next and country last. Intra-class correlations often suggest that most of the variation is at lower levels. For example, intra-class correlations in cross-cultural psychology research suggest that most variation in values is supposedly at the individual level, and not at the country level (typically about 90% versus 10%) (Fisher and Schwartz 2011). However, it is critical to keep in mind here that the measurement error is also at the individual or firm level; the larger the measurement error at this level, the higher the explained variance at this level, and the lower the relative variance explained at the other levels. Correcting for measurement error in multilevel models is therefore critical (Fox and Glas 2002). Authors using multilevel methods should take this into account when reporting their findings.

Fourth, in interaction models (with 'moderating effects'), the common practice is to select one standard deviation above the mean and one standard deviation below the mean, and then draw two lines as if these coefficients reflect the full range of possible scores of a moderating variable, implicitly assuming they do not have a confidence interval – i.e., that uncertainty regarding the interaction coefficient is absent. This approach – though common – is incomplete because it ignores the margin of error with which the interaction effect is estimated, and it does not show the marginal effect for the whole range of scores on the moderating variable (the one standard deviation below and above the mean may not necessarily be representative values).

Ideally, authors should report confidence intervals for interaction effects over the relevant range of the explanatory variable. For linear models, Brambor et al. (2006) and Kingsley et al. (2017) nicely explain how to do this; for nonlinear models, we refer to Haans et al. (2016). There

are various ways to provide more information on the nature and magnitude of the interaction effect. Here, using an otherwise standard STATA do-file (see Appendix 1 for details), in Fig. 4.2, we provide an example of a graph visualizing how to discuss interaction effects. This is just one (simple) example of how to unpack the nature of the interaction effect (see also Williams (2012), and Greene (2010), for alternative approaches, and Hoetker (2007), Wiersema and Bowen (2009), Zelner (2009) for a discussion of logit and probit models, also including STATA do files).

In this example with continuous variables, the two outer lines give the 95% confidence range for the interaction line, which shows the marginal effect of variable X on dependent variable Y for the full range of possible scores of the moderator variable M.[6] The small dots represent all observations for M in the sample (and not just a score one standard deviation below and above the mean of M). Only if the two lines reflecting the confidence interval are *both* below and above the horizontal zeroline, the interaction effect is significant. In Fig. 4.2, this is the case for values of M left of A and values of M right of B. This graph shows that, although the

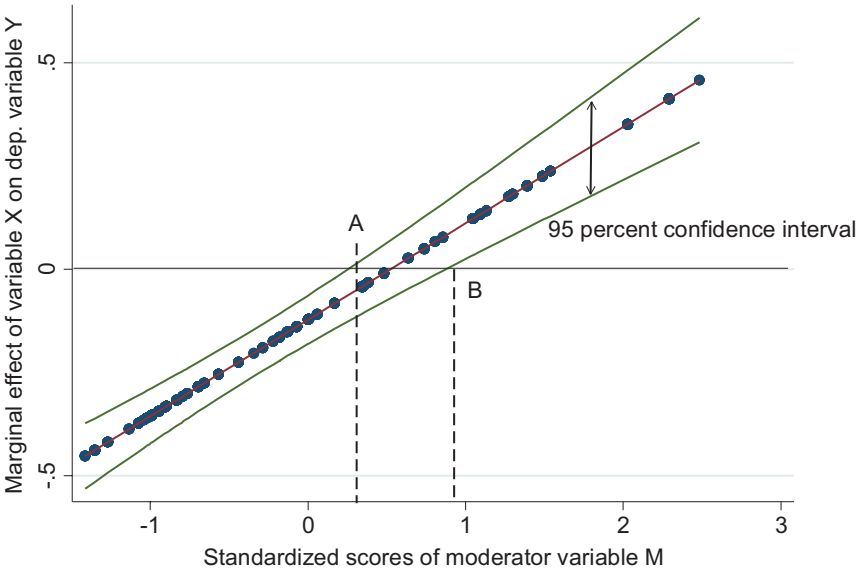

Fig. 4.2 Illustrating the effect size in interaction models

average interaction effect may turn out to be significant in a regression model, there is a range for the moderator variable M for which the effect is insignificant (between A and B), and the effect of X on Y conditional on M is negative for low values of M and positive for high values of M. Such a graph is (much) more informative than the standard practice of just showing the point estimate of X on Y for just two values of M. This observation is important, because by its very nature many international business studies are based on the starting point that key relations are moderated by contextual variation and contingent upon the external environment.

To summarize, we suggest the following more fine-grained guidelines with respect to effect-size reporting, further specifying guideline 4:

Guideline 4a: When discussing effect size, authors should take the confidence interval associated with the estimated coefficient into account as well as the minimum and maximum effect (not just one standard deviation above and below the mean), thus providing a range of the strength of a particular relationship. This may be done graphically for more complex models

Guideline 4b: When discussing effect sizes, where possible and relevant, authors should compare the range of the effect size of the variable of interest with other variables included in the regression model

The handling of outliers can significantly influence the result of regression analyses, especially if the underlying hypothesis test calls for a nonlinear relationship. Thus, as stated in guideline 5, authors should explain not only how they handled outliers, but also what the outliers tell with respect to their underlying theory. A popular but problematic approach is the use of winsorized datasets – i.e., datasets that have been transformed by eliminating extreme values (e.g., the highest and lowest 5% of the data) to reduce the effect of possibly spurious outliers. If this practice is used, non-winsorized datasets and datasets with different threshold levels for the winsorizing should be included in the robustness analysis, and discrepancies must be explained.

The results of nonsignificant or negative results can be of substantive interest for the creation of cumulative scientific knowledge, as indicated

in guideline 6. This applies in particular when results fail to fully confirm received knowledge from earlier research, and when the analysis is based on high-quality data, rigorous methods and sufficient statistical power.

Causality and Endogeneity

Another critical issue involves inference of causality and endogeneity from empirical analyses. In economics, for instance, it has become common to stop using terminology related to making any causal inference unless there is a solid identification strategy in place, which is a major challenge with any cross-sectional study design. Notwithstanding that theory normally suggests a causal direction that may be illustrated by directional arrows, many commonly used empirical techniques for cross-section data do not test for the direction of causality. Solid identification strategies are often not easy to find in nonexperimental social sciences. As a result, scholars have started to adapt their language more and more, using words like "association" and "relation" instead of "determinant" and "effect" or "affect" (cf. Reeb et al. 2012).

Apart from careful language, *JIBS* expects authors to deal with the issue of endogeneity to the extent possible (Rosnow and Rosenthal 1984; Shadish et al. 2002). Probably the best-known technical solutions are lagged explanatory variables and the instrumental variables method (Angrist and Krueger 2001). Such technical solutions can often support a causal interpretation, but cannot prove causality. After all, many of these solutions are still based on correlations, while lagged variables may be subject to inertia. Moreover, the instrument may fulfil the statistical criterion of "independence," but not at a more substantive level.

The above suggests two additional guidelines:

Guideline 7: In the absence of a clear strategy designed explicitly to identify causes and effects, authors should be careful in using terminology suggesting causal relationships between variables of interest, and accordingly adjust their language in the wording of the hypotheses and in the discussion of the empirical results

Guideline 8: To the extent feasible, authors should address issues of causality and endogeneity, either by offering technical solutions or by adopting an appropriate research design

Note that guideline 8 provides direction, and should not be interpreted as "must do." Given the difficulty of finding truly exogenous instruments, we would not want otherwise excellent papers to not make it to *JIBS* because the manuscript does not have a section on endogeneity.

Robustness Tests

In view of the challenges that may arise from using a single test statistic from a single regression equation to test a hypothesis, it is important that authors assess the evidence comprehensively. In particular, by conducting a variety of robustness tests, authors can show that a significant finding is not due to an idiosyncrasy of the chosen empirical model and/or estimation strategy. Evidence from empirical tests becomes more convincing when it is supported by appropriate robustness tests.

The discussion section of a paper provides opportunities for a comprehensive assessment of the evidence, beyond the statistical properties of the specific tests used in the focal regression analysis. What tests are appropriate varies with the design of the study and the nature of the data. It is a normal part of the reviewing process that reviewers suggest some additional robustness tests, and authors are expected to seriously engage with such suggestions.[7] If this additional work were to result in an excessive number of tables, an additional file with these tables and short explanation of them can be included in a supplement to the paper that will be made available on the *JIBS* website. Robustness tests may include, for example, additional analyses with

- alternative proxies of focal constructs (i.e., variables mentioned in the hypotheses as independent or explanatory variables), especially for those that involve abstract concepts that cannot be measured directly;
- alternative sets of control variables, especially when correlation is present in the dataset between a focal explanatory variable and a control variable; and/or

- alternative functional forms of the regression models, especially for the hypotheses that suggest nonlinear effects (Haans et al. 2016; Meyer 2009), or moderating or mediating effects (Andersson et al. 2014; Cortina et al. 2015).

A guiding principle to perform certain robustness tests is the importance to rule out alternative explanations for the same finding. In the discussion section, an informative reflection on the outcomes of the robustness analyses, in relation to the hypotheses and alternative theories, can be included to clearly identify the study's findings vis-à-vis the extant literature. While such robustness tests are common practice, we suggest that more can be done to effectively use such tests. This gives the next guideline:

Guideline 9: Authors are expected to conduct a variety of robustness tests to show that the significant finding is not due to an idiosyncrasy of the selected empirical measures, model specifications and/or estimation strategy

From HARKing to Developing Theory

Hypothesizing After the Results are Known (HARKing) in search of hypotheses for already known positive results is causing great harm to scientific progress (Bosco et al. 2016). We would like to note that HARKing is not the same as "playing with your data" to explore the nature of relationships and get better feeling for possibly interesting patterns in a dataset. HARKing refers to the practice of datamining and, after significant results are established, developing or adjusting theoretical arguments *ex post*, but presenting the theory as if already in place *ex ante*. The issue with HARKing is that we have no knowledge of the many nulls and negatives that were found but not reported along the way, and therefore readers cannot be sure as to the true power of the statistical evidence. While papers in business studies journals appear to confirm groundbreaking hypotheses, we rarely see reports about falsification outcomes.[8] As indicated in our opening paragraph, about 89% of all hypoth-

eses in *JIBS* (82%), *SMJ* (90%) and *Organization Science* (92%) were confirmed in the 2016 volumes.[9] Yet no information is provided about the many "interventions" applied to produce this abundance of positive results.

To tackle this problem, no journal can operate a policing force to monitor and sanction what is happening behind the closed doors of our authors' offices. Eliminating HARKing requires an orchestrated effort to seriously change deeply embedded practices in the scholarly community (Ioannidis 2005, 2012). What we can do, for now, is firstly to reduce the focus on single test statistics when assessing results in favour of comprehensive assessments, and thereby to reduce the incentives to engage in HARKing (hence our guidelines 1–9), and secondly to mentor and train a new generation of scholars to intrinsically dislike HARKing practices. Here, key is that established scholars lead by example. Of course, this also requires broader institutional change to remove some of the incentives that disproportionately reward scholars finding statistically significant results (Ioannidis 2012; van Witteloostuijn 2016). What journals can do boils down to, basically, two alternatives (or a combination of both).

For one, some journals are introducing the option to submit the theory first, and the empirical tests and results later (see, e.g., *Comprehensive Results in Social Psychology*, and *Management and Organization Review;* cf. Lewin et al. 2016). If the theory is accepted after a thorough review, the final manuscript will be published (of course, conditional on appropriate data and the state-of-the-art empirical analyses). This approach is nascent and it is still an open question how successful this two-step approach will be. For now, we therefore suggest that *JIBS* take the alternative route.

This second alternative is to encourage and recognize theorizing from empirical findings – i.e., the inductive leg of the development of theory. We expect papers (both submitted and published ones) to report the initial hypotheses honestly (that is, the ones drafted *before* running analyses). Developing theory *after* running analyses (to have a better explanation of the findings) is perfectly legitimate, but this could be done in a post hoc section, explicitly discussing this change of theory in relation to the results. Similarly, removing hypotheses because the evidence is weak can be problematic.

Empirical phenomena or relationships in conflict with established theories can be a powerful driver of new theoretical developments (Doh 2015). In (international) business studies, such building theory from data is more common in qualitative research, yet it is a valid methodology also with respect to quantitative data.[10] Thus, for example, a theoretical model may be motivated by connecting a surprising theoretical finding with a relevant stream of theoretical literature different than what motivated the study at the outset. Theory developed in this way *ex post* should be tested on another dataset, be it within the same paper or in a new study, similar to theory development in grounded theory research using qualitative data. The critical methodological issue here is that authors do not pretend to have a higher level of empirical support for their new theoretical ideas than what their empirical analysis provides. Thus, as editors, we encourage development or post hoc revisions of theory on the basis of empirical findings in the discussion section. This gives our final guideline:

Guideline 10: HARKing is a research malpractice. Theory developed by interpreting empirical phenomena or results should be reported as such (for example, in the discussion section)

The Role of Reviewers

In advancing international business research towards the standards reflected in the ten guidelines, we need the constructive engagement of reviewers. Firstly, this implies that we have to prevent reviewers from pushing authors towards practices that we critiqued above. This includes practices that we as editors occasionally see, such as demanding a different theoretical post hoc framing for the results already present in the paper, elimination of single hypotheses on the grounds of weak empirical support, and/or because they have been tested in prior research. Sharpening hypotheses or adding hypotheses is fine, but not around results already present in the original version. Offering post hoc alternative hypotheses to better align with findings is a natural step in the scientific research cycle, if done in the open.

Secondly, beyond avoiding negative practices, reviewers should look for positive contributions to enhance the rigor of a given study. For example, reviewers may suggest additional ways to illustrate empirical findings, or robustness tests that enhance the credibility of the results. At the same time, reviewers should avoid being perfectionists and, e.g., ask for tests that require nonexistent data, but use best practice in the given line of research as their benchmark when assessing how to evaluate the rigor of a paper under review.

Concluding Remarks

Empirical research is also the art of the feasible. In the theoretical world of an econometrics or statistics classroom, datasets have statistical properties that real-world datasets can rarely or never meet. While scholars should aspire to collect and work with high-quality datasets, as editors and reviewers we are realistic in setting our expectations. However, given these limitations, we as *JIBS* editors strongly believe that improvements are feasible, and are necessary to advance international business research to the next level and to address frequently voiced concerns regarding the validity of scholarly knowledge. This editorial has outlined what we consider good practices for conducting hypothesis testing research, and reporting and discussing the associated empirical results. We expect *JIBS's* editors, reviewers and authors to aspire to these standards. These guidelines are not written in stone, but offer benchmarks for both researchers and reviewers to enhance the quality of published international business work.

Standards are not set in stone also because they will be subject to continuous reassessment. This editorial is a clear sign of this. Debates among editors of (international) business studies journals are ongoing, and many journals are revising their editorial policies in view of these debates. This editorial has outlined concrete and actionable steps that we can take at *JIBS*. We are convinced that we, as a scholarly community, need to – and will be able to – change established research and publication practices to improve upon the current state of the art. We will all benefit from that, and will be ready to produce new and cumulative knowledge in interna-

tional business that will be impactful, from both academic and societal perspectives.

Acknowledgements We gratefully acknowledge the constructive comments from Editor-in-Chief Alain Verbeke, eleven editors of *JIBS*, as well as from Bas Bosma, Lin Cui, Rian Drogendijk, Saul Estrin, Anne-Wil Harzing, Jing Li, Richard Walker, and Tom Wansbeek. We also thank Divina Alexiou, Richard Haans, Johannes Kleinhempel, Sjanne van Oeveren, Britt van Veen, and Takumin Wang for their excellent research assistance. Sjoerd Beugelsdijk thanks the Netherlands Organization for Scientific Research (NWO grant VIDI 452-011-10). All three authors contributed equally to this editorial.

Appendix 1: Stata Do File to Create Fig. 4.2

Model:

Dependent variable = Y
Independent variable = X
Moderator variable = M
Interaction variable = X∗M

To generate Fig. 4.2:

```
predictnl me = _b[X] + _b[X∗M]∗M if e(sample),
se(seme)
gen pw1 = me–1.96∗seme
gen pw2 = me + 1.96∗seme
scatter me M if e(sample) || line me pw1 pw2 M if e(sample), pstyle(p2
   p3 p3) sort legend(off) ytitle (“Marginal effect of X on Y”).
```

Notes

1. In many disciplines contributing to international business research, conventional Type 1 error probabilities are $p < 0.05$ or 0.01. There are situations where a higher Type 1 error probability, such as $p < 0.10$, might

be justified (Cascio and Zedeck 1983; Aguinis et al. 2010), for example, when the dataset is small and a larger dataset is unrealistic to obtain.

2. Note that according to Dalton et al. (2012), the selection bias (or file-drawer problem) does not appear to affect *correlation tables* in published versus unpublished papers.
3. A "true" *p*-value would be the *p*-value observed in a regression analysis that was designed based on all available theoretical knowledge (e.g., regarding the measurement of variables and the inclusion of controls), and not changed after seeing the first regression results.
4. Brodeur et al. (2016) extensively test whether this assumption holds, as well as the sensitivity of the overall distribution to issues like rounding, the number of tests performed in each article, number of tables included, and many more. Similar to Brodeur et al. (2016), we explored the sensitivity of the shape of the distribution to such issues, and we have no reason to assume that the final result in Figure 4.1 is sensitive to these issues.
5. The spikes at *z*-scores of 3, 4, and 5 are the result of rounding and are an artefact of the data. As coefficients and standard errors reported in tables are rounded – often at 2 or 3 digits – very small coefficients and standard errors automatically imply ratios of rounded numbers, and as a consequence, result in a relatively large number of *z*-scores with the integer value of 3, 4, or 5. This observation is in line with the findings reported for Economics journals by Brodeur et al. (2016).
6. The data on which the graph is based are taken from Beugelsdijk et al. (2014).
7. If authors believe that certain suggested additional tests are not reasonable or not feasible (for example, because certain data do not exist), then they should communicate that in their reply. The editor then has to evaluate the merits of the arguments of authors and reviewers, if necessary bringing in an expert on a particular methodology at hand. If the latter is required, this can be indicated in the Manuscript Central submission process.
8. A laudable exception is the recent special issue of *Strategic Management Journal* on replication (Bettis et al. 2016b).
9. The grand total is heavily influenced by *SMJ* with 362 tested hypotheses, vis-à vis 164 in *JIBS* and 185 in *Organization Science*.
10. An interesting alternative may be abduction. For example, see Dikova, Parker, and van Witteloostuijn (2017), who define abduction as "as a

form of logical inference that begins with an observation and concludes with a hypothesis that accounts for the observation, ideally seeking to find the simplest and most likely explanation." See also, e.g., Misangyi and Acharya (2014).

References

Aguinis, H., S. Werner, J.L. Abbott, C. Angert, J.H. Park, and D. Kohlhausen. 2010. Customer-centric research: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods* 13 (3): 515–539.

Andersson, U., A. Cuervo-Cazurra, and B.B. Nielsen. 2014. Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies* 45 (9): 1063–1071.

Angrist, J.D., and A. Krueger. 2001. Instrumental variables and the search for identification: Form supply and demand to natural experiments. *Journal of Economic Perspectives* 15 (4): 69–85.

Angrist, J.D., and J.S. Pischke. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24 (2): 3–30.

Antonakis, J., S. Bendahan, P. Jacquart, and R. Lalive. 2010. On making causal claims: A review and recommendations. *Leadership Quarterly* 21 (6): 1086–1120.

Barley, S.R. 2016. 60th anniversary essay: Ruminations on how we became a mystery house and how we might get out. *Administrative Science Quarterly* 61 (1): 1–8.

Bedeian, A.G., S.G. Taylor, and A. Miller. 2010. Management science on the credibility bubble: Cardinal sins and various misdemeanors. *Academy of Management Learning & Education* 9 (4): 715–725.

Bettis, R.A. 2012. The search for asterisks: Compromised statistical tests and flawed theory. *Strategic Management Journal* 33 (1): 108–113.

Bettis, R.A., S. Ethiraj, A. Gambardella, C.E. Helfat, and W. Mitchell. 2016a. Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal* 37 (2): 257–261.

Bettis, R.A., C.E. Helfat, and M.J. Shaver. 2016b. Special issue: Replication in strategic management. *Strategic Management Journal* 37 (11): 2191–2388.

Beugelsdijk, S., H.L.F. de Groot, and A.B.T.M. van Schaik. 2004. Trust and economic growth: A robustness analysis. *Oxford Economic Papers* 56 (1): 118–134.

Beugelsdijk, S., A. Slangen, M. Onrust, A. van Hoorn, and R. Maseland. 2014. The impact of home-host cultural distance on foreign affiliate sales: The moderating role of cultural variation within host countries. *Journal of Business Research* 67 (8): 1638–1646.

Bhattacharjee, Y. 2013. The mind of a con man. *New York Times Magazine*, April 26.

Bobko, P. 2001. *Correlation and regression: Applications for industrial organizational psychology and management*. 2nd ed. Thousand Oaks: Sage.

Bosco, F.A., H. Aguinis, K. Singh, J.G. Field, and C.A. Pierce. 2015. Correlational effect size benchmarks. *Journal of Applied Psychology* 100 (2): 431–449.

Bosco, F.A., H. Aguinis, J.G. Field, C.A. Pierce, and D.R. Dalton. 2016. HARKing's threat to organizational research: Evidence from primary and meta – Analytic sources. *Personnel Psychology* 69 (3): 709–750.

Brambor, T., W.R. Clark, and M. Golder. 2006. Understanding interaction models: Improving empirical analyses. *Political Analysis* 14 (1): 63–82.

Branch, M. 2014. Malignant side-effects of null-hypothesis testing. *Theory and Psychology* 24 (2): 256–277.

Brodeur, A., M. Le, M. Sangnier, and Y. Zylberberg. 2016. Star wars: The empirics strike back. *American Economic Journal: Applied Economics* 8 (1): 1–32.

Buckley, P., T. Devinney, and J.J. Louviere. 2007. Do managers behave the way theory suggests? A choice-theoretic examination of foreign direct investment location decision-making. *Journal of International Business Studies* 38 (7): 1069–1094.

Cascio, W.F., and S. Zedeck. 1983. Open a new window in rational research planning: Adjust alpha to maximize statistical power. *Personnel Psychology* 36 (3): 517–526.

Choi, J., and F. Contractor. 2016. Choosing an appropriate alliance governance mode: The role of institutional, cultural and geographic distance in international research & development (R&D) collaborations. *Journal of International Business Studies* 47 (2): 210–232.

Cohen, J. 1969. *Statistical power analysis for the behavioral sciences*. New York: Academic Press.

Cortina, J.M., T. Kohler, and B.B. Nielsen. 2015. Restriction of variance interaction effects and their importance for international business. *Journal of International Business Studies* 46 (8): 879–885.

Crosswell, J.M., et al. 2009. Cumulative incidence of false positive results in repeated, multimodal cancer screening. *Annals of Family Medicine* 7 (3): 212–222.

Dalton, D.R., H. Aguinis, C.A. Dalton, F.A. Bosco, and C.A. Pierce. 2012. Revisiting the file drawer problem in meta-analysis: An empirical assessment of published and non-published correlation matrices. *Personnel Psychology* 65 (2): 221–249.

Dikova, D., S.C. Parker, and A. van Witteloostuijn. 2017. Capability, environment and internationalization fit, and financial and marketing performance of MNEs' foreign subsidiaries: An abductive contingency approach. *Cross-Cultural and Strategic Management* 24 (3): 405–435.

Doh, J. 2015. Why we need phenomenon-based research in international business. *Journal of World Business* 50 (4): 609–611.

Doucouliagos, C., and T.D. Stanley. 2013. Are all economic facts greatly exaggerated? Theory competition and selectivity. *Journal of Economic Surveys* 27 (2): 316–339.

Economist. 2014. *When science gets it wrong: Let the light shine in*. June 14. http://www.economist.com/news/science–and–technology/21604089-two-big-recent-scientific-results-are-looking-shakyand-it-open-peer-review. Accessed 23 Mar 2017.

Ferguson, C.J., and M. Heene. 2012. A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science* 7 (6): 555–561.

Fisher, R.A. 1925. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Fisher, R., and S. Schwartz. 2011. Whence differences in value priorities? Individual, cultural, and artefactual sources. *Journal of Cross-Cultural Psychology* 42 (7): 1127–1144.

Fox, P.J., and C.A.W. Glas. 2002. Modeling measurement error in a structural multilevel model. In *Latent variable and latent structure models*, ed. G.A. Marcoulides and I. Moustaki. London: Lawrence Erlbaum Associates.

Gerber, A.S., D.P. Green, and D. Nickerson. 2001. Testing for publication bias in political science. *Political Analysis* 9 (4): 385–392.

Gigerenzer, G. 2004. Mindless statistics. *Journal of Socio-Economics* 33 (5): 587–606.

Goldfarb, B., and A. King. 2016. Scientific Apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal* 37 (1): 167–176.

Gorg, H., and E. Strobl. 2001. Multinational companies and productivity spillovers: A meta-analysis with a test for publication bias. *Economic Journal* 111: F723–F739.

Greene, W. 2010. Testing hypotheses about interaction terms in nonlinear models. *Economics Letters* 107: 291–296.

Grieneisen, M.L., and M. Zhang. 2012. A comprehensive survey of retracted articles from the scholarly literature. *PLoS One* 7 (10): e44118. https://doi.org/10.1371/journal.pone.0044118.

Haans, R.F.P., C. Pieters, and Z.L. He. 2016. Thinking about U: Theorizing and testing U-and inverted U-shaped relationships in strategy research. *Strategic Management Journal* 37 (7): 1177–1196.

Head, M.L., L. Holman, R. Lanfear, A.T. Kahn, and M.D. Jennions. 2015. The extent and consequences of p–hacking in science. *PLoS Biology* 13 (3): e1002106. https://doi.org/10.1371/journal.pbio.1002106.

Henrich, J., S.J. Heine, and A. Norenzayan. 2010a. The weirdest people in the world? *Behavioral and Brain Sciences* 33 (2–3): 61–83.

———. 2010b. Most people are not WEIRD. *Nature* 466: 29.

Hoetker, G. 2007. The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal* 28 (4): 331–343.

Hubbard, R., D.E. Vetter, and E.L. Little. 1998. Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal* 19 (3): 243–254.

Hunter, J.E., and F.L. Schmidt. 2015. *Methods of meta-analysis: Correcting error and bias in research findings*. 2nd ed. Thousand Oaks: Sage.

Husted, B.W., I. Montiel, and P. Christmann. 2016. Effects of local legitimacy on certification decision to global and national CSR standards by multinational subsidiaries and domestic firms. *Journal of International Business Studies* 47 (3): 382–397.

Ioannidis, J.P.A. 2005. Why most published research findings are false. *PLoS Medicine* 2 (8): e124.

———. 2012. Why science is not necessarily self-correcting. *Perspectives on Psychological Science* 7 (6): 645–654.

John, L.K., G. Loewenstein, and D. Prelec. 2012. Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science* 23 (5): 524–532.

Kerr, N.L. 1998. HARKIng: Hypothesizing after results are known. *Personality and Social Psychology Review* 2 (3): 196–217.

Kingsley, A.F., T.G. Noordewier, and R.G. Vanden Bergh. 2017. Overstating and understating interaction results in international business research. *Journal of World Business* 52 (2): 286–295.

Kirk, R.E. 1996. Practical significance: A concept whose time has come. *Educational and Psychological Measurement* 56 (5): 746–759.

Leamer, E.E. 1985. Sensitivity analyses would help. *American Economic Review* 75 (3): 308–313.

Lewin, A.Y., C.Y. Chiu, C.F. Fey, S.S. Levine, G. McDermott, J.P. Murmann, and E. Tsang. 2016. The critique of empirical social science: New policies at *Management and Organization Review*. *Management and Organization Review* 12 (4): 649–658.

Lexchin, J., L.A. Bero, B. Djulbegovic, and O. Clark. 2003. Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *British Medical Journal* 326 (7400): 1167–1170.

Masicampo, E.J., and D.R. Lalande. 2012. A peculiar prevalence of *p*-values just below 0.05. *Quarterly Journal of Experimental Psychology* 65 (11): 2271–2279.

McCloskey, D.N. 1985. The loss function has been mislaid: The rhetoric of significance tests. *American Economic Review* 75 (2): 201–205.

McCloskey, D.N., and S.T. Ziliak. 1996. The standard error of regressions. *Journal of Economic Literature* 34: 97–114.

Meyer, K.E. 2006. Asian management research needs more self-confidence. *Asia Pacific Journal of Management* 23 (2): 119–137.

———. 2009. Motivating, testing, and publishing curvilinear effects in management research. *Asia Pacific Journal of Management* 26 (2): 187–193.

Misangyi, V.F., and A.G. Acharya. 2014. Substitutes or complements? A configurational examination of corporate governance mechanisms. *Academy of Management Journal* 57 (6): 1681–1705.

Mullane, K., and M. Williams. 2013. Bias in research: the rule rather than the exception? *Elsevier Journal.* http://editorsupdate.elsevier.com/issue-40-september-2013/bias-in-research-the-rule-rather-than-the-exception. Accessed 23 Mar 2017.

New York Times. 2011. *Fraud case seen as a red flag for psychology research*. November 2. http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html?-r=1&ref=research. Accessed 15 Jan 2017.

Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science*. https://doi.org/10.1126/science.aac4716.

Orlitzky, M. 2012. How can significance tests be deinstitutionalized? *Organizational Research Methods* 15 (2): 199–228.

Pashler, H., and E.-J. Wagenmakers. 2012. Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science* 7 (6): 528–530.

Peterson, M., J.L. Arregle, and X. Martin. 2012. Multi-level models in international business research. *Journal of International Business Studies* 43 (5): 451–457.

Pfeffer, J. 2007. A modest proposal: How we might change the process and product of managerial research. *Academy of Management Journal* 50 (6): 1334–1345.

Popper, K. 1959. *The logic of scientific discovery*. London: Hutchinson.

Reeb, D., M. Sakakibara, and I.P. Mahmood. 2012. From the editors: Endogeneity in international business research. *Journal of International Business Studies* 43 (3): 211–218.

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin* 86 (3): 638–641.

Rosnow, R.L., and R. Rosenthal. 1984. *Understanding behavioral science: Research methods for customers*. New York: McGraw-Hill.

Rothstein, H.R., A.J. Sutton, and M. Borenstein. 2005. *Publication bias in meta-analysis, prevention, assessment and adjustment*. New York: Wiley.

Sala-i-Martin, X. 1997. I just ran two million regressions. *American Economic Review* 87 (2): 178–183.

Shadish, W.R., T.D. Cook, and D. Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference*. New York: Houghton Mifflin.

Simmons, J.P., L.D. Nelson, and U. Simonsohn. 2011. Falsepositive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science* 22 (11): 1359–1366.

Sterling, T.D. 1959. Publication decision and their possible effects on inferences drawn from tests of significance – Vice versa. *Journal of the American Statistical Association* 54 (285): 30–34.

van Witteloostuijn, A. 2015. Toward experimental international business: Unraveling fundamental causal linkages. *Cross Cultural & Strategic Management* 22 (4): 530–544.

———. 2016. What happened to Popperian falsification? Publishing neutral and negative findings. *Cross Cultural & Strategic Management* 23 (3): 481–508.

Wasserstein, R. L., and N. A. Lazar. 2016. The ASA's statement on *p*-values: Context, process, and purpose. *American Statistician,* 70(2): 129–133. http://amstat.tandfonline.com/doi/full/10.1080/00031305.2016.1154108. (ASA = American Statistical Association).

Wiersema, M.F., and H.P. Bowen. 2009. The use of limited dependent variable techniques in strategy research: Issues and methods. *Strategic Management Journal* 30 (6): 679–692.

Williams, R. 2012. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal* 12 (2): 308.

Wonnacott, T.H., and R.J. Wonnacott. 1990. *Introductory statistics for business and economics*. New York: Wiley.

Zedeck, S. 2003. Editorial. *Journal of Applied Psychology* 88 (1): 3–5.

Zellmer-Bruhn, M., P. Caligiuri, and D. Thomas. 2016. From the editors: Experimental designs in international business research. *Journal of International Business Studies* 47 (4): 399–407.

Zelner, B. 2009. Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management Journal* 30 (12): 1335–1348.

5

What's in a *p*? A Commentary

Agnieszka Chidlow, William Greene, and Stewart R. Miller

Introduction

The *JIBS* editorial by Meyer, Witteloostuijn and Beugelsdijk (2017) presents a useful and readable guide for researchers, not only to emphasize the need for transparency and replicability of empirical findings, but also to

Meyer, K., A. van Witteloostuijn, and S. Beugelsdijk. 2017. What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies*, 48 (5): 535–551.

A. Chidlow (✉)
The University of Birmingham, Birmingham, UK
e-mail: A.Chidlow@bham.ac.uk

W. Greene
New York University, New York, NY, USA
e-mail: wgreene@stern.nyu.edu

S. R. Miller
University of Texas at San Antonio, San Antonio, TX, USA
e-mail: stewart.miller@utsa.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_5

lay out a very important path for responsible empirical practice within the international business (IB) community. The editorial should be taken seriously. We applaud the authors for preparing (in our view) such a long-overdue critique of the statistical elephant in the room in order to allow future scholarship to develop a more rigorous body of empirical research.

As we reflect on Meyer et al.'s (2017: 536) noble efforts to address "the recent surge of pleas to change the extant research practices," we comment on three of the best practices proposed in their work in order to further reinforce the trustworthiness of IB scholarship within social science research: (a) we examine a false sense of security with software advances offered to scholars, (b) we propose the use of existing criteria for selecting a statistical level of significance and (c) we point out some ethical issues associated with scholarly research.

A False Sense of Security About Modern Software

Without doubt, advances in modern analytical software, such as SPSS and STATA, have elevated the technical level of empirical social science research. They have provided access to new models and methods, automated the graphical and numerical analysis of effect sizes in models of all types, and enabled more intricate and precise uses of models including subtle analyses of effect sizes and interactions.

All of these features provide researchers with extensive resources to extract the most detailed available information in social science data with greater precision than conventional first-order descriptive statistics such as cross tabulations. This is very welcome. However, this computing power comes with a greater need for a detailed theoretical understanding on the part of researchers. For example, effect sizes and models with interaction terms are often complicated functions of more than one estimate. The idea of "statistical significance" needs to be considered very carefully in this case.

As new estimation techniques compel a significant level of theoretical expertise on the part of researchers, the very common tendency among

scholars is to outsource such expertise to the software developers. A ubiquitous example of such practice is "robust inference," where simple push button implementations—without any description of "why" or "how" such computations are done—often betray a shallow understanding of "robust" standard errors. (They do not redeem a flawed estimator!) Another example of such conduct is the many applications of the linear probability "model." Finally, there is the ubiquitous, compulsive convention to "cluster" on *something*, even in cross sections.

The ease with which software developers are able to layer thinly motivated options on important new tools raises the possibility of uninformed or, at worst, misleading post-estimation analysis. It is very important that researchers be able to distinguish informative new tools from bells and whistles when documentation presents them as if they were equally important. With open source technologies, estimation techniques can gain currency *because* they are associated with widely known statistical platforms. An example is the questionable "fixed effects vector decomposition (FEVD)" estimator that gained traction when and, arguably, because it was released as a user provided procedure in STATA (Breusch et al. 2011; Greene 2011).

Criteria for Selecting a Statistical Level of Significance

We strongly agree with Meyer et al. (2017: 537) that researchers should steer away from the "old asterisks habit" and instead report the obtained statistical level of significance in their models. This is because the classical statistical theory does not provide a set of rules for selecting the levels of 0.05, 0.01 and 0.001 (Fisher 1925; Gibbons and Pratt 1975; Greene 2018) due to the different nature and type of the research problems under investigations as well as the fact that the selection of a significance level is a complex process.

In order to assist IB researchers with determining the appropriate statistical level of significance we recommend the work of Skipper, Guenther and Nass (1967) and Stanford (1968) where the authors propose a

number of criteria pertaining to the selection of significance levels and how to report them. Skipper et al. (1967) strongly encourage scholars to think and reflect on the arbitrary nature of the conventionally accepted levels of significance because different classes of research may require different levels of alpha to report the actual levels obtained and to have an opinion if the obtained levels support (or not) tested hypotheses. Stanford (1968) goes further and urges scholars to: (a) take into consideration the practicality of the problem under examination and the gravity of errors available on the basis of value orientation; (b) consider the rational and empirical evidence from other studies when interpreting a significance level in order to avoid testing hypotheses in isolation; (c) carefully examine the relationship between the power of the test and the sample size as well as the size of the true difference; (d) think carefully about the robustness of the tests used, the degree of control in design as well as the confidence interval; and (e) be clear about both Type I and Type II errors in hypotheses testing as these (to some extent) vary inversely with one another. Even though such criteria should not be considered as definitive in any sense they could certainly give a good starting point to probe a deeper understanding into the rationale behind significance levels and undoubtedly assist scholars in diminishing the use of the "conventional wisdom" standards of 0.05, 0.01 and 0.001 as proof of hypotheses.

Ethical Issues Associated with Scholarly Research

Concerns stemming from HARKing and *p*-hacking, as pointed out by Meyer et al. (2017: 538), have increased scholarly awareness of "publication ethics" and highlighted "the pressure to adhere to proper statistical standards and enhance transparency." Their article has boosted the post-publication detection of poor practices within the business and management scholarly communities (Eden et al. 2018; Aguinis et al. 2017; Bergh et al. 2017) including significant ethical problems such as star chasing behaviour. Such actions, unfortunately, incentivize a range of common ethical practices across various social science disciplines; examples include

fudging coefficients and/or standard errors, skewing the data collection process, dropping unfriendly observations, excluding key control variables (John et al. 2012), selective or distorted reporting (Ioannidis 2005; Murphy and Aguinis 2019), and failure to perform the requisite specification diagnostics that reveal data issues (i.e. outliers and a serial correlation).

Questionable scholarly practices are driven not only by the publish-or-perish culture where the reward structure promotes practices inconsistent with those of statistical best practices (Wasserstein and Lazar 2016) but can also be learned from others. Propagating and endorsing questionable practices has a lasting effect on impressionable PhD students, who, after all, represent the next generation of scholars. As stated in Meyer et al. (2017), pressures to publish as PhD students and then as assistant professors may reinforce learned questionable practices. For example, dissertations based on a system of equations that, in turn, are sliced into multiple papers may fail to account for endogenous variables relationships in the analysis, which produces biased p-values. These concerns with unethical behaviours in research have contributed to a movement for "Responsible Research in Business and Management" that calls for sound methodology (https://rrbm.network/position-paper/principles-of-responsible-science/).

Conclusions

Building on Meyer et al. (2017) and our comments, the recommendation that follows from this commentary is that empirical models developed and tested by IB scholars should be complex enough to employ the useful modern technology but at the same time simple enough to be comprehensible—but not too simple. Credible research design should be executed with persuasive and authoritative precision.

In our opinion, there is no, "right" or "wrong" level of significance in hypotheses testing. A more rational approach ought to be used to report the actual level of significance by placing the burden of interpretive skill on the researcher. Such a policy would encourage scholars to give higher

propriety to selecting appropriate levels of significance for a given problem and would discourage the misleading culture of the "old asterisks habit."

What is more, any unethical research practices relating to HARKing and *p*-hacking need to be questioned and discouraged not just by journal "gate keepers" but also by PhD supervisors and co-authors. After all, these forms of behaviour adversely affect the credibility and integrity of the entire scholarly community.

References

Aguinis, H., W. Cascio, and R. Ramani. 2017. Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies* 48: 653–663.

Bergh, D., B. Sharp, H. Aguinis, and M. Li. 2017. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization* 15: 423–436.

Breusch, T., M. Ward, and H. Minh. 2011. FEVD: Just IV or just mistaken. *Political Analysis* 29: 165–169.

Eden, L., K. Lund Dean, and P. Vaaler. 2018. *The ethical professor: A practical guide to research, teaching, and professional life*. New York: Routledge.

Fisher, R. 1925. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.

Gibbons, J., and J. Pratt. 1975. P-values: Interpretation and methodology. *The American Statistician* 29: 20–25.

Greene, W. 2011. Fixed effects vector decomposition: A magical solution to the problem of time-invariant variables in fixed effects models. *Political Analysis* 19: 135–146.

———. 2018. *Econometric analysis*. Upper Saddle River: Pearson Education Inc.

Ioannidis, J. 2005. Why most published research findings are false. *PLoS Medicine* 2: 696–701.

John, L., G. Loewenstein, and D. Prelec. 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science* 23: 524–532.

Meyer, K., A. van Witteloostuijn, and S. Beugelsdijk. 2017. What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies* 48: 535–551.

Murphy, K., and H. Aguinis. 2019. HARKing: How badly can cherry picking and question trolling produce bias in published results? *Journal of Business and Psychology* 34: 1–17.

Skipper, J., Jr., A. Guenther, and G. Nass. 1967. The sacredness of .05: A note concerning the uses of statistical levels of significance in social science. *The American Sociologist* 3: 16–18.

Stanford, L. 1968. Criteria for selecting a significance level: A note on the sacredness of .05. *The American Sociologist* 3: 220–222.

Wasserstein, R., and N. Lazar. 2016. The ASA's statement on p-values: Context, process, and purpose. *American Statistician* 70: 129–133.

Part IV

Trusting Findings

6

From the Editors: Can I Trust Your Findings? Ruling Out Alternative Explanations in International Business Research

Alvaro Cuervo-Cazurra, Ulf Andersson, Mary Yoko Brannen, Bo Bernhard Nielsen, and A. Rebecca Reuber

A. Cuervo-Cazurra (✉)
Northeastern University, Boston, MA, USA
e-mail: a.cuervocazurra@neu.edu

U. Andersson
Malardalen University, Vasteras, Sweden

BI Norwegian Business School, Oslo, Norway
e-mail: ulf.r.andersson@mdh.se

M. Y. Brannen
Copenhagen Business School, Frederiksberg, Denmark

University of Victoria, Victoria, BC, Canada
e-mail: myb.egb@cbs.dk

B. B. Nielsen
The University of Sydney Business School, The University of Sydney, Sydney, NSW, Australia

Copenhagen Business School, Frederiksberg, Denmark
e-mail: bo.nielsen@sydney.edu.au

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_6

A. R. Reuber
University of Toronto, Toronto, ON, Canada
e-mail: reuber@rotman.utoronto.ca

Introduction

International business (IB) phenomena provide new opportunities for identifying interesting and important relationships that are often overlooked in other studies, adding many dimensions of complexity to the research we conduct. This additional complexity emerges from several sources: from the cross-border relationships that organizations engage in as they have to deal with differences in economic, political, social, and geographic conditions; from the cross-country comparison of relationships that take into account additional variation in how the environment shapes relationships; and from the inclusion of the country level of analysis that alters relationships at lower levels of analysis. Previous editorials have explained how to deal with some of these issues by, for example, providing suggestions on how to: explain interaction effects within and across levels of analysis (Andersson et al. 2014; Cortina et al. 2015), address multilevel challenges (Peterson et al. 2012), solve endogeneity problems (Reeb et al. 2012), improve qualitative research (Birkinshaw et al. 2011), address common method challenges (Chang et al. 2010), improve the theoretical identification of relationships (Bello and Kostova 2012; Thomas et al. 2011), and more generally how to benefit from, and deal with, the inherent interdisciplinary nature of IB research (Cantwell and Brannen 2011; Cantwell et al. 2014; Cheng et al. 2009; 2014).

We build on these ideas and focus on providing a better understanding of how to ensure that the findings coming out of empirical studies are trustworthy, i.e., "worthy of confidence" (Merriam-Webster Dictionary 2016; see also Lapan and deMar-rais 2003). Ensuring that the relationships identified in an empirical study are trustworthy is important in IB studies in particular and in management studies in general, because there is a very limited tradition of replicability that can help uncover researchers' biases and differences in empirical techniques (Bettis et al. 2016; Silberzahn and Uhlmann 2015). This limited replicability is due to several reasons. Many of the samples are proprietary and closely guarded by researchers

who want to extract the maximum number of publications out of their data collection effort. Even if the datasets are not proprietary, the specific samples may be difficult to replicate because researchers do not make their samples available to others, in contrast to studies in economics. Further, replication tends to be discouraged from publication in leading journals, including *Journal of International Business Studies* (JIBS), which prioritize novelty in ideas and analyses. In the absence of replication, each paper in and of itself has to demonstrate that it is worthy of confidence.

In doing so, a key task in empirical papers is ruling out alternate explanations for the phenomena under investigation. It is easy for this step to be neglected. Authors are encouraged to increase the storytelling nature of their articles (e.g., Haley and Boje 2014; Pollock and Bono 2013) and this includes developing a straightforward, accessible story line (Ragins 2012). It can be tricky to introduce the possibility of alternative explanations without deviating from the plot of the narrative. However, it is necessary to do so for a paper to be considered trustworthy.

The objective of this editorial is to provide guidance to help IB scholars address alternate explanations in their empirical manuscripts, and ensure that they have identified the correct relationships and mechanisms so that readers can place higher trust in their findings. The editorial is organized in two parts that address the particular challenges of two distinct empirical traditions: qualitative and quantitative. Part A deals with qualitative research methods. It discusses multiple and integrated techniques to strengthen readers' belief that the explanations arising from the analysis of one or few cases are the correct ones and not subject to alternative influences that emerge from data limitations or the inherent biases in the minds of the researchers. Part B deals with quantitative studies. It discusses large sample studies that test whether theoretically-derived relationships hold on a large number of individuals, teams, organizations, or countries. It provides suggestions on how to control for alternative explanations not only in the analyses of data, which has been the usual focus of the discussions of controls in quantitative studies, but also in the theoretical explanation of the hypothesized relationships as well as in the research design.

In discussing qualitative and quantitative research separately in two distinct sections, we recognize that we are inviting at least two types of criticism. The first is that there is no objective, clear-cut delineation between qualitative and quantitative research. As Small points out "the quantitative

versus qualitative opposition has been used to contrast many kinds of alternative studies: largen versus small-n, nomothetic versus idiographic, causal versus interpretive, variable-based versus case-based, explanatory versus descriptive, probabilistic versus deterministic, and numerous others" (Small 2011: 59). We agree that the two categories are not mutually exclusive. Further, we agree with Small that "qualitative" and "quantitative" can refer independently to data, to data collection, and to data analysis, which renders the binary classification of many studies difficult. For example, with increased accessibility of electronic text and software tools such as sentiment analysis, there is greater quantification of qualitative data (Kaplan 2015). However, we retain the distinction here for clarity of exposition, and for each section of the editorial we draw on scholarly authorities that are unambiguously about qualitative or quantitative research.

The second criticism we invite is that the relevant issues associated with the two research traditions are very different; in other words, we are discussing apples and oranges with only a loose connection under the umbrella of trustworthiness. However, we have two reasons for combining both research traditions in a single editorial. The first reason is to highlight the importance of being open minded. Too often scholars dismiss research that does not conform to their expected standards of analysis, and this is in part because of a natural tendency of paying attention to what one is familiar with. The current JIBS editorial team values theoretical and methodological pluralism to promote complementary ways to address new and difficult research questions and enhance the overall development of the field. The second reason is the value that one can gain from better understanding an alternative research tradition. We want to emphasize the value of comparing and contrasting research traditions next to one another. Although the specifics differ, both traditions face challenges in ensuring the identification of findings that other researchers can trust. Authors should recognize that reviewers will include experts in the methods they use, who will prioritize method-specific standards, but they are also likely to get reviewers who may be more familiar with other types of empirical methods. It is imperative that authors can explain to this second group of reviewers how they are establishing trustworthiness through the methods-related choices they make. An IB scholar keeping up with the literature needs to understand how trustworthiness is estab-

lished for empirical approaches they may have little experience with. Thus both parts of this editorial are relevant to IB scholars even if they self-identify with only one of the research traditions covered.

Part A: Producing Trustworthy Qualitative International Business Research

Research based on qualitative data has played a long and illustrious role in IB (Birkinshaw et al. 2011); yet, the proportion of qualitative research appearing in JIBS is lower than this track record might warrant. There are probably several interrelated reasons for this. There are few submissions of qualitative papers to JIBS. There is limited training in many PhD programs in qualitative methods, and so researchers may lack familiarity with them. There is a lack of established standards for analysing and presenting data (e.g., Bansal and Corley 2012; Pratt 2008), which makes the research process seem uncertain. It is time-consuming to embark on the long journey involved in collecting and analysing qualitative research, such as gaining access to research sites, conducting interviews, and analysing interview transcripts and documents. On top of all of this, there is the language challenge: primary data from interviews and participant observation often need to be conducted in more than one language, transcriptions must be done by a native speaker and at some point translated into English for publication in JIBS, and assuring meaning congruence and functional equivalence of terms is challenging.

In addition to these supply-based reasons, we believe that a factor constraining the publication of qualitative research papers is that they are having difficulty getting through the review process successfully. While the nature of the difficulties vary, we have noticed that a weakness common to many qualitative research submissions is that the authors have not paid sufficient attention to demonstrating the trustworthiness of their research. To address this, we provide guidelines as to how qualitative researchers in IB can establish this trustworthiness in their manuscripts.

At the outset we note that researchers wishing to use qualitative methods have many resources from which to draw inspiration. There was a JIBS Special Issue on Qualitative Methods (Birkinshaw et al. 2011), and

there have been recent JIBS articles on qualitative methods in general (e.g., Doz 2011) and on specific topics related to qualitative methods such as longitudinal historical research (Burgelman 2011), grounded theory (Gligor et al. 2016), case-based research (Welch et al. 2011), and ethnography (Westney and Van Maanen 2011). There are articles in other journals on topics particularly relevant to IB, such as process-based research (e.g., Langley 1999; Welch and Paavilainen-Mäntymäki 2014) and there are classic texts such as Corbin and Strauss (2008), Glaser and Strauss (2011), Miles and Huberman (1994), Marschan-Piekkari and Welch (2011), Piekkari, Welch and Paavilainen (2009), Van Maanen (1998) and Yin (2009). We encourage authors to consult these and other resources when they are making research design and analysis decisions, and to use them to justify these decisions when reporting research results in their manuscripts.

Our intention in this editorial is to highlight the importance of making explicit and consistent choices in order to establish trustworthiness in a qualitative manuscript submitted for publication. This requires rigor from the start of a research project, because the conceptualization and design of a project influences the nature of the analysis that can be undertaken, and therefore the findings that constitute a scholarly contribution to the field. There are three well-known paths that are unlikely to lead to successful outcomes. One such path is converting a teaching case into a research case, which is problematic because a teaching case will rarely have the theoretical relevance and the rich data required of a research case. A second questionable path can occur in situations where it is difficult to collect data from a sample large enough to establish statistical significance, and so a researcher collects data from several companies and attempts to establish generalization by showing that multiple companies are engaged in the same strategies. This use of case studies is an example of a theoretical contribution that small n studies cannot make. Small n studies cannot make frequency-based insights, such as the propensity to engage in a particular firm behaviour, because the frequency observed is highly dependent on the particular cases selected for examination. Moreover, small n studies can rarely explain outcomes such as performance, which are affected by many factors, because they cannot control for these factors as can large-scale quantitative studies. Finally, a third

questionable path is "convenient sample driven" research, or "squat ethnography" (Van Maanen 1998), where a researcher has access to a subject (individual, team, company, country) and starts collecting data. Once collected, the researcher starts analysing the data and thinking about what to do with it, hoping to have a *eureka* moment in which something that seems to be different emerges from the data. This approach tends to be justified with an argument along the lines of "with an open mind and with no prior biases I studied company x to be able to identify new patterns." However, such an approach mistakes having an open mind with having no clue about what to do!

None of these paths are likely to result in a trustworthy manuscript. Instead, trustworthiness needs to be built into the start of a manuscript and maintained consistently throughout it. We next provide some guidance as to how this can be done in the research context, research design and empirical analysis. Table 6.1 summarizes the ideas presented in these discussions.

Trustworthiness in Research Context

Qualitative methods are inherently embedded in context and so it is critical that the context of studies based on qualitative methods be explicitly defined. The type of context that is relevant to one study may be different from the type of context relevant to another study – for example, it could be an event, a type of environment, or a particular situational strength (Johns 2006) – but it is important that the contextual nature of the research be consistent across all aspects of the manuscript – the research question, the literature review, methodological choices and the theoretical interpretation of the findings. If this is done effectively, then the contextual delineation of the study bounds the theoretical claims that can be made, thereby providing clarity around what is and what is not explained.

Because context is so central to the theoretical and empirical aspects of qualitative research, it is incumbent on authors to justify the particular context they are studying. At a basic level, authors should consult the JIBS Statement of Editorial Policy, which describes the meaning of IB with respect to submissions to the journal. Beyond this, it is advanta-

Table 6.1 Recommendations for establishing trustworthiness in qualitative research

Dimension	Focus of attention	Question to answer	Checklist
Research context	Delineation of the boundaries of the theoretical claims	How is the theory contextualized?	Describe the underlying context in which the research question is embedded Justify why this context is interesting and relevant for international business scholars Clarify what is known and what is not known about the phenomenon under investigation in this context, in terms of extant theoretical arguments and empirical findings Identify the reasons for using qualitative methods and the expected outcomes of the research in theoretical terms
Research design	Theoretical rationale for selecting this site or case(s)	Why has this site or sample been chosen for study?	Justify a single case in terms of the theoretical insights it can provide Specify the theoretical bases on which the case(s) was chosen
	Data replication	What is the basis for assurance that the findings are based on multiple observations?	Specify the ways in which the data are replicated

(*continued*)

Table 6.1 (continued)

Dimension	Focus of attention	Question to answer	Checklist
	Data triangulation	What is the basis for assurance that the findings are based on multiple data sources?	Specify the nature of the different types of data collected and how each was collected Ensure that you show how each type of data is relevant and used in the analysis
Empirical analysis	Multilingual and multicultural boundaries	What is the basis for assurance that the analysis overcomes linguistic and cultural barriers?	Specify how you overcame linguistic and cultural barriers in interpreting the data collected
	Clarity of analysis	What is the basis for assurance that the findings are based on a rigorous, unbiased analysis?	Describe the analysis process in detail and show examples of work products such as coding schemes Identify and explain "negative" cases in the data: Cases that do not confirm to dominant patterns
	Reporting both evidence and theory	How do the findings reflect the empirical evidence as well as new theoretical constructs?	Decide what to show vs. what to tell Transcend description to reveal new theoretical constructs Develop convincing displays of the data and the theory

geous for authors to show that the specific context they are studying is theoretically interesting and relevant to current scholarly IB conversations.

In many qualitative studies, the motivation to study a particular context is based on observations of real world phenomena. For example, Brannen and Peterson (2009) justify their study of a Japanese acquisition

in the US by highlighting the high failure rate of cross-border mergers and acquisitions and the lack of theory to explain them. In the absence of prior theory, such as this, it is difficult to develop hypotheses to be tested in a large scale study, and so inductive, qualitative methods are used to generate or create theory (Edmondson and McManus 2007). Sometimes, however, the motivation to select a particular context is based on prior research and the questions it leaves unaddressed. For example, Jonsson and Foss (2011) justify their study of the Swedish furniture retailer IKEA by noting that although scholars understand the trade-offs between replication (scale) and local adaption, little is known about the processes through which both can be accomplished. It is interesting to note that in both of these papers, the context is just one organization. That is not always the case in qualitative studies, of course. For example, Caprar's (2011) study of the culture of local employees of MNEs is based on focus groups of employees of American MNEs in Romania. He frames this choice of context as relevant to culture – the key theoretical construct – since Romanians are both welcoming of foreign investment and sufficiently culturally distant from Americans to be theoretically interesting.

These examples illustrate that in justifying a research context, it is important to clarify what is and what is not known about the phenomena under investigation, and to be explicit about why a qualitative research approach is used. The first task, positioning a scholarly paper in prior literature, is beneficial regardless of the empirical method. However, Pratt (2008) points out that a particular challenge for qualitative researchers is to manage the tension between recognizing and drawing on existing theory, while also distancing from it to show that new theory has been generated. He suggests developing open theoretical frameworks that describe prior research while highlighting where prior research has been largely silent, in order to create a new space for an author's contribution. In creating these boundaries between what is known and what is not yet known, an author can credibly signal that alternative explanations for the paper's findings are unlikely.

The second task, justifying the use of qualitative methods, is important in conveying the overall theoretical objectives of the research. While articulating an explicit research question is beneficial in conveying the specific focus of the research, communicating the nature of the findings

in theoretical terms helps readers to follow the thread of the storyline. Are you using qualitative methods to extend theory in a particular direction or are you building new theory? Are you generating variance theory or process theory (Langley 1999)? Are you intending to develop testable propositions or reveal new interpretations of theoretical constructs or relationships? An important dimension of communicating the nature of your findings is being precise with respect to the outcome you are explaining; for example, learning processes within MNEs (e.g., Jonsson and Foss 2011), variation in SME internationalization practices (e.g., Lamb et al. 2011) or variation in managerial narratives (e.g., Haley and Boje 2014). Since choices among these theoretical objectives are connected with choices related to research design, empirical analysis and reporting of findings, expressing them clearly and early in the paper helps the reader understand the subsequent choices you make. This consistency therefore enhances the trustworthiness of the explanations offered as theoretical contributions.

Trustworthiness in Research Design

We have come across misperceptions that research based on quantitative data and deductive reasoning is empirical research, while research based on qualitative data and inductive reasoning is conceptual research. These perceptions are wrong. Both are empirical studies and in both the quality of the research design is crucial for establishing trustworthiness. Moreover, it is crucial to check for data quality in qualitative research, because there are no statistical tests to provide assurances about the operationalization of theoretical constructs and the strength of the relationships among them.

Three aspects of the design of qualitative research can substantially influence perceptions of its trustworthiness: site selection, data replication, and data triangulation. First, with respect to site or sample selection, the researcher needs to justify how and why they chose a single site (one case), or how and why they constructed a sample of multiple cases, such as individuals, teams, organizations, events, regions or countries. Whether one case or a sample of cases is selected, the basis of selection needs to be tightly coupled with the theoretical context of the study and the interpre-

tation of its findings in order for the choice to be seen as trustworthy. Single cases can be justified because they are extreme, unique, representative, revelatory or longitudinal (Yin 2009: 47–49) and it is important to embed the justification in the theoretical contribution of the paper. As Siggelkow (2007) points out, it is easier to justify a special case than a representative case because you can show that it was selected to allow you to gain insights that other cases would not provide. For example, in order to reveal insights about the liability of foreignness, Brannen (2004) chose the US entertainment firm Walt Disney Company as a research site because it was an extreme case of paradoxes regarding foreignness. When the objective is to investigate variance, it is important to justify the selection of several cases on the basis of theoretical diversity, so individual cases can serve as replications, contrasts and extensions to the emerging theory (Eisenhardt and Graebner 2007). For example, Lamb et al. (2011) wanted to capture the greatest possible variation in small firm internationalization and so they justified their cases by emphasizing that they reflected a variety of international experiences and histories within and across different wine export networks that helped better understand internationalization. In the field of IB it is not unusual to combine a single site with a theoretical diverse sample within that site. For example, Jonsson and Foss (2011) chose IKEA as a site because it exhibits a unique combination of format standardization and local adaptation, but to investigate variance in learning within IKEA, they interviewed employees in three markets (China, Japan and Russia) whose differing degrees of development were likely to be associated with variance in learning.

A second aspect of research design that influences the trustworthiness of a manuscript is data replication. Replication adds credibility to findings because it provides support that they are deeply grounded in diverse empirical evidence and not idiosyncratic to one particular case (Eisenhardt and Graebner 2007). As we have already pointed out, including multiple cases (interviewees, firms) in a sample provides replication. Researchers can also provide replication by collecting data more than once. For example, in Caprar's (2011) study of the culture of local employees, he conducted three focus groups, varying their composition and timing in order to be able to assess whether these factors impacted the findings. In process studies, replication can be provided through data collected on

multiple observations longitudinally (Langley et al. 2013). For example, Bingham (2009) captured data on processes associated with multiple foreign entries over time. In this case, the study was designed with replication across organizations (cases) and within organizations (entries), but longitudinal data collection can also provide within-case replication when the study is based on a single organization. In ethnographies, which are designed specifically to describe and understand how groups of individuals (cultures) function; their norms and patterns of behaviour, values and basic assumptions, replication is characterized by its continuous nature. The research outcomes of ethnography are detailed narrative accounts of cultural phenomena told as much as possible from the native's point of view, and so participant observation is a key aspect of the methodology. The ethnographer needs to find a role within the group under observation from which to participate in some manner, even if only as "outside observer." Participant observation, therefore, is limited to contexts where the community under study understands and permits it. Further, since the ethnographer's aim is to understand predominantly tacit, complex, contextually embedded, existential phenomena, the amount of time spent in the field must be substantial – to an anthropologist this means at least 1 year, though a year may be too brief if the research involves learning or perfecting a new language on the part of the researcher. Thus, rather than being characterized by discrete replications, ethnographic research is characterized by diverse and continuous data collection and it is important for the ethnographer to describe in detail both the research data and how data collection took place. For example, in studying a Japanese acquisition of an American manufacturing plant, Brannen and Peterson (2009) provide a rich description of the plant before and after the acquisition, as well as the nature of their participant observation activities and other data collection techniques that were used.

A third element of research design that enhances the trustworthiness of a manuscript is data triangulation. It is common for authors to state that they have supplemented interviews with archival data about the entities they study, but positioning such data as supplemental detracts from their credibility. If the data are not relevant to the analysis and the findings, it is preferable to leave them out of the discussion. It is rare when authors show how they incorporated diverse types of data in their analysis. If the

data are relevant, it is important to justify both how they were collected and how they were used. For example, in their study of MNE's storytelling, Haley and Boje (2014) describe their diverse data sources – including onsite observation, interviews, videos, TV commercials, and transcripts of legal disputes – and weave all of these into their discussion of the study's findings. Likewise, in their study of Englishisation in the provision of cross-border services, Boussebaa, Sinha and Gabriel (2014) carefully detail and justify collecting interview data from different types of employees, as well as data from internal documents, company intranet pages and onsite observation. In discussing their findings, they are able to deepen their interpretation of interview data by portraying it in conjunction with the company's human resources policies and with the physical work set-up that they observed.

Trustworthiness in Empirical Analysis

The empirical analysis of qualitative data can be enhanced, and thus the confidence of the scholarly IB community in the interpretation of the data presented, in three ways: navigating multilingual and multicultural boundaries, establishing clarity in the analysis, and reporting both evidence and theory and the links between the two.

First, multilingual and multicultural boundaries are particularly prevalent in the field of IB because much of the scholarly inquiry crosses national, cultural or linguistic lines. It is important for researchers to show how they navigate such boundaries effectively, because accurate data interpretation is so important in establishing the credibility of qualitative research findings. This navigation involves accurate translation of documents and interview transcripts. However, most qualitative IB researchers do not discuss their translation decisions in their manuscripts, even though there are substantial theoretical differences among approaches to translation (Chidlow et al. 2014). It also involves an intimate knowledge of the cultural milieus being examined, both to be sufficiently accepted to be able to collect meaningful data and to be sufficiently acclimatized to be able to interpret that data. This is often achieved by ensuring that someone on the research team has the required language skills and cultural familiarity.

Second, with respect to providing a clear analysis, authors can be overwhelmed by the quantity of data to be analysed and by the lack of prescriptions for how the analysis should be conducted, and for this reason they need to pay particular attention as to how to analyse data in the most effective way. In contrast to quantitative studies, in qualitative studies there are no standard formats for discussing the methods and findings sections (e.g., Bansal and Corley 2012; Pratt 2008). However, this does not mean that any approach for analysing data is valid. Indeed, qualitative researchers are recognizing that there are templates for distinct styles of qualitative research (e.g., Gioia et al. 2012; Langley and Abdallah 2011). Regardless of the type of analysis used, it is important that the reader understand in detail what was done and why. Too often, manuscripts go from a description of the sample to a description of the findings and provide little detail on how data were analysed. One way to show how data analysis was conducted is to show examples of work products, such as the coding schemes developed. This not only helps increase confidence in the analysis, but can also help other researchers improve their own research designs.

While data analysis in qualitative studies tends to be focused on identifying dominant patterns in the data, it is also important to recognize that there may be "negative cases" (Corbin and Strauss 2008: 84); i.e., cases that do not fit the dominant pattern. These are important to acknowledge and explain. Rather than detracting from a study's credibility, they can signal analytic rigor because rarely are dominant patterns universal. Moreover, negative cases can provide an opportunity to deepen the theoretical claims that are being made by taking exceptions into account.

Third, deciding how to report the findings of a qualitative study can be challenging, because there are no standardized tables that are expected, and because qualitative data do not always lend themselves to being summarized. One of the key issues that an author faces is deciding what to show and what to tell (Pratt 2009). Focusing on showing the data (the evidence for theoretical claims) can make the paper seem overly descriptive, while focusing on telling about the data (the theoretical interpretations) can make the theory seem unsubstantiated. Successful qualitative researchers address this difficulty by coming up with creative ways to

display their data (Bansal and Corley 2012). It is important for the reports of the findings to transcend description and indicate clearly the new theory that was generated from the investigation.

Towards More Trustworthy Qualitative Manuscripts

In Part A of this editorial we have provided suggestions for how IB scholars can enhance readers' confidence in research findings that are based on qualitative data. Scholarly insights are more trustworthy when they take into account extraneous factors that may have affected research results. As is discussed in Part B, on controls in large sample quantitative studies, the ruling out of alternative explanations is handled by controlling for them. In qualitative research, however, the likelihood and magnitude of alternative explanations cannot be measured. Instead, as we have explained, there are multiple and integrated mechanisms to strengthen a reader's belief that the explanations presented in a qualitative research study are accurate and valid. These mechanisms include ensuring that the boundaries of the theoretical claims are delineated, the research site is appropriate, the data are rich and robust and there is transparency in data analysis and the interpretation of the findings. Moreover, it is important that there be coherence and consistency across these mechanisms so that the thread from theoretical purpose to method to findings to theoretical contribution is clearly visible and easy to follow. We hope that these suggestions are useful for producing more sophisticated and trustworthy qualitative studies.

Part B: Using Controls in International Business Research

Trusting the findings from empirical analyses has a longer tradition and there are already several JIBS editorials that have analysed ways to handle the analysis of large samples (e.g., Andersson et al. 2014; Cortina et al. 2015; Peterson et al. 2012; Reeb et al. 2012). To complement and extend these ideas, in this editorial we analyse how to use controls in IB. Controls

are particularly important in quantitative IB research, which is characterized by analysing complex phenomena, often spanning multiple disciplines, theories and levels of analysis. The study of cross-border phenomena not only adds an additional layer of country-level influences to the relationships, but can also modify how such relationships operate as new mechanisms emerge that alter existing arguments (Andersson et al. 2014; Cortina et al. 2015). This complexity is the source of new insights on the behaviour of economic actors that extend not only IB theory but also theories developed with a single country in mind. However, despite its importance, this complexity needs to be controlled for to avoid confusion and ambiguity.

Controls are commonly used in large sample empirical studies to address spuriousness and hence enhance confidence in results. In these studies, the standard solution is to focus on a few focal influences and include controls for other characteristics that may have an additional impact on the dependent variable, but that are not the focus of interest of the particular study. However, in some cases these controls are included without due justification; often seemingly as a mechanical way of addressing potential reviewers' concerns rather than as a concerted effort to account for alternative influences that may pollute the proposed relationships. Yet the inclusion of controls does not by itself address the inherent complexity in IB research. In fact, the inclusion of the wrong controls, or exclusion of relevant controls, may seriously affect empirical results and cast in doubt the validity of a study.

In this editorial, we argue that including the appropriate controls is essential for the validity of a study and that researchers in general, and IB researchers in particular, need to pay more attention to the nature and role of controls when conducting their studies to increase the trustworthiness of the ideas and findings presented. Here we go beyond previous discussions of controls that have focused on their use in large sample studies (e.g., Becker 2005; Breaugh 2008; Spector and Bran-nick 2011; Moody and Marvell 2010) and propose that future research can improve by taking into account controls in three areas: theory, research design and empirical analysis. Table 6.2 summarizes the recommendations we discuss in this editorial. First, we explain how to use controls to theoretically establish the boundaries of arguments and dismiss alternative and com-

peting explanations of the proposed relationships. Second, we explain how to design studies to include a control group in the sample to facilitate the comparison to the group of interest in order to identify whether the arguments are general or apply only to certain groups. Third, we explain how to use appropriate statistical techniques which account for alternative influences on the dependent variable by including relevant control variables.

Table 6.2 Recommendations for using controls in quantitative research

Dimension	Focus of attention	Question to answer	Checklist
Theory	*Theoretical boundaries:* State the theoretical boundaries of the arguments	Under which conditions do the arguments hold?	Identify the complementary (or substituting) factors or characteristics at various levels (e.g., country, industry, firm, team or individual level) that you assume Discuss how the arguments proposed apply to certain types of individuals or companies
	Alternative explanations: Theoretically dismiss alternative explanations of the relationships discussed	How could these relationships be alternatively explained?	Identify alternative theories that may explain the proposed relationships Discuss how the mechanisms proposed by such alternative theories differ from the ones proposed by your arguments Argue and explain how the predictions driven by the theory are better than the predictions driven by the alternative theory

(*continued*)

Table 6.2 (continued)

Dimension	Focus of attention	Question to answer	Checklist
Research design	*Control group*: Establish whether arguments apply (a) to all individuals or companies in general or only to individuals or companies of a particular nature, and (b) to all individuals or companies in general but individual or companies of a particular nature exhibits some additional different behaviour	Do these relationships hold only for the individuals or companies analysed or do they hold for others as well?	Include in the research design a control group against which the relation of interest can be contrasted and compared Modify arguments and hypotheses so that they are presented in comparison to the control group and not just as general arguments
	Natural experiments: Take advantage of natural experiments to identify a control group, both in quantitative as well as qualitative analyses	Do the relationships hold for firms not affected by the natural experiment?	Use natural experiments when possible to compare firms subject to a treatment to those not subject to the treatment
Empirical analysis	*Inclusion of controls:* Include relevant and theoretically-justified control variables	What are the alternative influences on the dependent variable and how are they affecting it?	Avoid mimicry and provide justification for the inclusion of a variable as a control rather than as an independent variable Avoid methodological trickery and select controls that explain the dependent rather than independent variable(s)

(*continued*)

Table 6.2 (continued)

Dimension	Focus of attention	Question to answer	Checklist
	Exclusion of controls: Exclude controls that are independent of other controls and of the dependent variable and related dimensions of a construct	What are the main influences on the dependent variable?	Include all dimensions or explain why some dimensions should be excluded
	Measurement of controls: Provide full information on the control variables	How are the constructs measured? What are their characteristics?	Discuss which controls were included, how they were measured and where they come from Explain and justify if controls, their measurement or their treatment vary within the same study Report descriptive statistics for controls including means, standard deviations, range, and so on, and provide evidence of reliability and validity where appropriate
	Reporting of controls: Identify and report the relative impact of the controls on the dependent variable	How important are the alternative influences on the dependent variable?	Discuss how controls influence the dependent variable(s) and key relationships

(*continued*)

Table 6.2 (continued)

Dimension	Focus of attention	Question to answer	Checklist
	Compare the impact of the independent variables and all influences		Discuss the impact of controls in comparison to the impact of independent variables Compute economic significance of variables on the dependent variable Show model with controls only before adding explanatory variables Report significance levels and betas Show full model with and without controls

Trustworthiness Through Controls in Theoretical Development

Despite our quest for generalization, theoretical arguments rarely have universal applicability. Typically, a theory is developed with a particular, often rather narrow, set of assumptions regarding its boundaries and potential applicability. Though often not stated explicitly, such boundary conditions regarding the use of a theory may result in its' applicability being limited to particular contexts, for instance countries with democratic political systems and efficient market mechanisms, or individuals with a minimum level of education or income. When the context changes, as is often the case in IB research, the underlying theory or some of its arguments may need modification. Indeed, such modifications may constitute the very essence of the contribution that an IB study provides to the literature. Even in cases in which the contribution is the modification

of assumptions, the theoretical development may need two sets of controls: (1) theoretical boundaries that establish the limits of the applicability of the arguments and (2) clarifications that to account for the existence of alternative explanations of the arguments.

Establishing Theoretical Boundaries

Articles need a clear statement of the theoretical boundaries. Although the search for a generalizable argument is the objective that researchers aim to achieve, in reality most research has limited applicability, either because the researchers have not explained assumptions (Bello and Kostova 2012; Thomas et al. 2011), or because relationships depend on particular environmental conditions (Cuervo-Cazurra 2012). Thus a clear and explicit statement of the conditions under which the proposed relationships hold is needed as a first theoretical control.

A statement of theoretical boundaries is not the same as saying that the specified relationships only hold in the context in which they are later tested, but rather that the proposed arguments assume the existence of particular conditions. There is nothing wrong with having a study or arguments that assume certain conditions or specific contexts (Barkema et al. 2015). Such studies may provide important steps to our understanding of how a theory can be extended to explain situations that have not been considered in the initial development of the theory, but the conditions need to be made explicit.

To specify such boundary conditions we recommend the following. First, think about your unstated assumptions and the complementary (or substituting) factors or characteristics at various levels (e.g., individual, team, firm, country). Second, once you have identified these characteristics, discuss how the arguments proposed apply to certain types of individuals, companies or countries. You can do this with an initial paragraph before the theoretical arguments in which you acknowledge such boundary conditions with statements such as 'the theoretical boundaries of the arguments are the following. First, in the current paper we assume individuals or companies of [insert particular type, characteristics, etc.] and

thus the following arguments may need modification when analyzing individuals or companies of a different type or characteristics.'

In conceptual work you may include boundary conditions in the development of specific propositions, much in the same way as one would do in empirical work. For instance, one may specify that 'we expect Y to be positively influenced by X in emerging economies, whereas the relationship is reversed in advanced economies.' Note how explaining the boundary conditions of the theoretical arguments rather than simply stating 'other things being equal' provides a more precise application of the theory. At the same time, it provides future researchers with useful guidance for how to design an empirical study to test the relationship, and it even specifies variables that should be controlled for. Finally, in both empirical and conceptual work, you must return to the issue of the boundary conditions of theory when discussing results and implications as this provides the basis upon which the contributions should be judged.

Theoretically Controlling for Alternative Explanations

Once you have established the theoretical boundaries, a second level of theoretical controls involves theoretically accounting for alternative explanations of the proposed phenomenon. In many cases the propositions or hypotheses establish a relationship between independent variables and the dependent variable. However, these relationships can be explained with many alternative theories and theoretical arguments. Thus the burden falls on the researcher not only to explain the proposed relationship(s), but also to rule out alternative accounts for such relationship(s). To do this, you need first to identify alternative theories that may explain the proposed relationships, and then discuss how the mechanisms proposed by such alternative theories differ from the ones proposed by your preferred theory. After this, the next step is to argue and explain how the predictions driven by the theory proposed by you are better than the predictions driven by the alternative theory; especially

if this provides a simpler explanation with fewer assumptions (i.e., Occam's razor, Duignan 2015) and one that can be falsified with data (Popper 2002).

IB research may require alternative explanations because of differences in context or relationships across time and space (Dunning 1998). First, under the conditions established in the theoretical boundaries, the initial explanation may no longer hold and thus we need a more sophisticated explanation. It is your responsibility to provide proof that the new mechanisms are better than the old ones. You may explain how the previous arguments are theoretically constrained to particular situations, and a new explanation is needed for the new situation. Second, new influences and relationships may emerge, which previous theoretical explanations had not taken into account. In this case, you can explain how the previous mechanisms are too simplistic and extend theory to account for new conditions and assumptions.

Trustworthiness Through Controls in Research Design

Unlike the natural sciences, in management studies, with the exception of some psychology-based analyses, there are rarely random samples and limited opportunities for conducting experiments in which some firms are assigned to receive a treatment and others to be a control group (Banerjee and Duflo 2009; Cook and Campbell 1979). In order to encourage more experimental design in IB research, Zellmer-Bruhn, Caligiuri and Thomas (2016) outline these opportunities and explain the value and limitations of experiments in the IB context.

Despite these possibilities, however, most management studies use convenient samples that have data or surveys on companies or individuals that are easily accessible or effortlessly identified. If, in addition to not having a random sample, the researcher restricts the sample to firms or individuals that have a characteristic of interest, the researcher is in many cases bound to find the expected relationships. Without a control group the author cannot know whether this behaviour is exclusive to the group under analysis, or whether it is generalizable to other firms or individuals that were not included in the sample. For example, authors may argue

that emerging-market multinationals (EMNCs) are internationalizing quickly nowadays. If such arguments are tested on a sample that only includes EMNCs, researchers may find that this is indeed the case. However, if advanced economy multinationals are included as well, researchers may find that these firms are also internationalizing quickly thanks to, for example, advances in information and transportation technologies and the reduction of constraints on trade and investment. Hence the argument applies to all multinationals and not just EMNCs.

Including a Control Group

We recommend including a control group in the research design against which the relation of interest can be contrasted and compared. This helps understand whether: (1) the arguments presented apply to all individuals or companies in general or only to individuals or companies of a particular nature, and (2) the arguments presented apply to all individuals or companies in general but individuals or companies of a particular nature exhibits some additional different behaviours.

Including a control group also requires the modification of the arguments and hypotheses in the theoretical development, so that such arguments and hypotheses are presented in comparison to the control group and not just as general arguments. One interesting way of doing such comparison can be not only to include individuals or firms that do not have the required novel characteristic, but also, if data are available, to do a matched sample in order to identify how the characteristics of interest indeed drive the proposed relationships (Estrin et al. 2016; Reeb et al. 2012). Naturally, including a control group requires more work collecting data. This does not, however, excuse researchers from doing so. If the research question warrants the introduction of a control group in order to analyse a particular phenomenon, reviewers and editors must insist on such steps being taken. You will do well to consider this issue early on when designing your research in order to avoid rejection due to design issues. Inadequate attention to theoretical and empirical boundary conditions is often grounds for rejection in JIBS and excuses due to data limitations are not valid if data can be obtained.

Using Natural Experiments

In some instances researchers can take advantage of natural experiments to identify a control group. Natural experiments are examples of designs that are able to isolate (control) the effects of the focal (treatment) variable by eliminating the effects of extraneous factors. For example, Kogut and Zander (2000) analysed the ability to innovate of the German optics firm Carl Zeiss in different political environments in East and West Germany as a result of the division into two firms post-World War II. The rather dramatic division of Germany after the War provided a fertile ground for a natural experiment which utilized a matched-pair design of two entities that had hitherto been part of the same organization, thus avoiding some of the problems of conjectural causality (or multiple causes) inherent in comparative work (Ragin 2014).

Trustworthiness Through Controls in Empirical Analyses

A typical way of controlling for alternative explanations in large sample analyses is to include in the empirical model other variables that may influence the dependent variable but that are not the focus of discussion in the theoretical development. Unfortunately, some studies do not even include controls and merely use an analysis of differences in means between groups to test hypotheses; such analysis cannot be used to test theoretical arguments, because there may be many other alternative factors that influence behaviour beyond belonging to one group or another.

Even in cases when researchers include controls in empirical analyses, their inclusion often does not seem to be adequately justified or guided by theory. First, it appears that specific controls are sometimes included merely because previous papers have used them. In such cases one usually finds citations to previous work without an explanation of the reasons why such controls need to be included. On other occasions controls are included because they exert influence on some of the independent variables of interest. The inclusion of such controls raises two issues. One is the creation of multi-collinearity that results in the independent variables

of interest becoming statistically significant merely because some of the controls are included in the analysis. Another is a misunderstanding of the need to include controls; while controls need to be included as alternative explanations of the dependent variable, they should not serve as competitive explanations of other independent variable.

Third, there is a difference between *theoretically irrelevant* and *not statistically significant* controls. In the former case, if theory does not call for the inclusion of a variable in order to control for alternative influences on the dependent variable, it should not be part of the statistical model. If the variable happens to be statistically significant, it presents itself as an opportunity to develop theory (best case) or it represents a collinear relation with another independent variable (worst case). In the latter case, if theory calls for the inclusion of a particular control variable, it should be part of the statistical model irrespective of its statistical significance. In practice, selection of appropriate control variables may be difficult but should be guided by whether they satisfy the criteria for spuriousness based on theory, prior empirical studies, and common knowledge about the phenomenon under investigation. It is better to err on the side of caution by including all the theoretically relevant controls, even if many of them are not statistically significant, though such practice may result in unstable results due to overfitting of the model.

Based on our editorial experiences and to facilitate a better use of controls in large empirical analyses, we summarize the following observations on the common mistakes made in the use of controls in large sample empirical studies in IB, and provide some suggestions for solving them. We group them in four themes: Inclusion, exclusion, measurement and reporting.

Inclusion of Controls (1): Justified Controls

A common mistake is that there is often little or no theoretical justification for inclusion of specific controls, apart from inserting references to previous studies that have used the control. However, in many cases these references have little to do with the current dependent variable and may be contextually irrelevant.

Our recommendation it to include a theoretical justification. Avoid mimicry of other studies and instead provide sound theoretical reasoning for each and every control included. This should include a brief discussion of why a particular variable is a biasing (control) rather than a substantive (independent) variable in a particular model.

Inclusion of Controls (2): Relevant Controls

Impotent control variables are often included – for example ones that are uncorrelated with the dependent variable – without justification for inclusion. Unless a control can be legitimately justified as suppressor, it should be excluded as it will reduce power in the analysis. Alternatively, controls are sometimes included to improve the statistical significance of key relationships or to increase the model fit by reducing error terms. This includes instances where certain controls are included in some analyses but not in others, or the nature or even measurement of controls vary within the same study.

We recommend that you make sure to include the 'correct' controls. This should be driven by theory and not by previous research (which may be flawed or contextually different) or what works statistically. Also, avoid including too many controls in the pursuit of 'methodological trickery' – more is not necessarily better and each and every control must be theoretically and logically justified. Finally, select controls that explain the dependent variable, not those associated with independent variables.

Exclusion of Controls: Excluding Dimensions of Controls

Some studies conveniently exclude related dimensions of the independent construct, which may artificially inflate the significance of the dimension of the construct included. For example, when analysing the impact of culture on finance, many studies select individualism/collectivism as the key cultural variable and exclude the other dimensions of culture in the controls; if the other dimensions are included, the significance of individualism/collectivism may likely be affected, maybe to the point in which it loses its statistical significance.

Our recommendation is that instead of excluding certain dimensions of a construct, either include all dimensions or explain the logic for excluding the dimensions and potential biases. Also, if controls, their measurement or their treatment vary within the same study, this needs to be clearly explained and justified.

Measurement of Controls (1): Specify Controls

In some cases there is little information on the specific measurement of controls, using some vague indications rather than providing precise explanation of how the measure was created (e.g., discussing GDP per capita without specifying where data came from and how it was measured: using GDP in current dollar terms, in international dollar terms, in PPP terms, dividing GDP by the estimated or census population, etc.).

We suggest that you provide information and clearly discuss in the method section how controls were measured and why a particular measure is adequate for the context of the study. This may include a discussion of the validity and reliability of controls, as well as an explanation of choices of controls. Without specific knowledge about which controls were included, how they were measured and where they come from, replication is impossible.

Measurement of Controls (2): Describe Controls

Often some controls are not included in the correlation matrix. Their exclusion may be a sign of sloppy work or a sign of trying to conceal potential multicollinearity problems between controls and other independent variables. Also, in some cases the effect size is not provided for all controls in tables, which results in missing information.

We would recommend you report descriptive statistics for all controls including means, standard deviations, range, and so on, and provide evidence of reliability and validity where appropriate.

Reporting of Controls (1): Impact of Controls

There is often no discussion of the impact of controls in the results or discussion sections. Far too often we are left to speculate what significant controls may mean and almost never is the relationship between controls and the dependent variable explicitly discussed.

Our recommendation is to discuss in the results section how controls influence dependent variable(s) and key relationships in your model and offer insights for future researchers on what to control for in studies of a particular phenomenon.

Reporting of Controls (2): Importance of Controls

In many cases, controls account for more explanatory power than the main effects, but this is almost never discussed. This begs the question of whether the statistical significance of the variables of interest has any economic significance, which again is rarely computed and discussed.

We recommend that you explain the impact of the controls in comparison to the impact of the independent variables, and compute and discuss the economic significance of the variables on the dependent variable.

Reporting of Controls (3): Comparisons

Another common mistake is that the baseline model with only the impact of the controls on the dependent variable is excluded from the table of results, or the full model is not run both with and without controls. Thus we cannot fully assess how much explanatory power the inclusion of all the relevant independent variables provides beyond the controls.

Our recommendation is that you run a model that only includes controls before adding explanatory variables to models and report significance levels and betas. Also run full models with and without controls to rule out controls as potential explanation for results. Explain what it may mean if results differ markedly when controls are included and when they are not; this may help future researchers rule out potential biasing effects. Discuss the results in relation to the specific controls included using lan-

guage like 'controlling for A, B and C, the relationship between X and Y was…' and make sure to relate this to prior studies in the literature – this may include references to other studies of the same phenomenon in which certain controls were found to have similar, opposite, or no effects.

Towards More Trustworthy Quantitative Manuscripts

In Part B of this editorial we have provided suggestions on how to control for alternative influences in the complex phenomena analysed in IB research to increase the trustworthiness of the ideas and findings presented in the research. We argued that studies need to include controls at the level of theory, research design and analysis to account for alternative explanations and influences to understand this complexity, in addition to providing more sophisticated theoretical development and the explanation of the mechanisms (e.g., Bello and Kostova 2012; Thomas et al. 2011). The need to control for alternative explanations applies to (1) theoretical development, by explaining the boundaries of the analyses on the applicability of the theory; (2) research design, by including a control group against which to compare the characteristic of interest, and (3) the empirical analysis.

The overall intention of Part B is to make researchers aware of what actually controlling for alternative explanations entails, which goes beyond what has in many cases become an automatic or mechanistic process of adding a few variables to the statistical analysis. We do so by providing specific recommendations for selection and treatment of control variables in IB research. Hopefully these recommendations will result in better and more trustworthy quantitative studies in the future.

Looking Ahead: Mixed Methods and Ambidextrous IB Scholars

Both qualitative and quantitative research can improve their trustworthiness by paying attention to the theoretical development, research design and data analysis, to ensure that the insights gained from the analyses are

not subject to alternative, unaccounted influences. Although we have divided the discussion in this editorial into two parts to provide depth to the suggestions, both qualitative and quantitative data are complementary in developing IB as a field of scholarly inquiry. Mixed method approaches, in which researchers undertake both qualitative and quantitative studies to answer their research question, are worthy of greater scholarly attention than they have hitherto attracted. Using qualitative and quantitative methods in tandem can increase the trustworthiness of a study by compensating for the weaknesses inherent in any one method alone, and can yield a richer answer to a research question (Brannen and Peterson 2009; Kaplan 2015; Small 2011). This points to the benefit of scholars investing in becoming more ambidextrous with respect to their methods-related skills and/or in establishing ambidextrous research teams; and we hope that there are increasing numbers of these. However we need to heed the caution that a mixed methods scholar "risks being jack-of-all-trades and master of none" (Kaplan 2015: 431). Trustworthiness through cohesiveness, depth and rigor still needs to be incorporated into the design and analysis of all datasets in order for a mixed method study to be substantive and persuasive. We hope that this editorial provides a useful framework for sparking the interest in gaining expertise in a different tradition and creating more trustworthy studies that provide deeper insights.

Acknowledgements We are thankful for suggestions of improvements from Peter Buckley, John Cantwell, Jose Cortina and anonymous reviewers. Alvaro Cuervo-Cazurra thanks the Robert Morrison Fellowship and the Walsh Research Professorship for financial support.

References

Andersson, U., A. Cuervo-Cazurra, and B.B. Nielsen. 2014. Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies* 45 (9): 1063–1071.

Banerjee, A., and E. Duflo. 2009. The experimental approach to development economics. *Annual Review of Economics* 1 (1): 151–178.

Bansal, P., and K. Corley. 2012. Publishing in AM part 7: What's different about qualitative research? *Academy of Management Journal* 55 (3): 509–513.

Barkema, H.G., X.-P. Chen, G. George, Y. Luo, and A.S. Tsui. 2015. West meets East: New concepts and theories. *Academy of Management Journal* 58 (2): 460–479.

Becker, T.E. 2005. Potential problems in the statistical control of variables in organizational research: A qualitative analysis with recommendations. *Organizational Research Methods* 8 (3): 274–289.

Bello, D.C., and T. Kostova. 2012. Conducting high impact international business research: The role of theory. *Journal of International Business Studies* 43 (6): 537–543.

Bettis, R.A., S. Ethiraj, A. Gambardella, C. Helfat, and W. Mitchell. 2016. Creating repeatable cumulative knowledge in strategic management: A call for a broad and deep conversation among authors, referees, and editors. *Strategic Management Journal* 37 (2): 257–261.

Bingham, C.B. 2009. Oscillating improvisation: How entrepreneurial firms create success in foreign market entries overtime. *Strategic Entrepreneurship Journal* 3 (4): 321–345.

Birkinshaw, J., M.Y. Brannen, and R.T. Tung. 2011. From a distance and generalizable to up close and grounded: Reclaiming a place for qualitative methods in international business research. *Journal of International Business Studies* 42 (5): 573–581.

Boussebaa, M., S. Sinha, and Y. Gabriel. 2014. Englishization in offshore call centers: A postcolonial perspective. *Journal of International Business Studies* 45 (9): 1152–1169.

Brannen, M.Y. 2004. When Mickey loses face: Recontextualization, semantic fit, and the semiotics of foreignness. *Academy of Management Review* 29 (4): 593–616.

Brannen, M.Y., and M.F. Peterson. 2009. Merging without alienating: Interventions promoting cross-cultural organizational integration and their limitations. *Journal of International Business Studies* 40 (3): 468–489.

Breaugh, A. 2008. Important considerations in using statistical procedures to control for nuisance variables in non-experimental studies. *Human Resource Management Review* 18 (4): 282–293.

Burgelman, R.A. 2011. Bridging history and reductionism: A key role for longitudinal qualitative research. *Journal of International Business Studies* 42 (5): 591–601.

Cantwell, J., and M.Y. Brannen. 2011. Positioning JIBS as an interdisciplinary journal. *Journal of International Business Studies* 42 (1): 1–9.

Cantwell, J., A. Piepenbrink, and P. Shukla. 2014. Assessing the impact of IBS as an interdisciplinary journal: A network approach. *Journal of International Business Studies* 45 (7): 787–799.

Caprar, D.V. 2011. Foreign locals: A cautionary tale on the culture of MNC local employees. *Journal of International Business Studies* 42 (5): 608–628.

Chang, S., A. Van Witteloostuijn, and L. Eden. 2010. Common method variance in international business research. *Journal of International Business Studies* 41 (2): 178–184.

Cheng, J.L.C., W. Henisz, K. Roth, and A. Swaminathan. 2009. Advancing interdisciplinary research in the field of international business: Prospects, issues and challenges. *Journal of International Business Studies* 40 (7): 1070–1074.

Cheng, L.C., J. Birkinshaw, D.R. Lessard, and D.C. Thomas. 2014. Advancing interdisciplinary research: Insights from the IBS special issue. *Journal of International Business Studies* 45 (6): 643–648.

Chidlow, A., E. Plakoyiannaki, and C. Welch. 2014. Translation in cross-language international business research: Beyond equivalence. *Journal of International Business Studies* 45 (5): 562–582.

Cook, T.C., and D.T. Campbell. 1979. *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Corbin, J., and A. Strauss. 2008. *Basics of qualitative research*. 3rd ed. Thousand Oaks: Sage.

Cortina, J.M., T. Kohler, and B.B. Nielsen. 2015. Restriction of variance interaction effects and their importance for international business research. *Journal of International Business Studies* 46 (8): 879–885.

Cuervo-Cazurra, A. 2012. How the analysis of developing country multinational companies helps advance theory: Solving the Goldilocks debate. *Global Strategy Journal* 2 (3): 153–167.

Doz, Y. 2011. Qualitative research for international business. *Journal of International Business Studies* 42 (5): 582–590.

Duignan, B. 2015. Occam's razor. *Encyclopedia Britannica.* http://www.britannica.com/topic/Occams-razor. Accessed 27 July 2015.

Dunning, J.H. 1998. Location and the multinational enterprise: A neglected factor? *Journal of International Business Studies* 29 (1): 45–66.

Edmondson, A.C., and S.E. McManus. 2007. Methodological fit in management field research. *Academy of Management Review* 32 (4): 1155–1179.

Eisenhardt, K., and M. Graebner. 2007. Theory building from cases: Opportunities and challenges. *Academy of Management Journal* 50 (1): 25–32.

Estrin, S., K.E. Meyer, B.B. Nielsen, and S. Nielsen. 2016. Home country institutions and the internationalization of state owned enterprises: A cross-country analysis. *Journal of World Business* 51 (2): 294–307.

Gioia, D.A., K.G. Corley, and A.L. Hamilton. 2012. Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods* 16 (1): 15–31.

Glaser, B.G., and A.L. Strauss. 2011. *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Publishing Company.

Gligor, D.M., C.L. Esmark, and I. Golgeci. 2016. Building international business theory: A grounded theory approach. *Journal of International Business Studies* 47 (1): 93–111.

Haley, U.C.V., and D.M. Boje. 2014. Storytelling the internationalization of the multinational enterprise. *Journal of International Business Studies* 45 (9): 1115–1132.

Johns, G. 2006. The essential impact of context on organizational behavior. *Academy of Management Review* 31 (2): 386–408.

Jonsson, A., and N.J. Foss. 2011. International expansion through flexible replication: Learning from the internationalization experience of IKEA. *Journal of International Business Studies* 42 (9): 1079–1102.

Kaplan, S. 2015. Mixing quantitative and qualitative research. In *Handbook of qualitative organizational research: Innovative pathways and methods, chapter 40*, ed. K. Elsbach and R.M. Kramer. New York: Routledge.

Kogut, B., and U. Zander. 2000. Did socialism fail to innovate? A natural experiment of the two Zeiss companies. *American Sociological Review* 65 (2): 169–190.

Lamb, P., J. Sandberg, and P.W. Liesch. 2011. Small firm internationalization unveiled through phenomenography. *Journal of International Business Studies* 42 (5): 672–693.

Langley, A. 1999. Strategies for theorizing from process data. *Academy of Management Review* 24 (4): 691–710.

Langley, A., and C. Abdallah. 2011. Templates and turns in qualitative studies of strategy and management. *Building Methodological Bridges: Research Methodology in Strategy and Management* 6: 201–235.

Langley, A., C. Smallman, H. Tsoukas, and A.H. Van de Ven. 2013. Process studies of change in organization and management: Unveiling temporality, activity and flow. *Academy of Management Journal* 56 (1): 1–12.

Lapan, S.D., and K. de Marrais, eds. 2003. *Foundations for research: Methods of inquiry in education and the social sciences*. New York: Routledge.

Marschan-Piekkari, R., and C. Welch, eds. 2011. *Rethinking the case study in international business and management research*. Cheltenham: Edward Elgar Publishing.

Merriam-Webster Dictionary. 2016. Trustworthy. http://www.merriam-webster.com/dictionary/trustworthy. Accessed 18 Apr 2016.

Miles, M.B., and M.A. Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. 2nd ed. Thousand Oaks: Sage.

Moody, C.E., and T.B. Marvell. 2010. On the choice of control variables in the crime equation. *Oxford Bulletin of Economics and Statistics* 72 (5): 696–715.

Peterson, M.F., J.-L. Arregle, and X. Martin. 2012. Multilevel models in international business research. *Journal of International Business Studies* 43 (5): 451–457.

Piekkari, R., C. Welch, and E. Paavilainen. 2009. The case study as disciplinary convention: Evidence from international business journals. *Organizational Research Methods* 12 (3): 567–589.

Pollock, T.G., and J.E. Bono. 2013. Being Scheherazade: The importance of storytelling in academic writing. *Academy of Management Journal* 56 (3): 629–634.

Popper, K. 2002. *The logic of scientific discovery*. 2nd ed. New York: Routledge.

Pratt, M.G. 2008. Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier North American journals. *Organizational Research Methods* 11 (3): 481–500.

———. 2009. For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. *Academy of Management Journal* 52 (5): 856–862.

Ragin, C. 2014. *The comparative method: Moving beyond qualitative and quantitative strategies*. Oakland: University of California Press.

Ragins, B.R. 2012. Reflections on the craft of clear writing. *Academy of Management Review* 37 (4): 493–501.

Reeb, D., M. Sakakibara, and I.P. Mahmood. 2012. Endogeneity in international business research. *Journal of International Business Studies* 43 (3): 211–218.

Siggelkow, N. 2007. Persuasion with case studies. *Academy of Management Journal* 50 (1): 20–24.

Silberzahn, R., and E.L. Uhlmann. 2015. Crowdsourced research: Many hands make tight work. *Nature* 526 (7572): 189–191.

Small, M.L. 2011. How to conduct a mixed methods study: Recent trends in a rapidly growing literature. *Annual Review of Sociology* 37 (1): 57–86.

Spector, P.E., and M.T. Brannick. 2011. Methodological urban legends: The misuse of statistical control variables. *Organizational Research Methods* 14 (2): 287–305.

Thomas, D.C., A. Cuervo-Cazurra, and M.Y. Brannen. 2011. Explaining theoretical relationships in international business research: It's about the arrows, NOT the boxes. *Journal of International Business Studies* 42 (9): 1073–1078.

Van Maanen, J. 1998. *Qualitative studies of organizations*. Thousand Oaks: Sage.

Welch, C., and E. Paavilainen-Mäntymäki. 2014. Putting process (back) in: Research on the internationalization process of the firm. *International Journal of Management Reviews* 16 (1): 2–23.

Welch, C., R. Piekkari, E. Plakoyiannaki, and E. Paavilainen-Mäntymäki. 2011. Theorizing from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies* 42 (5): 740–762.

Westney, D.E., and J. Van Maanen. 2011. The casual ethnography of the executive suite. *Journal of International Business Studies* 42 (5): 602–607.

Yin, R. 2009. *Case study research: Design and methods*. 4th ed. London: Sage.

Zellmer-Bruhn, M., P. Caligiuri, and D.C. Thomas. 2016. From the editors: Experimental design in international business research. *Journal of International Business Studies* 47 (4): 399–407.

7

Can I Trust Your Findings? A Commentary

Jonathan P. Doh

Introduction

Academic journals in business and management have taken an increasing interest in providing advice, commentary, criticism, and guidance regarding the use and application of various empirical methods. In this regard, the *Journal of International Business Studies* (*JIBS*) has published several editorials on various aspects of methods and empirical interpretation. In one of these, "From the Editors: Can I trust your findings? Ruling out alternative explanations in international business research," Cuervo-Cazurra, Andersson, Brannen, Nielsen, and Reuber offer guidance to scholars as to how to improve the validity and generalizability of empirical research in international business (IB), especially with respect to

Cuervo-Cazurra, A., U. Andersson, M.Y. Brannen, B.B. Nielsen, and A.R. Reuber. 2016. Can I trust your findings? Ruling out alternative explanations in international business research. *Journal of International Business Studies* 47 (8): 881–897.

J. P. Doh (✉)
Villanova University, Villanova, PA, USA
e-mail: jonathan.doh@villanova.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_7

ruling out alternative explanations. In this editorial, the authors, each of whom has extensive editorial experience in their own right, tackle a range of strategies to ensure greater "trustworthiness" in qualitative and quantitative studies in IB. In my commentary, I will briefly summarize and comment upon the two major sections of this editorial, provide some of my own perspectives on insights we might derive from these contributions, and highlight some broader implications. I will then offer some concluding thoughts.

Ruling Out Alternative Explanations in Qualitative Studies

In their discussion of qualitative studies, the authors emphasize the somewhat idiosyncratic path that often leads to the initiation of qualitative research, namely, the fascination and curiosity around a specific phenomenological context. As such, they stress the importance of justifying the research context, the overall research design, and perhaps most importantly, the choice to use qualitative (versus other) methods. They review a series of questions scholars should ask related to the context, research design, and methodology employed. Much of this guidance is fairly straightforward, however, one "tip" stands out: "Identify and explain 'negative' cases in the data: cases that do not confirm to dominant patterns" (p. 885). It strikes me that too often scholars are fearful of revealing any findings or cases that are contrary to their overall hypotheses (whether formal or informal) and may somewhat subconsciously or unknowingly suppress this countervailing information. These authors wisely discourage that tendency, suggesting that null results can often provide significant insight, especially when the study is replicating some earlier, often widely accepted findings.

They also acknowledge that qualitative research papers often have a particular challenge in getting through the review process. One response to this problem is provided by this article, namely, to ensure that scholars have paid sufficient heed to demonstrating the trustworthiness of their research. Another is to explore ways to act on the often expressed concern

by qualitative researchers in IB (and more generally) that editors routinely assign reviewers unfamiliar with those methods. In this regard, a recent *Journal of World Business* Perspectives article by Welch and Piekkari (2017) offers a useful complement to the subject article by providing a similar set of recommendations and guidance to reviewers and editors of qualitative research in IB.

Ruling Out Alternative Explanations in Quantitative Studies

The discussion of quantitative studies focuses primarily on the use and misuse of control variables. Here the authors run through a series of "to dos" and "not to dos" related to control variables. The series follows the same pattern as in the qualitative section, focusing on questions that should be addressed in theory, in empirical analysis, and in reporting of results. To me, the most important contribution here has to do with the anticipation of boundary conditions, and therefore pre-identification of potential controls during the theoretical development of an article rather than as a *pro-forma* afterthought. The authors provide sound recommendations related to both unwarranted inclusion as well as unwarranted exclusion of controls, urging scholars to identify controls on theoretical grounds, not just because they have been used by others under similar circumstances.

In a more extended version of this discussion, Nielsen and Raswant (2018) explore the selection, use, and reporting of control variables in IB. They find that a majority of the studies they review do not meet basic minimum standards in terms of the use of controls. While the use and misuse of controls is a critical concern in empirical studies in IB, it is more a symptom than a problem. Rather, the core challenge in IB and other business and management research concerns some of the generally accepted norms, practices, and assumptions that undergird what we consider to be an acceptable empirical exposition.

Discussion: Broadening the Methodological Conversation

In this regard, several developments and innovations have occurred alongside or since this editorial was published, each of which serve to strengthen its recommendations and overall contribution and to position it within a broader debate.

First, the fields of IB and management more broadly have lately been reconsidering several structures associated with our research tradition. Specifically, there is renewed interest in challenging how we report our findings, determining whether and how studies can be replicated, and in the broader aims and purpose of our research. In an influential *Strategic Management Journal* editorial, Bettis, Ethiraj, Gambardella, Helfat, and Mitchell (2016: 257) provide a rethinking of the "*knowledge and norms around the use and interpretation of statistics.*" In this editorial, the authors (also editors) criticize the overreliance on "p values" in strategic management research in relation to the overall "interestingness" of the research. Further, they attempt to shift the focus away from statistical significance based on specific confidence intervals to the overall size of the effect, the potential for it to be replicated, and the relative uninterest on the part of journals to publish studies that contain null or insignificant findings.

This critique launched a wave of responses from other journals, including *Journal of World Business* and *JIBS*, each of which expressed policies that encourage the specification of effect size and specific p values as opposed to cut-offs (See Hahn and Ang 2017; Meyer 2017). More broadly, IB, like other fields, has been wrestling with questions about its relevance and connection to real-world phenomena (See Doh 2015; Delios 2017).

The second, related—but more operational—development has to do with the increasing expectation (at some journals now a requirement) that authors make their datasets available to reviewers and readers as part of the submission or acceptance process. This practice goes a long way to allowing for the examination of the assumptions and overall approach to the research design, as well as easing opportunities for replication and extension. In so doing, it serves two purposes: it provides a validity

"check" on the empirical conclusions of our research and also offers guidance to others who may want to explore similar themes using the same basic data source.

Taking these points together, a growing movement is underway to reorient and redirect IB scholarship towards (1) research that is replicable and can contribute to a broader body of knowledge and understanding, and (2) research that is more relevant and tethered to real-world problems and challenges. This, in my view, is a promising direction for IB, a trend that I, along with co-authors, have discussed extensively elsewhere (Buckley et al. 2017). In particular, I and my co-authors have suggested that societal "grand challenges" offer an organizing principle for IB research that seeks to achieve relevance, rigour, and real-world contribution. At the same time, however, we acknowledge that researching grand challenges is itself challenging in that it may require interdisciplinary approaches, multilevel methods, and consideration of a diverse range of societal actors and influences.

Summary and Conclusion

In sum, the "Can I Trust Your Findings?" editorial makes a solid contribution to the various editorials and commentaries that have offered guidance, direction, and best practices in empirical methods in IB, especially as they relate to founding assumptions and approaches to context in qualitative methods, and the use (and misuse) of control variables in quantitative ones. Further, this contribution adds to the broader debate about the value of replication, the relevance of null findings, and the need for IB to tackle bigger, bolder challenges that often span levels of analysis and lend themselves to multi-method approaches.

References

Bettis, R., S. Ethiraj, A. Gambardella, C. Helfat, and W. Mitchell. 2016. Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal* 37 (2): 257–261.

Buckley, P.O., J.P. Doh, and M. Benischke. 2017. Towards a renaissance in international business research? Big questions, grand challenges, and the future of IB scholarship. *Journal of International Business Studies* 48 (9): 1045–1064.

Delios, A. 2017. The death and rebirth (?) of international business research. *Journal of Management Studies* 54 (3): 391–397.

Doh, J.P. 2015. From the editor: Why we need phenomenon-based research in international business. *Journal of World Business* 50 (4): 609–611.

Hahn, E.D., and S.H. Ang. 2017. From the editors: New directions in the reporting of statistical results in the Journal of World Business. *Journal of World Business* 52 (2): 125–126.

Meyer, K. 2017. What's in a *p*? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies* 48 (5): 535–551.

Nielsen, B., and A. Raswant. 2018. The selection, use, and reporting of control variables in international business research: A review and recommendations. *Journal of World Business* 53 (6): 958–968.

Welch, C., and R. Piekkari. 2017. How should we (not) judge the 'quality' of qualitative research? A re-assessment of current evaluative criteria in International Business. *Journal of World Business* 52 (5): 714–725.

8

Additional Thoughts on Trusting Findings: Suggestions for Reviewers

Mary Yoko Brannen, Alvaro Cuervo-Cazurra, and A. Rebecca Reuber

After reflecting on our *JIBS* editorial (Cuervo-Cazurra et al. 2016) on 'trusting findings' from research based on qualitative and quantitative methods, thinking about not only what we said but also how we went about writing it, Jonathan Doh's (Chap. 7, this volume) thoughtful commentary, and recent discussions on this topic, two observations stand out that warrant further thought.

The first observation is how difficult it is for scholars to take up the challenge of tackling significant, bold, real-world phenomena with an

M. Y. Brannen
Copenhagen Business School, Frederiksberg, Denmark

University of Victoria, Victoria, BC, Canada
e-mail: myb.egb@cbs.dk

A. Cuervo-Cazurra (✉)
Northeastern University, Boston, MA, USA
e-mail: a.cuervocazurra@neu.edu

A. R. Reuber
University of Toronto, Toronto, ON, Canada
e-mail: reuber@rotman.utoronto.ca

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_8

open-minded, interdisciplinary, multi-methods approach. Even the process of agreeing upon the format and content of our editorial was not easy from a mixed-methods perspective. The second observation is the challenge of reviewing mixed-methods articles. Few researchers are trained in both methods, and even fewer can properly establish bridges between the two. In hopes of helping international business (IB) scholars to become more ambidextrous, especially when it comes to reviewing mixed-methods articles, we offer the following three suggestions.

First, become familiar with the norms and standards of the methods used in the manuscript in order to provide insightful suggestions for improvement. There are a wide variety of qualitative and quantitative methods. Both types of methods offer different benefits to a study, as summarized in Table 8.1. Reviewers can assess the manuscript on how well the authors leverage these benefits to contribute to knowledge. In addition to assessing the use of each method individually, pay attention to what each method adds to the storyline of the manuscript's theoretical contribution, and how the methods complement each other.

Second, pay attention to how data are used to build evidence. With large quantities of diverse data and few standards to prescribe how mixed-methods data should be depicted, authors can be unsure of how to present the data to build evidence. Two unfortunate outcomes can result. One outcome is data overwhelming the storyline. Authors are often reluctant to pare down their data, in order to showcase data that may have been difficult and costly to collect and analyse, and to provide as much evidence for their claims as they can. A second unfortunate outcome is the opposite. Authors hold back on presenting data to highlight

Table 8.1 Characteristics of qualitative and quantitative methods

Qualitative research	Quantitative research
Useful for capturing meaning	Useful for testing relationship
Focus on processes and events	Focus on relationships between variables
Validity and authenticity are key	Reliability and replicability are key
Contextualization is critical	Generalization beyond context
Few subjects	Large number of subjects
Intentional sampling	Random sampling
Thematic analysis	Statistical analysis
Researcher immersion	Researcher detachment

the theoretical discussion, and the paper ends up with little persuasive evidence for the theoretical claims. The reviewer's role then becomes not only a judge of the quality of the data through how it was collected and analysed, but also a judge of the contribution of the data to the theoretical story being told.

Third, allow time to identify the theoretical storyline if the data are interesting and have the potential to yield an insightful contribution. If the authors have a large volume of diverse data, there are likely to be multiple stories in the data set, and it may take several iterations to narrow it down. The suggestions of the reviewers are especially valuable at this point. However, the suggestions may not be fully followed in the paper's revision. Reviewers have the best of intentions and provide developmental feedback based on their expertise but have incomplete knowledge of the authors' data. Bearing in mind that the authors receive input from other reviewers and an editor, when the revised paper comes back, reviewers should be open to the possibility that the storyline might have taken a different route from what they had suggested. The review process offers a rich and multi-vocal opportunity to establish a conversation with the authors on how to make the most of their data. Mixed-methods papers are particularly demanding to execute and to review. Our hope is that our editorial and these additional thoughts can help to increase methodological ambidexterity among IB scholars.

Reference

Cuervo-Cazurra, A., U. Andersson, M.Y. Brannen, B.B. Nielsen, and A.R. Reuber. 2016. Can I trust your findings? Ruling out alternative explanations in international business research. *Journal of International Business Studies* 47 (8): 881–897.

Part V

Theorizing from Cases

9

Theorising from Case Studies: Towards a Pluralist Future for International Business Research

Catherine Welch, Rebecca Piekkari, Emmanuella Plakoyiannaki, and Eriikka Paavilainen-Mäntymäki

Introduction

The case study has an established place in qualitative international business (IB) research.[1] A recent review of articles published in four core IB journals over a 10-year period found case studies to be the most popular

C. Welch
The University of Sydney Business School, The University of Sydney, Sydney, NSW, Australia
e-mail: catherine.welch@sydney.edu.au

R. Piekkari (✉)
School of Business, Aalto University, Espoo, Finland
e-mail: rebecca.piekkari@aalto.fi

E. Plakoyiannaki
Leeds University Business School, University of Leeds, Leeds, UK
e-mail: E.Plakoyiannaki@leeds.ac.uk

E. Paavilainen-Mäntymäki
Turku School of Economics, University of Turku, Turku, Finland
e-mail: eriikka.paavilainen-mantymaki@utu.fi

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_9

qualitative research strategy (Piekkari et al. 2009). This prevalence of the case study is not surprising, given its potential to generate novel and groundbreaking theoretical insights. Yet our contention in this paper is that the theorising potential of case studies has not been fully realised in the field of IB. We attribute this to the entrenched belief that case research is suited only to inductive theory-building. In this paper, we seek to challenge this belief by offering alternatives to inductive theorising and broadening the possibilities in IB for theorising from case studies. In order to do so, we consider how the case study generates causal explanations and how it incorporates context – two features of the case study that are often regarded as being incompatible. By challenging common preconceptions about case studies, we see our audience as comprising not just qualitative researchers, but also the wider community of scholars who are often called upon to evaluate the theoretical contribution of case research.

Our paper contributes to the debate over theorising in IB research, which recognises that alongside the perennial epistemological dilemma faced by social scientists – namely, how to develop robust explanations about phenomena in the social world – IB scholars have to contend with the question of how to ensure that their theories are sensitive to diverse national contexts. Consequently, we would argue that IB is a highly appropriate field in which to discuss the development of rigorous, yet context-sensitive, theory. There is growing concern that, in the pursuit of robust explanations, contextualisation has suffered. Greater use of qualitative research has been suggested as a remedy for this imbalance, thus placing approaches such as the case study squarely on the agenda for IB theory. Yet our contention in this paper is that, in IB research, the dominant view of the case study as a tool solely for *inductive theory-building* has restricted its theorising potential, both in terms of generating causal explanations and of contextualising theory.

In this paper, we challenge this dominant view by constructing a typology that offers alternatives to inductive theory-building. The first alternative views the case study as a *natural experiment* for confirming or modifying existing theory. This method attributes greater explanatory power to the case study than does inductive theory-building. The second alternative, by conceiving case research as a form of *interpretive*

sensemaking, affirms the value of contextualisation to theorising. However, these two alternatives are both potentially limiting, we argue, because they set up a trade-off between the strengths of internal validity on the one hand and thick description on the other. In this paper, we outline a third alternative – a recent development in the methodological literature and new to IB – that rejects this trade-off, and instead emphasises the ability of the case study to generate *contextualised explanation*. By comparing these four approaches to theorising, we show that the differences between each type are fundamentally epistemological and philosophical (i.e., paradigmatic) in nature. Broadening the possibilities for theorising from the case study therefore requires an appreciation of how these underlying paradigmatic assumptions both enable and constrain empirical research.

We develop our arguments in this paper as follows. First, we review growing concerns about the decontextualised nature of theorising in IB research, arguing that inductive theory-building reinforces rather than resolves this dilemma. We then detail our analytical approach: a qualitative content analysis of case studies published in *Journal of International Business Studies (JIBS), Academy of Management Journal (AMJ)* and *Journal of Management Studies (JMS)* that allowed us to examine how researchers have conceptualised and utilised the contextualising and explanatory potential of the case study. This qualitative approach to textual analysis, while rarely used in IB research to date (Duriau et al. 2007), offers the strengths of iterative conceptual development, contextualised insights and access to the linguistic features of texts. We then present the typology generated from our content analysis, discussing first its foundations in theory (i.e., the methodological literature) and then its grounding in data (i.e., published case studies). We conclude by specifying how future IB researchers might enhance the theorising potential of the case study by combining contextual richness and explanatory rigour.

In this paper we will be using terms – notably context, theory, explanation and causation – that have been greatly contested in the social sciences, yet whose meaning is too often taken for granted by researchers. These terms will be developed during the course of our discussion, but we will offer our own definitions upfront. By context, we are referring to the

contingent conditions that, in combination with a causal mechanism, produce an outcome. Explaining a phenomenon we take to mean showing what makes it what it is. Explanation need not necessarily be causal (Ruben 1990: 233), but causal explanations are our focus in this paper given their centrality to the debate over the theorising potential of case studies. An explanation is causal if it makes claims about the capacities of objects and beings to make a difference to their world (adapted from Kakkuri-Knuutila et al. 2008; Sayer 1992). We take theory to mean a form of explanation that offers a coherent, examined conceptualisation of a phenomenon (based on Sayer 2000). Our subsequent discussion in this paper will reveal that these definitions are heavily influenced by critical realism, and that they have profound implications for our understanding of how to theorise from case research.

In the IB field, however, these foundational elements of the scientific endeavour receive little scrutiny (for an exception, see Redding 2005). We contend that a reassessment of these fundamental concepts is needed for the theorising potential of case studies to be realised. As Sayer (1992) has persuasively argued, methodology should not just be regarded as a matter of choosing among different methods of data production[2] and analysis; rather, it is about choosing among competing methods of theorising. Yet we contend that much of the methodological literature on case studies in IB, as well as in the social sciences generally, has focused on methods of data production and analysis rather than methods of theorising. In this paper, we show that alternative methods of theorising from case studies are available, and argue for their application to IB. Ultimately, by contrasting different theorising methods for case researchers, we are advocating paradigmatic pluralism (for a similar argument, see Brannen and Doz 2010; Morgan 1980; Van Maanen 1995). We have been influenced by contemporary philosophers of science who argue that pluralism is a necessary precondition for scientific discovery and theoretical advances (Kellert et al. 2006); accordingly, we argue that the IB field would benefit from diversity in approaches to theorising. To establish the need for greater diversity, we turn to the growing concern that context has been neglected in IB and management research.

Context and Theorising in IB Research

We would argue that the ground we cover in this paper – the tension between scientific explanation and context – is a concern for any research, but that it is particularly visible and pressing in IB, given the field's cross-border nature. How to account for context has been a recurring, but unresolved, question for IB scholars (Brannen and Doz 2010; Redding 2005). For example, researchers on China, facing an institutional environment very different from the Western origins of most management theories, have been conducting a lively debate on the need for contextualised theories and research processes (e.g., Child 2000; Shapiro et al. 2007; Tsui 2006). The internationalisation of the general management community has also sparked interest in contextualisation (see, e.g., Rousseau and Fried 2001; Tsui 2007). In this literature, context is typically defined as "the surroundings associated with phenomena which help to illuminate that phenomena [sic]" (Cappelli and Sherer 1991: 56).

Advocates for "contextualized knowledge" (Tsui 2004) concede that they face barriers to change in the scholarly community. Above all, they point to the entrenched belief that "context-free", universalist knowledge is superior to that of "context-valid", localised knowledge (Blair and Hunt 1986; Tsui 2004; Whetten 2009). Bamberger (2008: 844) observes that dominant beliefs "may be forcing us to overweight generalizability and, in the process, underweight contextual sensitivity". Suggestions of how to encourage contextualised research range from the modest – for example, adding context effects as moderating variables (Whetten 2009) – to the more radical, such as Tsui's (2006) call to explore non-Western methodological tools. There is an emerging consensus that "context-oriented" qualitative research forms part of the answer (Bamberger 2008; Johns 2006; Rousseau and Fried 2001; Tsui 2004).

However, proposals for modest change do not confront the underlying reasons for the dominance of decontextualised research: namely, the positivist assumptions that are still taken for granted in the IB field (Brannen and Doz 2010; Jack et al. 2008; Redding 2005). Any discus-

sion of contextualisation is necessarily embedded in a complex web of beliefs concerning the nature of theorising in the social sciences: the question of how to contextualise is ultimately about how to theorise, and the answer depends on one's philosophical orientation. The arguments in favour of contextualising IB research are therefore well established, but the solutions are more contested, and ultimately highly value laden.

By seeking to investigate how contextualising and explaining can be brought together in the case study, our paper addresses a gap in the existing literature on contextualising IB research. To date, this literature, while placing the need for more qualitative research on the agenda, has not examined its theorising potential in any depth. Qualitative research is referred to in very general terms, with no differentiation among its many traditions, and its contribution is simply assumed to lie in rich description and exploratory, inductive theorising. However, in relation to the case study, there are two problems with these assumptions. The first is that because the generalisability of case study findings is low, its theorising potential is ultimately regarded as inferior to that of hypothesis-testing research. Case studies are therefore confined to the initial, exploratory phase of research, and their potential for generating causal explanations is overlooked. Second, while all qualitative research is commonly assumed to be context sensitive, a strong trend towards decontextualisation has in fact prevailed in much case research. The rich context that is the essence of a case study is ultimately regarded as a hindrance to theorising. Since to theorise is to generalise away from context, "explaining" and "contextualising" are regarded as being fundamentally opposed. In this paper, we challenge these perceptions about theorising from case studies, and suggest how context and explanation might be reconciled. In this way, we are contributing not just an enhanced understanding of the theorising potential of the case study, but also a means of enriching the context orientation of IB research. In the next section, we detail how we ourselves used a qualitative approach to develop our conceptual understanding of the theorising potential of case studies in IB.

Qualitative Content Analysis of Published Case Study Articles

We commenced this study with a broad research question, namely "How do IB case researchers theorise from case studies?" In this section, we detail how and why we took the approach of a qualitative content analysis, and how we selected our dataset. Our approach to qualitative content analysis would best be termed "directed": that is, the analysis commences with an initial coding scheme (Hsieh and Shannon 2005). In contrast to a quantitative content analysis, even though the codes are selected in advance, they do not remain fixed during the analysis, but rather are refined through successive iterations between theory and data (Berelson 1971; Ryan and Bernard 2000). One strength of this approach is therefore that it allows for fresh conceptual understanding that is also grounded in empirical data.

A key task in any form of textual analysis is to decide on the appropriate sample; in other words, which texts to analyse (Krippendorff 2004). Our journal selection followed the qualitative principle of purposeful sampling, which allows the content analyst to select the units of investigation relevant to the study (Krippendorff 2004). We initiated our analysis with *JIBS;* as the leading journal in IB, we can expect that the case studies it publishes, although few in number (see Appendix), will influence research standards in the field (Clark and Wright 2007). We examined the period 1999–2008 in order to capture the most recent case study practices. We found little diversity in the methods of theorising from case studies in *JIBS.* At this point, we added a research question, namely: "What are the alternatives to theorising from case studies, and what is their potential contribution to IB research?"

Consequently, we expanded our dataset to include two management journals – *AMJ* and *JMS* – that are comparable with, yet contrasting to, *JIBS.* Like *JIBS,* they are highly ranked journals that influence, and also publish articles on, IB; yet they have had different editorial policies towards qualitative research. In contrast to *JIBS*, *AMJ* has published numerous editorials (e.g., Gephart 2004; Pratt 2009; Suddaby 2006) encouraging and providing advice on qualitative submissions. *JMS*, the

most highly ranked European-based management journal,[3] has also promoted discussion on qualitative research standards (e.g., Shah and Corley 2006). The inclusion of these two journals provided us with greater diversity of theorising practices, while at the same time still limiting ourselves to publications that are of relevance to IB scholars.

A central challenge in assembling our dataset was to identify all case studies in the three journals. We categorised articles as case studies if they met the definition proposed by Piekkari et al. (2009: 569): "a research strategy that examines, through the use of a variety of data sources, a phenomenon in its naturalistic context, with the purpose of 'confronting' theory with the empirical world". In order to identify a case study, we read the entire paper, not just its title and abstract. All articles were categorised independently by at least two members of the research team, and differences in opinion led us back to the "raw data", the case study articles themselves. Classification of some of the articles was hampered by the omission of essential details – even a methodological section. Our analysis also confronted the issue that "case study" is a contested term, and difficult to distinguish from other qualitative approaches (Wolcott 2001), so our categorisation of articles did not always agree with that of their authors.[4]

Having settled on a final dataset of 199 case studies (see Appendix), we then proceeded by analysing their contents qualitatively. In contrast to quantitative approaches, whose concern is the enumeration of categories, the aim is a holistic interpretation of the text that goes beyond its literal meaning. This enabled us to remain consistent with the objective of our paper, namely to analyse methods of theorising "in context" rather than "away from context". A more quantitative content analysis would not have been meaningful, because some authors used methodological terms very loosely. For example, "grounded theory" was commonly mentioned as a technique for data analysis and coding, but only rarely as a methodology for inductive theory-building. Therefore a frequency count of this term would have misrepresented the popularity of inductive theory-building approaches in the dataset. Instead of frequency counts, we used other techniques to aid our categorisation of texts (Berelson 1971), chiefly intra-content comparison (i.e., comparing different parts of the

same text, which allowed us to detect common themes as well as inconsistencies), comparisons between different texts (i.e., comparing across articles), and comparison of the textual content with a standard (in our case, our evolving typological categories).

As well as classifying the case studies according to their method of theorising, we wrote an analytical memo about each article, which helped us to proceed systematically and consistently (Miles and Huberman 1994). The memo addressed the following questions:

- Do the authors of the case articles state the theoretical objectives of the study and, if so, how?
- How do they relate theory to empirical data?
- Do they integrate the research context into the theoretical explanation of the case?
- Do they refer to methodological sources, and which ones?
- Do they generalise from case data?
- Do they make causal claims?
- Do they analyse the case holistically, or construct process explanations?
- What theoretical language do they use?

Given the focus of this paper, we did not analyse other aspects of the case study design and write-up, such as the methods of data production or analysis. In sum, the memos encouraged us to take advantage of the strengths of qualitative research to focus on the linguistic elements of the texts, representations of the theorising process, and the assumptions made by their authors.

At least two of us coded each article, independently first and then jointly, and we conducted two separate rounds of coding. The repeated double-coding of all units (i.e., articles) is not a common practice in content analysis, because of the time and cost involved (see Kolbe and Burnett 1991; Potter and Levine-Donnerstein 1999). We took this step because our objective, consistent with our qualitative approach, was to enrich our analysis with multiple interpretations and achieve an inter-subjective understanding across coders (for a similar argument, see Barbour 2001; Krippendorff 2004; Yardley 2000). Any divergence in coding was talked

through, as qualitative content analysis values "the content of disagreements and the insights that discussion can provide for refining coding frames" (Barbour 2001: 1116). Articles that seemed to challenge our existing coding scheme led us to further scrutinise and elaborate on our evolving categories (for a similar approach to coding, see Locke and Golden-Biddle 1997).

Accordingly, our approach to content analysis enabled successive iterations between theory (i.e., the methodological literature) and our dataset. Each iteration led to a modification to and enrichment of our conceptual understanding. Our initial review of the methodological literature generated two broad categories: *positivist* approaches to case studies vs *alternative* traditions. Close textual reading of the case studies in our dataset challenged this dichotomous view, and at one stage of our analytical process we worked with six distinct categories of theorising methods. These were eventually collapsed into four categories, which then required us to elucidate the commonalities and differences between them. The typology we present in this paper was progressively developed in the course of our analysis. After developing the first version of our typology, we conducted a final round of coding in order to refine the key dimensions of the typology and attributes of each category.

Content analysis faces the challenge that interpretivists and critical realists term the *double hermeneutic* (Giddens 1984; Noorderhaven 2004): all research is an interpretation of an already interpreted world. In our study, we were interpreting published studies without additional information about the authors' original motivations or the modifications they made during the review process. Moreover, the four methods of theorising are, of necessity, ideal types: in actual research practice, we found that authors do not necessarily adhere to a consistent philosophical position, methodological approach or even research design. We neither claim that ours is the only possible classification of these articles, nor that our categorisation necessarily matched authorial intentionality.

Having discussed the qualitative, interpretive nature of our analysis, we now present the insights we gained into the four methods of theorising. We turn first to the support for our typology that we found in the methodological literature (i.e., the theoretical foundations of our typol-

ogy), followed by an analysis of how the four methods were used in research practice (i.e., the empirical foundations of our typology).

Constructing the Typology: Theoretical Foundations

In this section we turn to the typology generated from our content analysis, and trace the foundation for each method of theorising to the extant methodological literature on case studies, and to distinctive philosophical traditions. Three of the methods – inductive theory-building, natural experiment and interpretive sensemaking – are well established, while the fourth – which we label contextualised explanation – is a more recent addition to the methodological literature. We provide an overview of each method and its underlying philosophical orientation (see Table 9.1 for a summary), paying particular attention to how explanation and context are framed. Two dimensions of the case study, namely contextualisation and causal explanation, form the basis of our typology, which we bring together in a two-by-two matrix.

Case Study as Inductive Theory-Building

Proponents of this method identify the main potential of the case study as lying in its capacity to induce new theory from empirical data. Eisenhardt (1989), the methodological authority most closely associated with this position (see also Bonoma 1985; Leonard-Barton 1990), explicitly grounds her defence of the case study's inductive strengths in a "positivist view" of science, whose aim is "the development of testable hypotheses and theory which are generalizable across settings" (Eisenhardt 1989: 546; see also Eisenhardt and Graebner 2007: 28).[5] The dilemma is that the small-N case study would seem to be incompatible with this objective, which aspires to uncover regularities or laws of behaviour by emulating the methods of the natural sciences. Eisenhardt (1989) does not question this "law-explanation orthodoxy" (Outhwaite 1987: 7) –

Table 9.1 Comparing four methods of theorising from case studies

Dimension	Inductive theory building	Natural experiment	Interpretive sensemaking	Contextualised explanation
Philosophical orientation	Positivist (empiricist)	Positivist (falsificationist)	Interpretive/ constructionist	Critical realist
Nature of research process	Objective search for generalities	Objective search for causes	Subjective search for meaning	Subjective search for causes
Case study outcome	Explanation in the form of testable propositions	Explanation in the form of cause–effect linkages	Understanding of actors' subjective experiences	Explanation in the form of causal mechanisms
Strength of case study	Induction	Internal validity	Thick description	Causes-of-effects explanations
Attitude to generalisation	Generalisation to population	Generalisation to theory (analytic generalisation)	"Particularisation" not generalisation	Contingent and limited generalisations
Nature of causality	Regularity model: Proposing associations between events (weak form of causality)	Specifying cause–effect relationships (strong form of causality)	Too simplistic and deterministic a concept	Specifying causal mechanisms and the contextual conditions under which they work (strong form of causality)
Role of context	Contextual description a first step only	Causal relationships are isolated from the context of the case	Contextual description necessary for understanding	Context integrated into explanation
Main advocate	Eisenhardt	Yin	Stake	Ragin/Bhaskar

namely, that to explain an event or phenomenon means to subsume it under a general law – rather, she creates a legitimate role for the case study by casting it as the "natural complement" to deductive theory-testing. While case studies cannot provide nomothetic, law-like generalisations, they can generate the theoretical propositions upon which large-scale quantitative testing is based.

Eisenhardt's model of the theorising process is strongly positivist in terms of its empiricism, in that she regards observation as the basis for theory development, and theory induced from data is likely to be more valid as "it closely mirrors reality" (Eisenhardt 1989: 547). She assumes that this process of observation can be objective, with the researcher achieving validity and reliability through the minimisation of bias. Eisenhardt distances her inductive theory-building approach from other qualitative traditions that avoid generalisability and universal claims in favour of "rich, complex description" (Eisenhardt 1989: 546). She regards "contextual description" as "a necessary first step" in case research, but on its own it does not lead to generalisable theory (Eisenhardt 1991: 626). Rather, case researchers need to escape the "idiosyncratic detail" of individual cases and conclude with "only the relationships that are replicated across most or all of the cases" (Eisenhardt and Graebner 2007: 30). Thus there is a shift from context-bound detail to context-free propositions.

Eisenhardt's acceptance that explanation takes the form of law-like generalisations affects her view of causality, as well as of context. While she notes that qualitative data can provide insight into "why or why not" particular relationships occur, this is not the primary focus of enquiry, which is rather to identify generalisable patterns for further testing. She avoids the use of terms such as "causal" or "causation", instead simply referring to "relationships" between variables and constructs. This evokes the regularity model of scientific explanation commonly traced back to the philosopher David Hume: namely, that the goal of scientific explanation is to uncover "constant conjunction" or covariation between variables (Brady 2008). Positivist philosophers of science have frequently been uncomfortable with the notion of causality, given that it is ultimately unobservable and therefore nonempirical: thus Eisenhardt's

avoidance of the concept of causality and preference for covariational terms is in keeping with this tradition (see Abbott 1998, on the same trend in sociology). The main aim of inductive theory-building research is to propose associations between constructs and variables that can then be tested. This can be seen as a weak form of causality, in that it seeks to establish regularities rather than the reasons behind them.

Case Study as Natural Experiment

Yin (2009) agrees that case studies are well suited to exploratory theory-building, but unlike Eisenhardt he does not confine case studies to this early stage in the theorising process. Much of his influential book on case studies (Yin 2009) is devoted to an account of how case studies can be used for "explanatory" rather than "exploratory" purposes. In fact, he regards case studies as best suited to "how and why" questions that "deal with operational links needing to be traced over time" (Yin 2009: 9). "Explanatory" case studies use deductive logic to test propositions, adjudicate among rival explanations, revise existing theories and establish causal relationships; in other words, they are suited to verification and not just discovery of new theory (see also Eckstein 2000). Flyvbjerg (2006: 227) goes so far as to claim that case studies are "ideal" for falsification, which Karl Popper regarded as central to theory development.

Although Yin (2009) is not explicit about his philosophical assumptions, he does not question the goals of generalisability, validity and reliability. However, despite sharing these core philosophical commitments with Eisenhardt, he nevertheless regards the possible contribution of the case study very differently (Table 9.1). In his view, the explanatory logic of the case study shares many features with the laboratory experiment. As a "natural experiment" (Lee 1989), the strength of the case study lies in its high degree of internal validity, so long as appropriate procedures are followed in its design and implementation. Many of the procedures that Yin (2009) advocates – such as replication logic, pattern matching and time-series analysis – are adaptations of experimental techniques. Similarly, his reply to concerns about the case study's generalisability is to argue that, like the experiment, the case study generalises to theoretical

propositions and not to populations (Yin 2009). Lack of statistical generalisability does not preclude case studies from having a strong explanatory contribution to offer.

Yin (2009: 143) concedes that the process for developing causal explanations with a case study "has not been well documented in operational terms". He also notes that "causal links may be complex" (2009: 141) and involve multiple independent and even dependent variables. Yet while case evidence is holistic and complex, attention to the research design and proper application of analytical techniques enable the researcher to converge on a set of causal relationships, isolating them from the broader context of the case. Other cases can then be investigated to establish whether the causal patterns occur as predicted, just as multiple experiments are used to refine and test theory. Given this experimental logic, Yin is comfortable with the use of explicit causal language (see also Hillebrand et al. 2001).

Case Study as Interpretive Sensemaking

The notion that case studies are a form of interpretive sensemaking is part of a rich tradition of "idiographic" rather than nomothetic social science; in other words, a social science that seeks to understand the particular rather than generate law-like explanations. Unlike positivist epistemology, which insists on the unity of the social and natural sciences, interpretive approaches[6] emphasise the uniqueness of the social sciences, in which subjects ascribe meaning to their own behaviour, and researchers are part of the world they study (see Table 9.1). Given that human activity can be understood only by accessing how it is intended and experienced, case researchers in this tradition – echoing a controversy that dates back to the nineteenth century – argue that in the social sciences, the scientific ideal of *erklären* (explaining an action by attributing it to exogenous causal factors) needs to be replaced by *verstehen* (understanding an action through the actor's subjective experience of it) (Johnson and Duberley 2000). Stake (1995: 38), a prominent advocate of interpretive sensemaking (see also Feagin et al. 1991), insists on "the difference between case studies seeking to identify cause and effect relationships and those seeking

understanding of human experience" (see also Prasad 2005, for a similar view in management). Case studies are well suited to the latter, as Lincoln and Guba (1985) argue, because they enable the rich contextual description essential to understanding.

Given this philosophical commitment, Stake (1995) proceeds to question the ideals upheld in positivist case traditions, including generalisability, causality and objectivity. In his view, "particularization" is the goal of case studies: that is, an understanding of the uniqueness of the case in its entirety. In contrast to researchers aiming at generalisable explanations, who seek "to nullify context" and "to eliminate the merely situational", researchers in the interpretive tradition embrace context, narratives and personal engagement on the part of the researcher (Stake 1995: 39, 40). Establishing cause–effect relationships is regarded as "simplistic" in the face of this complexity (Stake 2005: 449); instead, the aim is "thick description" – in other words, an appreciation of how the social context imbues human action with meaning (Table 9.1). Stake also disputes the notion that objectivity on the part of the researcher is possible, and argues that, when adjudicating among competing interpretations, "there is no way to establish, beyond contention, the best view" (Stake 1995: 108).

Case Study as Contextualised Explanation: An Emerging Alternative

In this section, we introduce contextualised explanation to the IB field (see Table 9.1). Given this method is a more recent development, it exhibits less consistency and uniformity than can be found in the other three methods we have profiled. Critical realism forms the ontological basis for this method, but social scientists are still debating how to apply this philosophy in practice. Meanwhile, researchers have pioneered analytical procedures for generating contextualised explanations, such as process tracing and qualitative comparative analysis. However, these scholars do not necessarily show an explicit or consistent philosophical commitment. Accordingly, we discuss the philosophical and methodological innovations associated with contextualised explanation separately.

How to Explain in Context: Philosophical Insights from Critical Realism The philosophical foundation for contextualised explanation is distinct from the other methods of theorising, as it lies in critical realism (see Table 9.1). There are multiple variants of critical realism, so the focus in this paper will be on the most influential: Roy Bhaskar (e.g., 1998), and those who have introduced his philosophies to practising social scientists (for applications to management see, e.g., Ackroyd and Fleetwood 2000; Reed 2005; Tsoukas 1989). Bhaskar is realist in the sense that he acknowledges the existence of a reality that is independent of our perceptions of it, but he also regards our comprehension of reality as theory-laden and subjective, and social phenomena as concept-dependent (in other words, constituted by the meanings we attach to them). Bhaskar regards explanation of social phenomena as being "both causal (as does the positivist) and interpretive (as does the hermeneuticist)" (Collier 1994: 167). In other words, Bhaskar provides a way to reconcile explanation *(erklären)* and understanding *(verstehen).*

Bhaskar rejects the empiricist assumption that sensory observation is the only basis for explanation, instead arguing that causality can be understood only with reference to "transcendental", or unobservable, causal mechanisms. In Bhaskar's philosophy, the concept of "causal mechanism" refers to the causal powers (or liabilities) of objects, structures and entities. Objects (whether physical, human or social) have causal powers by virtue of their intrinsic nature: an object and its causal powers are necessarily or internally related (Sayer 1992). However, in an open system such as that of the social world, the relationship of causal mechanisms to their effects is contingent and external, rather than necessarily and internally related. That is to say, whether a causal mechanism is activated depends on the conditions in which it operates: mechanisms are tendencies that may not be actualised, and even if actualised, may not be empirically observable. Only in a closed system, which is carefully manufactured in an experimental situation, can a causal mechanism potentially be isolated from other generative processes, and regular effects produced and observed. In open systems, in contrast, there can be no symmetry between explanation and prediction: "The same causal power … can

produce different outcomes … [or] different causal mechanisms can produce the same result" (Sayer 2000: 15). This means that explanation needs to account for the spatio-temporal context in which causal mechanisms operate.

As a result, causation is not about the search for event regularities: social scientists need to go beyond events to understand the nature of objects, and cause–effect relationships do not consistently produce regularities in an open system. Causal explanation lies rather in understanding the constituent nature of objects: in other words, what objects are capable of doing. Causal explanations are developed not by *collecting* observations, but rather by *digging* beyond the realm of the observable to understand the necessity inherent in objects (Collier 1994). The appeal to empirical observation – either through inductive theory-building or through deductive theory-testing – does not satisfy a critical realist.

Bhaskar's critical realism rejects the determinism and reductionism that are inherent in the regularity model. He ascribes causal power to human agency: that is, an actor's reasons for acting can play a role in causing that action (Collier 2005; Outhwaite 1987). Yet, at the same time, explanations cannot be reduced solely to human intentionality and agency, because human actors operate within already existing social structures. Social structures condition our actions, yet through our actions we (re-)produce these very social conditions. Explanatory accounts therefore need to encompass human intentionality – the articulated reasons of social actors – as well as an actor's position in the social structure. Therefore, while human action is inherently meaningful and purposeful, a causal explanation cannot be built solely from actors' own understandings and interpretations.

In addition, critical realism challenges the possibility of a purely inductive or deductive process of theory development. Lawson (2003) proposes that an explanation often starts with a surprising contrast, triggered by the realisation that an observed outcome is different from what had been anticipated (provoking the question, "why not X?"). This suggests that a new causal factor is in operation, or the observation domain was not as well understood as initially thought, or existing understandings of

causal mechanisms need to be refined. This is essentially an abductive process: the starting point is a perceived mismatch between an empirical observation and an existing theory, leading to a "redescription" or "recontextualisation" of the phenomenon (Danermark et al. 2002). In this view, theorising is a process not of discovery but of conceptualisation (Sayer 1992).

While critical realism offers a distinctive ontology and epistemology, it does not align itself to a specific research methodology. However, Sayer (1992: 243) argues that enquiries into causes (as opposed to regularities) – typified by questions such as "What produces a certain change?" – require an "intensive" research strategy, typically involving a qualitative, in-depth study of "individual agents in their causal contexts". Accordingly, case studies are well suited to developing causal explanations and "exposing" generative mechanisms (Danermark et al. 2002), while conversely the "explanatory penetration" of "extensive" large-N studies is likely to be weak. Yet the application of critical realism to case studies remains "underdeveloped" (Elger 2010: 256). In the meantime, recent years have seen methodological innovations in case research that question positivist forms of explanation and can be seen as consistent with a critical realist approach (Ragin 2009).

How to Explain in Context: Methodological Insights Consistent with critical realism as a philosophy, methodological approaches to contextualised explanation are concerned with accounting for why and how events are produced. Understanding how the outcome in a particular case was brought about (e.g., "A led to E through steps B, C, D") entails working backwards from events (causes-of-effects explanations) rather than estimating the net effects of causes (effects-of-causes explanations) (Mahoney and Goertz 2006). The technique of working backwards – of identifying the intervening causal process between two "variables" – has been termed process tracing (George and Bennett 2004; Gerring 2007b; Hall 2006). It involves a careful construction of a causal chain of evidence from observations that (unlike much data used in the social sciences) are noncomparable, because they are not from a uniform population (Gerring 2007a). Such an approach to causality has been defended as providing stronger explanatory power than the "weak" correlational form ("if X changed by

a certain amount, then Y will have changed by a related amount") offered by the regularity model (Elliott 2005; George and Bennett 2004; Roberts 1996).

Reconstructing causal chains of events suggests a historical approach, yet history is open to the charge that it only provides an explanation of the particular. Proponents of contextualised explanation question this neat separation between the particular/historical and general/theoretical: "Case studies typically partake of both worlds" (Gerring 2007a: 76). As George and Bennett (2004) contend, generalities are routinely used – and refined – to make sense of the particular (see also Hall 2006). Researchers make sense of particular events by classifying them as belonging to a class or broader phenomenon, and by making reference to existing theories, generalities and known patterns in order to "connect the discrete steps in an explanatory narrative" (Roberts 1996: 54). In the process of iterating between the particular and the general, theories can be refined and reassessed, or even rival explanations proposed. Equally, just as researchers require an understanding of the general to make sense of the particular, so too is the latter essential to explanatory accounts. History, then, is not opposed to general theory; rather, "theory cannot escape history" (Calhoun 1998: 860), in that explanations of actions require them to be situated in "social time" and "social place" (Abbott 1998).

So far, our discussion has assumed that there is a single causal chain or pathway to be investigated. However, case-oriented researchers question the assumption of causal homogeneity made by positivist research traditions; in other words, "the idea that causal factors operate in the same way for all cases" (Ragin 2000: 51). Instead of regarding causation as uniform, Ragin (2000; Rihoux and Ragin 2009) proposes a "multiple conjunctural" view as the foundation for case-based research. By "conjunctural", he means that case researchers explain by factoring in the combination of conditions found in the case rather than seeking to measure the net effect of an isolated variable. This is because a single variable may have a very different effect, depending on the configuration of variables with which it is combined in a case. Thus, for example, in combination with A and C, B may cause Y, but in other circumstances Y may occur only in B's

absence (expressible in Boolean algebra as Y = (A AND B AND C) OR (NOT B) AND D AND E). Understanding the effect of B therefore requires putting it in its spatial-temporal context. Because B may produce one effect in a particular context, but a different effect in another situation, "it is not useful to generalize about the overall effect of B without saying something about the context (i.e., other variable values) in which B appears" (Mahoney and Goertz 2006: 235). Ragin (e.g., Rihoux and Ragin 2009) has proposed a formal method based on set theory, qualitative comparative analysis, in order to analyse cases holistically as combinations of conditions. Ragin (2000) positions this "configurational view" as having a fundamentally different explanatory logic from that of the positivist approach, which assumes away causal heterogeneity.

As well as being "conjunctural" in nature, causality is "multiple", given that the same outcome may be produced by different causal pathways (also known as equifinality) (Rihoux and Ragin 2009). Again, Ragin argues that causality should be conceived in set-theoretic rather than probabilistic terms. In Boolean algebra, multiple causation can be expressed in terms such as Y = (A AND B) OR (C AND D) (Mahoney and Goertz 2006). Given the heterogeneous nature of causation, case researchers have concluded that generalisations are not universalities; they are always necessarily limited. Generalisation therefore involves "careful setting of scope" (Byrne 2009: 9): causal explanations require an understanding of the conditions under which they do – and do not – operate. Researchers can aim for no more than "contingent generalizations"; in other words, propositions such as "if circumstances A, then outcome O" (Gerring 2007a). In this way, what Eisenhardt regards as "idiosyncratic detail", to be removed from the explanation, now becomes part of the causal fabric of an explanatory account.

The Four Methods Compared

Figure 9.1 is a visual representation of our typology of case study theorising. The four methods of theorising are positioned in relation to each other on the basis of whether their emphasis is on causal explanation (*y*-axis) and/or contextualisation (*x*-axis). The figure encapsulates our

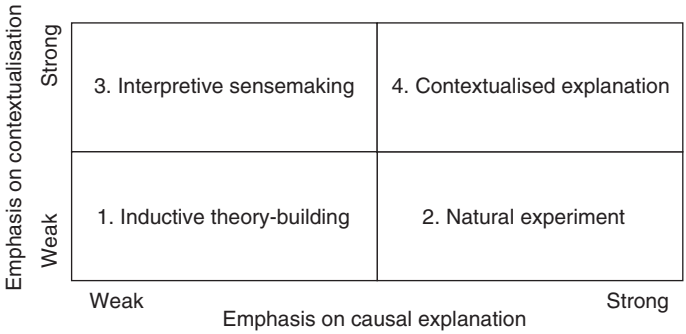

Fig. 9.1 Four methods of theorising from case studies

argument as to why contextualised explanation offers potential. In the inductive theory-building approach (Quadrant 1), the emphasis on both causal explanation and contextualisation is weak. This avoidance of causal claims and context can be traced back to the pursuit of nomothetic, law-like generalisations, which privileges the search for regularities rather than causes, and for context-free rather than context-sensitive knowledge. The established alternatives to inductive theorising redress these limitations, but only by accepting the traditional trade-off between causality and contextualisation. The method of the natural experiment (Quadrant 2) is a welcome development, in that it provides a defence of the case study's ability to generate causal, internally valid explanations. However, as we have seen, the emphasis on contextualisation is still weak. In the sensemaking tradition (Quadrant 3), the "rich story" that troubles Eisenhardt and Graebner (2007) is transformed into the case study's main strength, but at the cost of any claims to causal explanation. All three methods reinforce rather than question longstanding divisions in the social sciences: between *erklären* and *verstehen,* explanation and understanding, nomothetic and idiographic, objective and subjective, inductive and deductive, general and particular, context-free and context-specific.

The fourth method of theorising, contextualised explanation (Quadrant 4), represents an escape from the explanation–contextualisation trade-off. This method of theorising is based on the assertion that case studies

can generate causal explanations that preserve rather than eradicate contextual richness. Proponents insist that explanatory accounts are necessarily context-bound: as Sayer (1992: 60) has written, "making sense of events requires that we 'contextualize' them in some way". While positivist traditions abstract away from time and place, contextualised explanation is a way of explaining "without laws" (see, e.g., Abbott 1997). The regularity model of causation is rejected in favour of a more complex understanding that recognises the contingent nature of cause–effect relationships.

In this section, we have confined our attention to how the four methods of theorising have been developed in the methodological literature on case studies, as well as their philosophical foundations. The question still remains, however, as to how the four methods in our typology are used in research practice. Accordingly, we now shift our focus from the ideal types discussed by methodologists and philosophers (how researchers *should* use case studies to theorise) to theorising practices (how researchers *do* use case studies to theorise). We will argue that grounding our typology in research practice allows for a more nuanced understanding of each method and its potential in future IB research.

Constructing the Typology: Empirical Foundations

In this section, we will discuss the insights from our case study dataset into each method of theorising, reporting on findings about each category in our typology (see Fig. 9.1). Inductive theory-building was clearly the most popular of the four methods of theorising, as Table 9.2 shows.

Table 9.2 Number of case studies in AMJ, JIBS and JMS, 1999–2008

Theorising typology	AMJ	JIBS	JMS	Total
Inductive theory-building	23	12	61	96
Natural experiment	1	5	21	27
Interpretive sensemaking	8	0	44	52
Contextualised explanation	5	2	17	24
Total	37	19	143	199

JMS was the exception among the journals, in having both the highest number of case studies and the highest percentage of case studies that used alternatives to theory-building. For each method of theorising we highlight the key elements of the theorising process that have been central to our discussion throughout the paper: philosophical assumptions (while authors' philosophical orientation was typically not made explicit, it was reflected in how they represented the theorising process), causality, context and generalisability.

Quadrant 1: Inductive Theory-Building (*N* = 96)

The case studies that we classified as falling into this quadrant were all positivist in their assumptions, and identified their theoretical contribution as being exploratory. Despite these similarities, this category was the most diverse of the four. In particular, articles differed in the extent of theory development that they reported had occurred prior to entering the field. At one extreme, we identified articles characterised by a grounded theory approach, in which the introduction would often be followed by the methods section instead of the literature review to underline the inductive nature of theorising (e.g., Pratt et al. 2006). While most papers in this category were written up deductively, authors would demonstrate the inductive nature of their work by presenting a model or a set of propositions as an outcome rather than as a starting point of the research (e.g., Harvey et al. 2002; Maitlis and Lawrence 2003). At the other extreme were those, such as Danis and Parkhe (2002), who while positioning their contribution as theory-building, nonetheless included a priori propositions. In between were authors such as Gilbert (2005), who refined existing theory. While he adhered to the Eisenhardt approach in a faithful way, it was more common for authors to "cherrypick": that is, apply some guidelines from Eisenhardt, but not use her model as a template. While Yin (2009) was also heavily cited by Gilbert (2005) and others in this quadrant, the references were to his exploratory (rather than explanatory) case study.

We would argue that most articles in this quadrant used the term "inductive" very loosely. Moreover, only a few articles made the process

of "inductive" theory-building explicit and transparent. We found an exception in Tolich, Kenney, and Biggart (1999: 594), who "from their interviews … were able to draw out inductively what [they] believe were four variables that had a significant effect [on X] …" Similarly, Denis et al. (2000) distinguished between the "skeletal conceptual framework" with which they commenced the study, and the insights they gained from their fieldwork. Otherwise, it seemed that the majority of the authors were not explicit in specifying which of their insights were inductively derived from their fieldwork and which were more theory-driven.

While inductive theory-building articles typically did not explicitly aim to study causal chains or relationships, they nevertheless used causal vocabulary, as can be judged from expressions such as "influenced", "interacts with", "critical determinants of", "centrally facilitative in", "leads to", "trigger" (e.g., Côté et al. 1999; Faems et al. 2008; Harvey et al. 2002; Tolich et al. 1999). This causal vocabulary typically conformed to that of the regularity model, with authors using terms such as "associated with" and "moderating influence" (Wilkinson et al. 2001; Wong and Ellis 2002), while propositions were framed in correlational terms such as "the higher … the greater" (Büchel 2000).

The articles in this category were united in their descriptive treatment of the research context as a first step before analysing data. Thus these authors tended to have a separate section outlining the research setting, but the discussion of it remained at a descriptive rather than analytical level (e.g., Boxall and Steeneveld 1999). Context is a feature of these studies, but not as a means of providing explanation. Despite this dominant trend, we did identify authors who showed greater sensitivity towards context. They tended to pursue process research (e.g., Côté et al. 1999) or grounded theorising (e.g., Ravasi and Schultz 2006; Rodrigues and Child 2008). However, in these studies the rich process data did not translate into process theorising: for example, Maritan (2001) concludes with propositions that are correlational rather than processual in nature.

A common concluding point for articles in this quadrant was the standard warning about the lack of generalisability due to the small-N nature of the study and the context-specific nature of the findings. Like Eisenhardt, many authors in this quadrant portray their contribution as being a "bridge" to theory-testing, and end with a series of propositions

(e.g., Coviello 2006; Orr and Scott 2008). In such articles, the setting of the study can be portrayed as a limitation: thus Noda and Collis (2001) describe the "context … of the study" as potentially introducing bias (see also Beverland 2005). Some case researchers were less apologetic, and merely made the transferability of their findings a matter for discussion (e.g., Boxall and Steeneveld 1999). However, few case researchers in this quadrant specified the contexts to which the findings could be transferred, or why. An exception can be found in Pratt et al. (2006: 259), who, while presenting the "unique nature" of the setting as a weakness in their quest for generalisability, nevertheless argue that "careful" transferability is possible, and specify the settings that offer "easy-to-see parallels" to their own study.

Quadrant 2: Natural Experiment (*N* = 27)

Authors in this quadrant subscribed to positivist assumptions, but they positioned their theoretical contribution as being theory-testing, applying an established theory or providing rival explanations (although they typically did not explicitly use the metaphor of a natural experiment). Buck and Shahrim (2005) test a causal proposition by using a "least likely" case (as recommended by George and Bennett 2004), which they selected "in order to maximise the possibility of refutation" (Buck and Shahrim 2005: 58). At the same time, Markóczy (2000) is a clear example of articles that offer a rival explanation to that favoured by existing theory. She concludes with a very bold statement – namely, that cultural differences are overrated as an explanatory factor – and offers what she terms "novel factors" in accounting for beliefs. A "rival explanation" study can also take the form of a re-evaluation of a previously reported case that questions how it has been interpreted in existing literature (e.g., Howells 2002). These different forms of challenging existing theory were not widely used in our dataset. Nor were articles that sought to apply existing theory in order to provide a causal account, Collinson and Rugman (2008) being an exception. Accordingly, we would argue that the ability of the case study to modify, verify, test and challenge existing theory and offer rival explanations has been underexploited.

In this quadrant we detected the frequent use of causal language, which was often more explicit and pronounced compared with inductive theory-building articles. For example, Taylor (1999: 858) states as his objective "to determine what affects the degree of control … and to what extent control is related to [X] …". The correlational associations between variables are the dominant approach to causal relationships in this paper, but the author also concedes that "there is a complex pattern" at play, "one in which a combination of factors … affect the outcome" (Taylor 1999: 866). While Taylor's (1999) variable-oriented approach to causality was typical of this category, the authors of one paper developed a more nuanced argument regarding the level of "path-effects", concluding that while "historical patterns of development" were relevant, they "do not fully explain the present" (Hill et al. 2000).

Authors whose studies were classified into this category dealt with context in different ways. In some articles, the empirical context was seen to offer the advantage of a "natural" laboratory setting, as suggested by our label for this quadrant. In light of this aim, researchers took great care to select the naturally occurring conditions that were the most appropriate test of the theory, given that, unlike in an artificial laboratory setting, the environment cannot be controlled by the researcher. For example, De Boer, Van den Bosch, and Volberda (1999: 389) chose publishing firms moving into multimedia because they are "right at the heart of the evolutionary process in which technologies stemming from various industries are converging into hybrid forms". While we found almost decontextualised case studies in this category (e.g., Stiles 2001), we identified others which drew on contextual factors to generate explanation. For example, Ogbonna and Harris (2002) succeed in enriching a case study of two change initiatives by offering a context-sensitive account as to why the cross-case differences had occurred. In their study, context enhances the internal validity of the study, since it controls for environmental effects and promotes a replication logic. Yet, even in these articles, context was decomposed into a set of variables that had an effect on the phenomenon under study, rather than capturing the influence of context more holistically. Moreover, context tended to be de-emphasised or even isolated from the findings and conclusions, given that authors made generalising

claims and provided context-free models as their contribution (e.g., Johnson-Cramer et al. 2003).

Quadrant 3: Interpretive Sensemaking (*N* = 52)

Researchers in this quadrant, particularly those adhering to a social constructivist approach (Hodgson 2002; McCabe 2000; Watson and Watson 1999), were typically very explicit about their philosophical stance and their rejection of positivist assumptions. A distinct approach to reflecting on and representing the theorising process could often be found in these papers. While some authors used descriptors such as "exploratory" and "theory-building", in many other articles, a rigid distinction between theory and evidence was not upheld; instead these two elements were interwoven throughout the paper, in keeping with the belief that theory and observation cannot be separated (Chreim 2005; McInerney 2008). The authors of one paper explicitly described the theorising process as "a mixture of both deductive and inductive methods" (Noon et al. 2000: 504). More commonly, authors did not use either "induction" or "deduction" to denote their theorising. However, there were articles that claimed to be interpretive but nevertheless revealed traces of the positivist theory-building tradition, for example by acknowledging the shortcomings of a single case study in developing generalisable theory (e.g., Coupland and Brown 2004).

Authors presented their theoretical objectives in terms of illuminating and providing insight, for example: "we are interested in the worldviews of organizational members" (Maguire and Phillips 2008: 380). Portraying worldviews is not just a descriptive effort; Ram (1999) provides a rich narrative which is infused with theoretical concepts. These authors would often start their article with a vignette or a personal encounter from the field. Researchers with a more social constructivist approach sought not only to understand participants' meaning, but also how these meanings were constructed. For example, Yakura (2002) shows how time is "constructed" in multiple ways in a consulting firm, while Lindgren and Packendorff (2006: 841) view project work "as an ongoing construction of patterns of femininity and masculinity in society".

We found no explicit reference to causes, unless authors were referring to the causal models employed or constructed by research participants. However, we encountered extensive use of causal and explanatory language (e.g., Heracleous and Barrett 2001), despite the absence (in the main) of positivist language. Instead, another vocabulary was in use: managers "enact", power has "effects", hegemony is "produced" and meaning is "constructed" (Barry et al. 2001; Benjamin and Goclaw 2005). Salaman and Storey (2002: 163) conclude that the managers they studied "are both producers as well as products of the corporate culture": in other words, they address the agency-structure question discussed above in relation to critical realism. Narratives, which were a commonly used device, were also much more than descriptions; rather they had deep, explanatory purposes, as Ng and de Cock (2002: 40) state explicitly: "Story interpretation requires an answer to the deceptively simple question: 'Why did things turn out the way it did?'" In this way, as Kakkuri-Knuutila et al. (2008) have observed, "to understand" and "to explain" are not as opposed as they may seem.

Researchers in this quadrant tended to include themselves as part of the context, rather than taking an objective stance. In particular, authors adopting a social constructivist approach often did not just seek to examine how research participants constructed meaning, but also scrutinised their own sensemaking: "we unashamedly present our paper as our own construction and are happy to make explicit the discursive resources which we bring both to our research design and data analysis" (Watson and Watson 1999: 485). Similarly, Dick and Cassell (2002: 958) add that "the researcher's own role in the production of knowledge needs to be accounted for". Authors who took this approach did not seek to claim neutrality: "we rejected the idea that an objectively verifiable reality can be accessed through research" (Dick and Cassell 2002: 960–961), and presented theirs as just one possible "reading" (Chreim 2005: 589). Generalisability was often not discussed or even, in the case of Ng and de Cock (2002: 43), rejected explicitly: they state there is no need "to provide law-like theories with their attractive elegance and highly glossed accounts".

Quadrant 4: Contextualised Explanation (*N* = 24)

Overall, case studies that emphasised causal explanation (Quadrants 2 and 4 in Fig. 9.1) were in the minority. In our content analysis, we paid attention to how authors in this quadrant were able to combine the inherent strength of the case study to contextualise with its explanatory potential. Compared with the "natural experiment" quadrant, these articles aimed to generate explanation, but without strong adherence to positivist assumptions. While most of the authors did not explicitly reveal their philosophical assumptions, we found evidence of social constructivism/interpretivism (e.g., Ogbonna and Wilkinson 2003), and mild positivism (e.g., Wilkinson et al. 2001), as well as three papers whose authors were explicitly drawing on the critical realist tradition (Chung 2001; O'Mahoney 2007; Pajunen 2006).

Although there was variation in terms of how authors presented their theorising process, they tended not to separate theory-building and theory-testing. For example, Jacobides (2005: 486) – one of the few articles with the ambitious scope of directly challenging an established theory – searched for new analytical insights inductively, "without being bound by existing theory". Sminia (2003) aims to explain why a TV channel failed. This seemingly atheoretical purpose is countered by a discussion of how existing explanations are insufficient to shed light on failed ventures. The author takes a more deductive approach in combining insights from existing theoretical perspectives to arrive at a novel, more convincing explanation. Several papers fall in between these two extremes, as they recognise both inductive and deductive elements in theorising. Denis, Lamothe, and Langley (2001: 812), for example, write that "[o]ur approach was and is also partly deductive (theory inspired) and partly inductive (data inspired)".

In this quadrant, authors were more open about the explanatory aims of their paper (e.g., Sminia 2003; White and Liu 2001). Clark and Soulsby (1999: 537) set out to "offer a feasible and credible explanation of the spread of the [multidivisional organisation] in the Czech Republic". In a similar vein, Farjoun (2002: 848) builds "an explanatory model of institutional development". The explanatory aim is also made clear by

West (2008: 1508), who "sought to explain the commercialization of Shannon's theory during its first quarter century". In other articles "explanation" was not used, with authors (Mota and Castro 2004; O'Mahoney 2007) instead referring to "opposing and driving forces", "nonlinear relationships" and "multiple pathways" – all terms suggesting a complex view of causality.

Despite these differences, what typifies the authors' language is a very particular view of causality as a complex and dynamic set of interactions that are treated holistically. For example, Perlow et al. (2002) introduce "mutual causality" and "causal loop diagrams" to capture the connections between speed and decision-making in an Internet start-up. Jacobides (2005: 492), in turn, states that "[t]his is a study of a particular industry and … it focuses on understanding the causal dynamics of a particular setting". White and Liu (2001) offer "alternative transition trajectories" for firms operating within the industry under study. As we have shown, there is a well-established causal vocabulary in the critical realist tradition, but references to this literature are rarely made. O'Mahoney (2007: 1345) is one of the few to adhere explicitly to the critical realist tradition in developing an "explanatory theory". Another author, Chung (2001), uses Ragin's Boolean algebra to systematically compare cases.

For the authors of these articles context was a necessity, not a problem, in constructing rich explanations. As Perlow et al. (2002: 949) write, "[o]ur findings suggest the importance of examining decisions and their relationship within the context in which they happen". In this category theorising was viewed primarily as a localised explanation. Similarly, Clark and Soulsby (1999: 555) weave context into their theoretical interpretation: "the roles of institutional and strategic choice factors could only be understood in their mutual interaction". In their paper, explaining in context took the form of a real-time processual analysis that was very different from the process studies found in the inductive theory-building quadrant. The context was used to generate an explanation for the motives of the managers in the study. They argue that factors typically treated in other studies as "independent variables" should rather be understood as a "recursive

process" (Clark and Soulsby 1999: 556). In this quadrant, history and process become essential to developing a causal account. In defending her historical methodology, Farjoun (2002: 871) argues that "by its nature, historical analysis particularly attends to continuity and process, to diverse influences and to context". Denis et al. (2001: 815) emphasise that the time periods they identify are not predictable stages, but allow for more complex explanations such as "multidirectional causality".

Based on their in-depth knowledge of the context, authors in this category discuss the scope of the generalisable claims they are making, and identify specific contextual aspects that would expand – rather than reduce – the transferability of their findings. For example, White and Liu (2001: 122) conclude that "[d]eregulating or privatizing industries and breaking up monopolies represent environmental contexts that are conceptually similar to China's transitional economy in which the central plan has been discarded". Other authors warn against "overgeneralizing", such as Jacobides (2005), who emphasises the industry-specificity of his study. Finally, some authors in this category did not seek generalisability at all. Instead, they "sought to embrace all the richness and complexity of a real … setting" to generate a localised explanation and invite the reader to evaluate the applicability of their results in other situations (Ogbonna and Wilkinson 2003: 1159).

Overall, our content analysis revealed considerable versatility in theorising practices, and has pointed to variations within each typological category. We identified articles with greater context sensitivity, even in the positivist quadrants that typically have a weak emphasis on contextualisation. Equally, case studies using inductive theory-building and interpretive sensemaking methods might make causal claims, despite not acknowledging this explicitly, and despite differences in the causal vocabulary in use. Overall, our findings suggest that case researchers lack an established vocabulary to express the theorising process or its outcome. In the following section we will consider the wider implications of these findings.

Conclusion and Future Directions

In this paper, we have sought to expand the possibilities for theorising from case studies in IB research by constructing a typology of theorising methods. To enrich the future of qualitative research in IB, we have employed qualitative research ourselves – an in-depth qualitative content analysis of 199 case study articles published in three journals during 1999–2008. Our typology – which contrasts inductive theory-building, interpretive sensemaking, natural experiment and contextualised explanation – was developed by iterating between the existing methodological literature and a dataset of published case studies. We would argue that this typology enriches the potential of the case study, both for contextualising and for producing causal explanations. In this concluding section we draw out these possibilities, and the implications for future case research in IB. Consistent with the aims of our paper, our focus is on theorising, even though we recognise there are other dimensions to the case studies we analysed.

We commenced this paper by arguing that the current dominance of inductive theory-building in *JIBS* (as indeed, in the other journals we analysed) may be hindering the potential for case studies to contextualise theory and generate causal explanations. As we have discussed, the call for greater contextualisation of IB theory has been repeatedly made by scholars in the field (Brannen and Doz 2010). However, we have argued that the dominant method of inductive theorising places little emphasis on context: articles in this tradition treated context descriptively rather than analytically. In this method, context is seen as a limitation, given that the goal is law-like explanation. Simply conducting more case studies – which has been advocated in the literature on contextualising IB theory – would therefore not necessarily lead to more context-sensitive theorising. Rather, scholars need to consider the implications of their choice of theorising method carefully, because these methods differ in their emphasis on contextualisation.

As we have seen, the decontextualised nature of inductive theory-building studies is not the only concern. In addition, the widespread assumption that the role of the case study lies only in the exploratory,

theory-building phase of research downplays its potential to propose causal mechanisms and linkages, and test existing theories. The danger is that these legitimate uses of the case study are underutilised or even questioned. In IB, Yin (2009) has been used largely to justify the exploratory role of case studies, overlooking the strong emphasis he places on explanatory case studies. In this paper we have highlighted this neglected dimension of Yin's work, and have argued that the case study has an important role to play in refining, verifying, testing and challenging existing theory. Our content analysis revealed examples of case studies that effectively performed this role and placed a strong emphasis on causal explanation, although they were in the minority. This application of the case study is worth further examination, given its potential to interrogate existing theories.

Set against this background, we have proposed that the method of contextualised explanation, while rarely found in our dataset of published case studies, holds promise in that it offers a high degree of contextualisation without sacrificing the goal of causal explanation. We would argue that, above all, the value of this approach lies in its different view of how to generate theories about the social world: the rejection of the regularity model of causation, scepticism towards the possibility of meaningful law-like generalisations, and a defence of context as being an essential component of, rather than a hindrance to, explanation. As a result of redefining the theorising process in this way, proponents of contextualised explanation seek to explain "without laws". They offer a way of reconciling context and explanation by acknowledging the complexity of the social world, the bounded scope and contingency of causal relationships, and the simultaneous operation of multiple interaction effects. The possibility of such a reconciliation is also an abiding theme in IB, given that as a field its *raison d'être* is to explain phenomena in diverse national, cultural and institutional contexts.

While we suggest that greater application of contextualised explanation would benefit the IB field, we are not, however, advocating it as *the* sole method for theorising from case studies. We are not seeking to replace one method (i.e., inductive theory-building) with another (i.e., contextualised explanation). Rather, we have adopted a pluralist stance: that a field benefits from the diversity of, and even tension between,

different approaches. A broad paradigmatic consensus restricts methodological innovation and limits the range of methods, research problems, choice of topics – and ultimately theories. Yet while the research phenomena in IB are diverse in nature, we have not detected the same degree of diversity in theorising about these phenomena. The pursuit of law-like explanations remains the taken-for-granted approach to theorising in IB, leading to decontextualised research methods and consequently decontextualised theories (Brannen and Doz 2010). In this paper we have argued that the goal of more contextualised theories requires IB scholars to rethink their assumptions about the role of the case study. The search for greater pluralism led us beyond IB to examine key journals in the general management field, allowing us to gain insights into theorising from other research traditions.

Our paper contributes to a more pluralist future for IB by offering alternative methods of theorising from case studies. We have demonstrated that case researchers have a choice about how to theorise, just as they have a choice about how to produce and analyse data. Our typology encourages researchers to reflect critically upon their own and others' preconceived views of how to theorise from case studies; to explore possibilities for theorising that go beyond that of inductive theory-building; and to escape the trade-off between internal validity and thick description that is found in positivist and interpretive paradigms. From the perspective of those reviewing case research, the typology provides the means to evaluate theoretical contributions. By articulating and more explicitly specifying their method of theorising, researchers can foster greater mutual understanding of the theoretical purpose of their studies.

Yet at the same time as illustrating the flexibility that the case study methodology offers with respect to theorising, we have also suggested the need for coherence when reporting the theorising process. Each method of theorising entails distinct philosophical assumptions, research objectives and outcomes, as well as a vocabulary for describing the theorising process and articulating a study's theoretical contribution. By scrutinising the linguistic elements of texts, we found that case researchers were not always clear and consistent in the way they wrote up their theorising purpose and process. Methodological rigour has traditionally been reflected in selecting methods and research designs that fit the research

question of the study. Based on our content analysis, we would argue that methodological rigour is also evidenced by methodological self-awareness, transparency and careful linguistic choices in reporting the theorising process.[7] We hope that one contribution of our paper is to encourage case researchers to (re)consider their own approach to theorising from case studies, and be consistent in following their choice throughout their study.

In a pluralist field of research, greater methodological consistency and adherence to a particular method of theorising need not restrict authors from engaging with, learning from and being influenced by other methods. Thus, while we have highlighted the benefits for IB of exploring critical realism, our content analysis has implications for researchers operating within positivist and interpretive traditions as well. For those researchers taking a more positivist approach, we would suggest that considerable potential lies in exploiting the range of "natural experiments" that we identified in our content analysis: testing theory, proposing rival explanations, reanalysing cases, and applying or challenging existing theory. Such case studies go beyond the posing of covariational propositions, to providing explanations for causal relationships. In addition to strengthening the emphasis on causal explanation, there is also potential for researchers operating within positivist traditions to be sensitised to context when theorising from case studies. In our content analysis we found that authors achieved this by detailing how contextual factors produced the outcome, and how their findings might be transferred to other settings. These authors were able to use context to specify the boundary conditions of their explanations.

For researchers taking an interpretive approach, we would stress the advantages to recognising and making more explicit the explanatory fabric that permeates their contributions. In our content analysis we found that, while researchers in this tradition used a causal language different from that associated with the other methods of theorising, interpreting and understanding the social world also involves offering an explanation as to why events occur in the way they do. Some authors did succeed in sensitively combining contextualised explanation with a range of philosophical traditions, including interpretivism and even a moderate form of

positivism that seeks limited, contingent generalisations rather than universalities. Therefore we feel there is potential in exploring how insights from contextualised explanation could inform research more broadly.

In the years since the publication of Eisenhardt's (1989) article on theorising from case studies, greater appreciation has emerged in the methodological literature as to how authors can explain in context. At its best, contextualised explanation can provide novel theoretical accounts that incorporate rather than deny complexity. While IB scholars have so far not contributed to the emerging methodological debate on case studies, we would hope that the next ten years of *JIBS* will see a growing diversity and innovation in approaches to theorising. In a more pluralist field, case researchers would approach theorising differently. They would move beyond the conformity to the inductive theory-building that prevailed in the *JIBS* case studies of the previous decade. At the same time, they would question the trade-off between internal validity and thick description that characterises both positivist and interpretive paradigms. In this more pluralist scenario, case researchers, regardless of their paradigmatic stance, would be able to combine context sensitivity with explanatory rigour in their theorising. By arguing that contextualisation and rigorous explanation can be complementary rather than contradictory outcomes, we have proposed a future for the case study in IB that stands in contrast to the limited role to which it has traditionally been assigned.

Acknowledgements We thank Snejina Michailova, Marja-Liisa Kakkuri-Knuuttila, Kalle Pajunen, Kristina Rolin, Ben Tipton, Denice Welch, Lawrence Welch, Ian Wilkinson and Yorgos Zotos for their helpful comments on earlier versions of this manuscript; as well as students in our PhD case research classes at Aalto University and University of Sydney. Particular thanks are due to Geoff Easton and Ricardo Morais, who have been responsible for our interest in critical realism. Early versions of the paper benefited from feedback at the 2009 European International Business Academy annual conference, and at Academy of Management symposia in 2009 and 2010. We are extremely grateful for the constructive and thoughtful guidance provided by Mary Yoko Brannen as special issue editor, as well as the anonymous *JIBS* reviewers.

Appendix

Table 9.3 Categorisation of journal articles 1999–2008

Journal	Year	Quantitative	Mixed	Qualitative[a]	Case	Cases % of empirical[b]	Cases % of total[b]	Articles per volume[c]	NE[d]
AMJ	1999	40	0	0	0	0.0	0.0	43	3
AMJ	2000	60	4	4	2	2.9	2.8	72	2
AMJ	2001	57	2	3	9	12.7	12.7	71	0
AMJ	2002	59	1	1	8	11.6	11.6	69	0
AMJ	2003	44	0	1	2	4.3	4.2	48	1
AMJ	2004	50	1	0	3	5.6	5.5	55	1
AMJ	2005	52	0	0	5	8.8	6.9	72	15
AMJ	2006	47	0	5	3	5.5	4.9	61	6
AMJ	2007	52	1	3	3	5.1	3.5	85	26
AMJ	2008	49	1	3	2	3.6	3.6	55	0
Total		510	10	20	37	6.4	5.9	631	54
JIBS	1999	22	0	0	2	8.3	6.5	31	7
JIBS	2000	26	3	0	1	3.3	3.3	30	0
JIBS	2001	27	1	1	1	3.3	2.6	39	9
JIBS	2002	25	0	0	4	13.8	12.5	32	3
JIBS	2003	26	0	0	1	3.7	2.9	35	8
JIBS	2004	16	0	0	2	11.1	8.7	23	5
JIBS	2005	24	0	0	2	7.7	6.1	33	7
JIBS	2006	35	1	0	1	2.5	1.9	52	12
JIBS	2007	47	1	1	2	3.7	3.0	66	12
JIBS	2008	56	0	2	3	5.0	3.9	76	16
Total		304	6	4	19	5.6	4.6	417	79
JMS	1999	9	0	4	13	50.0	30.2	43	17

(*continued*)

Table 9.3 (continued)

Journal	Year	Quantitative	Mixed	Qualitative[a]	Case	Cases % of empirical[b]	Cases % of total[b]	Articles per volume[c]	NE[d]
JMS	2000	13	1	8	9	29.0	19.1	47	16
JMS	2001	15	2	5	11	33.3	22.9	48	15
JMS	2002	14	0	4	13	41.9	27.1	48	17
JMS	2003	20	1	11	25	43.9	31.3	80	23
JMS	2004	16	0	8	15	38.5	22.4	67	28
JMS	2005	19	1	5	19	43.2	27.9	68	24
JMS	2006	21	2	5	13	31.7	20.0	65	24
JMS	2007	20	0	7	12	30.8	17.1	70	31
JMS	2008	30	1	5	13	26.5	22.0	59	10
Total		177	8	62	143	36.7	24.0	595	205
Grand total		991	24	86	199	15.2	12.1	1643	338

[a]Excluding case studies
[b]Average values
[c]Excluding editorials, commentaries and notes
[d]Non-empirical

Notes

1. Although we recognise that case studies can be mixed and even quantitative (for a discussion in IB see, e.g., Nummela and Hurmerinta-Peltomäki 2006), in this paper we are concerned with case studies as a qualitative research strategy.
2. In this paper we follow interpretivists and critical realists in acknowledging that research is an act of interpretation. We use terms such as data "production" and typology "construction", rather than seeking to conceal the role of the researcher.
3. We decided to include a journal originating in Europe because it has been suggested that case study traditions are more firmly established there than in the US (Bengtsson et al. 1997).
4. An illustrative example is the article by Nutt (2000), who positioned his study as a multiple case investigation ($N = 376$) of strategic decisions. This paper was not included in our analysis, since it treated the cases as observations, rather than investigating the phenomenon in its natural setting.
5. There are many variants of positivism (Halfpenny 1982), including logical positivism, logical empiricism and falsificationism. The similarities rather than the differences among these traditions are our focus in this paper. However, it is worth noting that we would characterise Eisenhardt's empiricism as distinct from the assumptions behind the "natural experiment" approach, which rather follows a falsification logic, as advocated by Karl Popper (for a discussion of empiricism vs falsificationism, see Johnson and Duberley 2000).
6. In this paper we will use "interpretive" in a broad sense to refer to research traditions that include postmodernism, postcolonialism, critical theory and social constructivism.
7. We thank an anonymous reviewer for this insight.

References

Abbott, A. 1997. Of time and space: The contemporary relevance of the Chicago school. *Social Forces* 75 (4): 1149–1182.

———. 1998. The causal devolution. *Sociological Methods & Research* 27 (2): 148–181.

Ackroyd, S., and S. Fleetwood, eds. 2000. *Realist perspectives on management and organisations*. London: Routledge.

Bamberger, P. 2008. Beyond contextualization: Using context theories to narrow the micro-macro gap in management research. *Academy of Management Journal* 51 (5): 839–846.

Barbour, R. 2001. Checklists for improving rigour in qualitative research: A case of the tail wagging the dog? *British Medical Journal* 322 (5): 1115–1117.

Barry, J., J. Chandler, and H. Clark. 2001. Between the ivory tower and the academic assembly line. *Journal of Management Studies* 38 (1): 87–101.

Bengtsson, L., U. Elg, and J.-I. Lind. 1997. Bridging the transatlantic publishing gap: How north American reviewers evaluate European idiographic research. *Scandinavian Journal of Management* 13 (4): 473–492.

Benjamin, O., and R. Goclaw. 2005. Narrating the power of nonstandard employment: The case of the Israeli public sector. *Journal of Management Studies* 42 (4): 737–759.

Berelson, B. 1971. *Content analysis in communication research*. 2nd ed. Glencoe: Free Press.

Beverland, M.B. 2005. Crafting brand authenticity: The case of luxury wines. *Journal of Management Studies* 45 (5): 1003–1029.

Bhaskar, R. 1998. *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. 3rd ed. London: Routledge.

Blair, J.D., and J.G. Hunt. 1986. Getting inside the head of the management researcher one more time: Context-free and context-specific orientations in research. *Journal of Management* 12 (2): 147–166.

Bonoma, T.V. 1985. Case research in marketing: Opportunities, problems, and a process. *Journal of Marketing Research* 22 (2): 199–208.

Boxall, P., and M. Steeneveld. 1999. Human resource strategy and competitive advantage: A longitudinal study of engineering consultancies. *Journal of Management Studies* 36 (4): 443–454.

Brady, H.E. 2008. Causation and explanation in social science. In *The Oxford handbook of political methodology*, ed. B. Steffenmeier, H.E. Brady, and D. Collier, 217–270. Oxford: Oxford University Press.

Brannen, M.Y., and Y. Doz. 2010. From a distance and detached to up close and personal: Bridging strategic and crosscultural perspectives in international management research and practice. *Scandinavian Journal of Management* 26 (3): 236–247.

Büchel, B. 2000. Framework of joint venture development: Theory-building through qualitative research. *Journal of Management Studies* 37 (5): 637–661.

Buck, T., and A. Shahrim. 2005. The translation of corporate governance changes across national cultures: The case of Germany. *Journal of International Business Studies* 36 (1): 42–61.

Byrne, D. 2009. Case-based methods: Why we need them; what they are; how to do them. In *The Sage handbook of case-based methods*, ed. D. Byrne and C.C. Ragin, 1–10. London: Sage.

Calhoun, C. 1998. Explanation in historical sociology: Narrative, general theory, and historically specific theory. *American Journal of Sociology* 104 (3): 846–871.

Cappelli, P., and P.D. Sherer. 1991. The missing role of context in OB: The need fora meso-level approach. In *Research in organizational behavior*, ed. B.M. Staw, vol. 13, 55–110. Stanford: JAI Press.

Child, J. 2000. Theorizing about organization cross-nationally. In *Advances in international comparative management*, ed. J.L.C. Cheng and R.B. Peterson, vol. 13, 27–75. Stanford: JAI Press.

Chreim, S. 2005. The continuity–change duality in narrative texts of organizational identity. *Journal of Management Studies* 42 (3): 567–593.

Chung, C. 2001. Markets, culture and institutions: The emergence of large business groups in Taiwan, 1950s–1970s. *Journal of Management Studies* 38 (5): 719–745.

Clark, E., and A. Soulsby. 1999. The adoption of the multidivisional form in large Czech enterprises: The role of economic, institutional and strategic choice factors. *Journal of Management Studies* 36 (4): 535–559.

Clark, T., and M. Wright. 2007. Reviewing journal rankings and revisiting peer reviews: Editorial perspectives. *Journal on Management Studies* 44 (4): 612–621.

Collier, A. 1994. *Critical realism: An introduction to Roy Bhaskar's philosophy*. London: Verso.

———. 2005. Critical realism. In *The politics of method in the human sciences: Positivism and its epistemological others*, ed. G. Steinmetz, 327–345. Durham: Duke University Press.

Collinson, S., and A.M. Rugman. 2008. The regional nature of Japanese multinational business. *Journal of International Business Studies* 39 (2): 215–230.

Côté, L., A. Langley, and J. Pasquero. 1999. Acquisition strategy and dominant logic in an engineering firm. *Journal of Management Studies* 36 (7): 919–952.

Coupland, C., and A.D. Brown. 2004. Constructing organizational identities on the web: A case study of Royal Dutch/Shell. *Journal of Management Studies* 41 (8): 1325–1347.

Coviello, N.E. 2006. The network dynamics of international new ventures. *Journal of International Business Studies* 37 (5): 713–731.

Danermark, B., M. Ekström, L. Jakobsen, and J.C. Karlsson. 2002. *Explaining society: Critical realism in the social sciences*. London: Routledge.

Danis, W.M., and A. Parkhe. 2002. Hungarian–Western partnerships: A grounded theoretical model of integration processes and outcomes. *Journal of International Business Studies* 33 (3): 423–455.

De Boer, M., F.A.J. Van den Bosch, and H.W. Volberda. 1999. Managing organizational knowledge integration in the emerging multimedia complex. *Journal of Management Studies* 36 (3): 379–398.

Denis, J.-L., A. Langley, and M. Pineault. 2000. Becoming a leader in a complex organization. *Journal of Management Studies* 37 (8): 1063–1099.

Denis, J.-L., L. Lamothe, and A. Langley. 2001. The dynamics of collective leadership and strategic change in pluralistic organizations. *Academy of Management Journal* 44 (4): 809–837.

Dick, P., and C. Cassell. 2002. Barriers to managing diversity in a UK constabulary: The role of discourse. *Journal of Management Studies* 39 (7): 953–976.

Duriau, V.J., R. Reger, and M.D. Pfarrer. 2007. A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational Research Methods* 10 (5): 5–34.

Eckstein, H. 2000. Case study and theory in political science. In *Case study method*, ed. R. Gomm, M. Hammersley, and P. Foster, 119–164. London: Sage.

Eisenhardt, K.M. 1989. Building theories from case study research. *Academy of Management Review* 14 (4): 532–550.

———. 1991. Better stories and better constructs: The case for rigor and comparative logic. *Academy of Management Review* 16 (3): 620–627.

Eisenhardt, K.M., and M.E. Graebner. 2007. Theory building from cases: Opportunities and challenges. *Academy of Management Journal* 50 (1): 25–32.

Elger, T. 2010. Critical realism. In *Encyclopedia of case study research*, ed. A.J. Mills, G. Durepos, and E. Wiebe, vol. 1, 253–557. Los Angeles: Sage.

Elliott, J. 2005. *Using narratives in social research*. London: Sage.

Faems, D., M. Janssens, A. Madhok, and B. van Looy. 2008. Toward an integrative perspective on alliance governance: Connecting contract design, trust dynamics, and contract application. *Academy of Management Journal* 51 (6): 1053–1078.

Farjoun, M. 2002. The dialectics of institutional development in emerging and turbulent fields: The history of pricing conventions in the on-line database industry. *Academy of Management Journal* 45 (5): 848–874.

Feagin, J.R., A.M. Orum, and G. Sjoberg. 1991. *A case for the case study*. Chapel Hill: University of North Carolina Press.

Flyvbjerg, B. 2006. Five misunderstandings about case-study research. *Qualitative Inquiry* 12 (2): 219–245.

George, A.L., and A. Bennett. 2004. *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.

Gephart, R. 2004. From the editors: Qualitative research and the *Academy of Management Journal*. *Academy of Management Journal* 47 (4): 454–462.

Gerring, J. 2007a. *Case study research: Principles and practices*. New York: Cambridge University Press.

———. 2007b. The mechanismic worldview: Thinking inside the box. *British Journal of Political Science* 38: 161–179.

Giddens, A. 1984. *The constitution of society: Outline of the theory of structuration*. Berkeley: University of California Press.

Gilbert, C. 2005. Unbundling the structure of inertia: Resource versus routine rigidity. *Academy of Management Journal* 48 (5): 741–763.

Halfpenny, P. 1982. *Positivism and sociology: Explaining social life*. London: George Allen & Unwin.

Hall, P.A. 2006. Systematic process analysis: When and how to use it. *European Management Review* 3 (1): 24–31.

Harvey, J., A. Pettigrew, and E. Ferlie. 2002. The determinants of research group performance: Towards mode 2? *Journal of Management Studies* 39 (6): 747–774.

Heracleous, L., and M. Barrett. 2001. Organizational change as discourse: Communicative actions and deep structures in the context of information technology implementation. *Academy of Management Journal* 44 (4): 755–778.

Hill, S., R. Martin, and M. Harris. 2000. Decentralization, integration and the post-bureaucratic organization: The case of R&D. *Journal of Management Studies* 37 (4): 563–585.

Hillebrand, B., R.A.W. Kok, and W.G. Biemans. 2001. Theory-testing using case studies: A comment on Johnston, leach, and Liu. *Industrial Marketing Management* 30 (8): 651–657.

Hodgson, D. 2002. Disciplining the professional: The case of project management. *Journal of Management Studies* 39 (6): 803–821.

Howells, J. 2002. The response of old technology incumbents to technological competition: Does the sailing ship effect exist? *Journal of Management Studies* 39 (7): 887–906.

Hsieh, H.-F., and S.E. Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative Health Research* 15 (9): 1277–1288.

Jack, G.A., M.B. Calás, S.M. Nkomo, and T. Peltonen. 2008. Critique and international management: An uneasy relationship? *Academy of Management Review* 33 (4): 870–884.

Jacobides, M.G. 2005. Industry change through vertical disintegration: How and why markets emerged in mortgage banking. *Academy of Management Journal* 48 (3): 465–498.

Johns, G. 2006. The essential impact of context on organizational behavior. *Academy of Management Review* 31 (2): 386–408.

Johnson, P., and J. Duberley. 2000. *Understanding management research*. London: Sage.

Johnson-Cramer, M.E., R.L. Cross, and A. Yan. 2003. Sources of fidelity in purposive organizational change: Lessons from a re-engineering case. *Journal of Management Studies* 40 (7): 1837–1870.

Kakkuri-Knuutila, M.L., K. Lukka, and J. Kuorikoski. 2008. Straddling between paradigms: A naturalistic philosophical case study on interpretive research in management accounting. *Accounting, Organizations and Society* 33 (2–3): 267–291.

Kellert, S.H., H.E. Longino, and C.K. Waters, eds. 2006. *Scientific pluralism*. Minneapolis: University of Minnesota Press.

Kolbe, R.H., and M.S. Burnett. 1991. Content analysis research: An examination of applications with directives for improving research reliability and objectivity. *Journal of Consumer Research* 18 (2): 243–250.

Krippendorff, K. 2004. *Content analysis. An introduction to its methodology*. 2nd ed. Thousand Oaks: Sage.

Lawson, T. 2003. *Reorienting economics*. London: Routledge.

Lee, A.S. 1989. Case studies as natural experiments. *Human Relations* 42 (2): 117–137.

Leonard-Barton, D. 1990. A dual methodology for case studies: Synergistic use of a longitudinal single site with replicated multiple sites. *Organization Science* 1 (3): 248–266.

Lincoln, Y.S., and E.G. Guba. 1985. *Naturalistic inquiry*. Newbury Park: Sage.

Lindgren, M., and J. Packendorff. 2006. What's new in new forms of organizing? On the construction of gender in project-based work. *Journal of Management Studies* 43 (4): 841–866.

Locke, K., and K. Golden-Biddle. 1997. Constructing opportunities for contribution: Structuring intertextual coherence and "problematizing" in organizational studies. *Academy of Management Journal* 40 (5): 1023–1062.

Maguire, S., and N. Phillips. 2008. Citibankers at Citigroup: A study of the loss of institutional trust after a merger. *Journal of Management Studies* 45 (2): 372–401.

Mahoney, J., and G. Goertz. 2006. A tale of two cultures: Contrasting qualitative and quantitative research. *Political Analysis* 14 (3): 227–249.

Maitlis, S., and T.B. Lawrence. 2003. Orchestral manoeuvres in the dark: Understanding failure in organizational strategizing. *Journal of Management Studies* 40 (1): 109–139.

Maritan, C.A. 2001. Capital investment as investing in organizational capabilities: An empirically grounded process model. *Academy of Management Journal* 44 (3): 513–531.

Markóczy, L. 2000. National culture and strategic change in belief formation. *Journal of International Business Studies* 31 (3): 427–442.

McCabe, D. 2000. Factory innovations and management machinations: The productive and repressive relations of power. *Journal of Management Studies* 37 (7): 931–953.

McInerney, P.-B. 2008. Showdown at Kykuit: Field-configuring events as loci for conventionalizing accounts. *Journal of Management Studies* 45 (6): 1089–1116.

Miles, M.B., and M. Huberman. 1994. *Qualitative data analysis: An expanded sourcebook*. 2nd ed. Thousand Oaks: Sage.

Morgan, G. 1980. Paradigms, metaphors, and puzzle solving in organization theory. *Administrative Science Quarterly* 25 (4): 605–622.

Mota, J., and L.M. Castro. 2004. A capabilities perspective on the evolution of firm boundaries: A comparative case example from the Portuguese moulds industry. *Journal of Management Studies* 41 (2): 295–316.

Ng, W., and C. de Cock. 2002. Battle in the boardroom: A discursive perspective. *Journal of Management Studies* 39 (1): 23–49.

Noda, T., and D.J. Collis. 2001. The evolution of intraindustry firm heterogeneity: Insights from a process study. *Academy of Management Journal* 44 (4): 897–925.

Noon, M., S. Jenkins, and M.M. Lucio. 2000. Fads, techniques and control: The competing agendas of TPM and TECEX at the Royal Mail (UK). *Journal of Management Studies* 37 (4): 499–520.

Noorderhaven, N.G. 2004. Hermeneutic methodology and international business research. In *Handbook of qualitative research methods for international business*, ed. R. Marschan-Piekkari and C. Welch, 84–108. Cheltenham: Edward Elgar.

Nummela, N., and L. Hurmerinta-Peltomäki. 2006. Mixed methods in international business research: A value-added perspective. *Management International Review* 46 (4): 439–459.

Nutt, P.C. 2000. Decision-making success in public, private and the third sector organizations: Finding sector dependent best practice. *Journal of Management Studies* 37 (1): 77–108.

O'Mahoney, J. 2007. The diffusion of management innovations: The possibilities and limitations of memetics. *Journal of Management Studies* 44 (8): 1324–1348.

Ogbonna, E., and L.C. Harris. 2002. Organizational culture: Aten year, two-phase study of change in the UK food retailing sector. *Journal of Management Studies* 39 (5): 673–706.

Ogbonna, E., and B. Wilkinson. 2003. The false promise of organizational culture change: A case study of middle managers in grocery retailing. *Journal of Management Studies* 40 (5): 1151–1178.

Orr, R.J., and W.R. Scott. 2008. Institutional exceptions on global projects: A process model. *Journal of International Business Studies* 39 (4): 562–588.

Outhwaite, W. 1987. *New philosophies of social science: Realism, hermeneutics and critical theory*. Basingstoke: Macmillan.

Pajunen, K. 2006. Stakeholder influences in organizational survival. *Journal of Management Studies* 43 (6): 1261–1288.

Perlow, L.A., G.A. Okhuysen, and N.P. Repenning. 2002. The speed trap: Exploring the relationship between decision making and temporal context. *Academy of Management Journal* 45 (5): 931–955.

Piekkari, R., C. Welch, and E. Paavilainen. 2009. The case study as disciplinary convention: Evidence from international business journals. *Organizational Research Methods* 12 (3): 567–589.

Potter, W.J., and D. Levine-Donnerstein. 1999. Rethinking validity and reliability in content analysis. *Journal of Applied Communication Research* 27 (3): 258–284.

Prasad, P. 2005. *Crafting qualitative research: Working in the postpositivist traditions*. Armonk: M.E. Sharpe.

Pratt, M. 2009. From the editors: For the lack of a boilerplate: Tips on writing up and (reviewing) qualitative research. *Academy of Management Journal* 52 (5): 856–862.

Pratt, M.G., K.W. Rockmann, and J.B. Kaufmann. 2006. Constructing professional identity learning cycles in the customization of identity among medical residents. *Academy of Management Journal* 49 (2): 235–262.

Ragin, C.C. 2000. *Fuzzy-set social science*. Chicago: University of Chicago Press.

———. 2009. Reflection of casing and case-oriented research. In *The sage handbook of case-based methods*, ed. D. Byrne and C.C. Ragin, 522–534. London: Sage.

Ram, M. 1999. Managing consultants in a small firm: A case study. *Journal of Management Studies* 36 (6): 875–897.

Ravasi, D., and M. Schultz. 2006. Responding to organizational identity threats: Exploring the role of organizational culture. *Academy of Management Journal* 49 (3): 433–458.

Redding, G. 2005. The thick description and comparison of societal systems of capitalism. *Journal of International Business Studies* 36 (2): 123–155.

Reed, M. 2005. Reflections on the "realist turn" in organization and management studies. *Journal of Management Studies* 42 (8): 1621–1644.

Rihoux, B., and C.C. Ragin, eds. 2009. *Configurational comparative methods*. Thousand Oaks: Sage.

Roberts, C. 1996. *The logic of historical explanation*. University Park: Pennsylvania State University Press.

Rodrigues, S., and J. Child. 2008. The development of corporate identity: A political perspective. *Journal of Management Studies* 45 (5): 885–911.

Rousseau, D.M., and Y. Fried. 2001. Location, location, location: Contextualizing organizational research. *Journal of Organizational Behavior* 22 (1): 1–13.

Ruben, D.H. 1990. *Explaining explanation*. London: Routledge.

Ryan, G.W., and H.R. Bernard. 2000. Data management and analysis methods. In *Handbook of qualitative research*, ed. N.K. Denzin and Y.S. Lincoln, 2nd ed., 769–802. Thousand Oaks: Sage.

Salaman, G., and J. Storey. 2002. Managers' theories about the process of innovation. *Journal of Management Studies* 39 (2): 147–165.

Sayer, A. 1992. *Method in social science: A realist approach*. 2nd ed. London: Routledge.

———. 2000. *Realism and social science*. London: Sage.

Shah, S., and K.G. Corley. 2006. Building better theory by bridging the quantitative-qualitative divide. *Journal of Management Studies* 43 (8): 1821–1835.

Shapiro, D.L., M.A. Von Glinow, and Z. Xiao. 2007. Toward polycontextually sensitive research methods. *Management and Organization Review* 3 (1): 129–152.

Sminia, H. 2003. The failure of the Sport7 TV-channel: Controversies in a business network. *Journal of Management Studies* 40 (7): 1621–1649.

Stake, R.E. 1995. *The art of case study research*. Thousand Oaks: Sage.

———. 2005. Qualitative case studies. In *Handbook of qualitative research*, ed. N.K. Denzin and Y.S. Lincoln, 3rd ed., 443–466. Thousand Oaks: Sage.

Stiles, P. 2001. The impact of the board on strategy: An empirical examination. *Journal of Management Studies* 38 (5): 627–650.

Suddaby, R. 2006. From the editors: What grounded theory is not. *Academy of Management Journal* 49 (4): 633–642.

Taylor, B. 1999. Patterns of control within Japanese manufacturing plants in China: Doubts about Japanization in Asia. *Journal of Management Studies* 36 (6): 853–873.

Tolich, M., M. Kenney, and N. Biggart. 1999. Managing the managers: Japanese management strategies in the USA. *Journal of Management Studies* 36 (5): 587–607.

Tsoukas, H. 1989. The validity of idiographic research explanations. *Academy of Management Review* 14 (4): 551–561.

Tsui, A.S. 2004. Contributing to global management knowledge: A case for high quality indigenous research. *Asia Pacific Journal of Management* 21 (4): 491–513.

———. 2006. Contextualization in Chinese management research. *Management and Organization Review* 2 (1): 1–13.

———. 2007. From homogenization to pluralism: International management research in the academy and beyond. *Academy of Management Journal* 50 (6): 1353–1364.

Van Maanen, J. 1995. Fear and loathing in organization studies. *Organization Science* 6 (6): 687–692.

Watson, T., and D. Watson. 1999. Human resourcing in practice: Managing employment issues in the university. *Journal of Management Studies* 36 (4): 483–504.

West, J. 2008. Commercializing open science: Deep space communications as the lead market for Shannon theory, 1960–73. *Journal of Management Studies* 45 (8): 1506–1532.

Whetten, D.A. 2009. An examination of the interface between context and theory applied to the study of Chinese organizations. *Management and Organization Reviews* 5 (1): 29–55.

White, S., and X. Liu. 2001. Transition trajectories for market structure and firm strategy in China. *Journal of Management Studies* 38 (1): 103–124.

Wilkinson, B., J. Gamble, J. Humphrey, J. Morris, and D. Anthony. 2001. The new international division of labour in Asian electronics: Work organization and human resources in Japan and Malaysia. *Journal of Management Studies* 38 (5): 675–695.

Wolcott, H.F. 2001. *Writing up qualitative research*. 2nd ed. Thousand Oaks: Sage.

Wong, P.L.-K., and P. Ellis. 2002. Social ties and partner identification in Sino-Hong Kong international joint ventures. *Journal of International Business Studies* 33 (2): 265–289.

Yakura, E.K. 2002. Charting time: Timelines as temporal boundary objects. *Academy of Management Journal* 45 (5): 956–970.
Yardley, L. 2000. Dilemmas in qualitative health research. *Psychology and Health* 15 (2): 215–228.
Yin, R.K. 2009. *Case study research*. 4th ed. Thousand Oaks: Sage.

10

Theorizing from Cases: A Commentary

Kathleen M. Eisenhardt

Welch, Piekkari, Plakoyiannaki, and Paavilainen-Mäntymäki (2011) raise the critical issue of *context* in international business research. Context differences such as across nations and cultures are often enormously powerful, and yet frequently not well conceptualized. Their article serves an important role by putting the spotlight on context. In addition, the article offers other contributions beyond this central one. Like others (Yin 1984; Eisenhardt 1989), they note that case studies can range from theory building to theory testing. Yet the article also reminds us that most case research remains near the inductive end of the spectrum, and that there is often an inappropriate conflation of cases with theory building that is limiting the use of cases. This is problematic because case methods can be particularly helpful in bringing context into international business

Welch, C., R. Piekkari, E. Plakoyiannaki, and E. Paavilainen-Mäntymäki. 2011. Theorising from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies* 42 (5): 740–762.

K. M. Eisenhardt (✉)
Stanford University, Stanford, CA, USA
e-mail: kme@stanford.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_10

research. The article also provides a useful reminder of the relevance of causal explanations, and the importance of a pluralistic view of methods that recognizes the strengths of each. Finally, the article contributes an analysis of the use of case study methods in several major journals, indicating just how widely used these methods have become.

While these contributions are helpful, the article's central typology and 2 × 2 create artificial distinctions. Its interpretation of theory building cases combines cherry-picked phrases with an eighteenth century view of positivism. In reality, theory building from cases does not assume that researchers are fully objective, observation is the sole basis of theory development, or that context should be dismissed.

Instead, the article itself misses context—that is, by failing to recognize that early writing on the theory building from cases occurred in the context of contrasting with arm-chair theorizing and justifying the method vis à vis the dominant paradigm of econometric theory testing. The piece also misses context around causal language—that is, role of the prevailing norms of the time, field, journal, or even specific reviewers. These practical considerations are often at play rather than a philosophical stance inherent in the method itself.

Theory building from cases also does not dismiss context, but rather focuses on the major facets (but not every facet) of the context, given the focal research question and data. Consistent with theory building, boundary (i.e., scope) conditions are always germane (although not always provided) and tie tightly with context.

Finally, the article misses the close relationship between theory building from cases (Eisenhardt 1989; Eisenhardt and Graebner 2007) and Yin's (2009) work. Since theory building from cases builds on Yin's original work (1984), it uses replication logic, embedded units of analysis, the notion of experiments, pattern matching, and so on. Overall, there was and still is broad agreement that cases can be used for theory building or theory testing.

Since 2011, the use of case study methods has advanced. Researchers are more frequently using intermediate forms of theorizing like abduction and elaboration between the endpoints of theory building and testing (Bingham and Eisenhardt 2011). They are choosing along this spectrum primarily according to the research question and depth of

existing research related to it. Exemplary theory building from cases (and case research broadly) has often used and now increasingly sharpens causal mechanisms, consistent with the broad trend towards causal identification throughout organizations and strategy research. These mechanisms, for example, indicate the underlying theoretical logic that links constructs together.

Also consistent with the broad trend towards causal identification, case study researchers are more frequently using natural experiments such as racing research designs, and identifying the configurations, equifinality, and complex causality that may occur in the data (Battilana and Dorado 2010; Hallen and Eisenhardt 2012; Pache and Santos 2013; Kaplan et al. 2016). The point is that the use of theory building from cases and broadly case methods is diverse and increasingly rich.

In addition, inductive methods (i.e., approaches to generate theoretical insight from data) have advanced since 2011. Influential methodologists are rejecting attempts to define artificial boxes and boundaries. For example, Walsh and colleagues including Barney Glaser (Walsh et al. 2015) offer a critique of work that tries to define and confine grounded theorizing too narrowly and excessively. Other researchers are combining discussions of interpretivist and case approaches to provide a more accurate understanding of the actual differences (Gehman et al. 2018). Indeed, a consensus is emerging that the various inductive methods are complements (Eisenhardt et al. 2016). The three core methods—that is, cases, interpretivist, and ethnography—have substantial similarities such as rich data about focal phenomena over time, theoretical sampling, and grounded theory building.

A consensus is also emerging about the differences among the three core methods (Eisenhardt et al. 2016). *Theory building from cases* (Eisenhardt 1989) blends insights from the logics of cases (Yin 2009) and grounded theorizing (Glaser and Strauss 1967). While initially framed as positivist, philosophically diverse researchers use the method. They view cases as independent experiments where replication logic is germane and seek to develop underlying theoretical logic (Eisenhardt and Graebner 2007). Many researchers use multiple cases and provide both theoretical and actionable insights.

Interpretivist studies take a naturalist view, emphasizing that knowledge and understanding are socially constructed (Gioia et al. 2013). This method values the authentic representation of people's perception of their own lived experience. Thus, the focus is often on concepts like identity, sense making, and sense giving with an emphasis on data from interviews that give informants voice.

Ethnography has its roots in understanding culture (Van Maanen 1988). It typically relies on observations to reveal day-to-day practices, often in ways that people cannot or will not express (Bechky 2011). These observations may illuminate non-verbal cues, rituals and ceremonies, the role of artefacts, and the use of physical space. As such, these observations clarify interaction patterns and ways of working.

Overall, many methodologists are spending less effort advocating for a specific method and instead are trying to escape narrow definitions and confining boxes. There is an emerging recognition of the similarity across inductive methods as well as the relevance of specific methods for different types of research questions and contexts (Gehman et al. 2018). An interpretivist approach works particularly well when the research question involves the lived experiences of participants. As noted above, interpretivists often focus on concepts like identity and sense making and are concerned with how those participants socially construct reality. Ethnography is particularly relevant for research questions that involve culture and the nature of work. Theory building from cases addresses the broadest array of potential research questions and accommodates questions that explore differences across cases (e.g., variance such as performance differences) and similarities. Often although not always, the emphasis is frequently on processes and changes over time.

At the same time, context remains a critical issue in research, one that Welch et al. (2011) highlight for international business research. An important and long-standing approach for research regarding context is to be explicit about assumptions and boundary conditions. In my own work, for example, identifying *assumptions* and *boundary conditions* has been particularly helpful in distinguishing the established firm v. venture context and the nascent v. mature market context.

A subtler approach to taking context seriously is diligence in the choice of *construct names*. Much research conversation occurs in English, but

particular concepts that are germane in contexts using other languages (e.g., guanxi in China, the notion of face in Japan) may not have an English language equivalent. Such constructs require particular care in naming and defining. Similarly, common constructs that are well understood in English like trust and authority may not translate well to other languages and cultures. Measures pose a similar challenge.

That said, since the various inductive methods (e.g., theory building from cases, interpretivist) make limited use of a priori specification of constructs and relationships, they are likely to be particularly helpful in addressing novel contexts. They readily enable the researcher to develop fresh, context-relevant constructs. For example, inductive methods may offer better understanding of under-studied contexts and related constructs such as for family businesses (e.g., highly relevant in much of South America) and unique roles of the state (e.g., China) than deductive, econometrics-based methods.

Finally, machine learning and big data offer substantial promise for the future. Indeed, there is already activity around machine learning to expand and improve case methods and inductive methods, broadly. For example, topic modelling is quickly emerging as a useful approach for creating and measuring constructs from large textual data sets. Combining small N case studies with machine learning using large quantitative data sets is emerging as a new way to develop and elaborate theory (Tidhar and Eisenhardt 2019).

Theory building from cases and machine learning are unexpectedly similar—that is, both focus on pattern recognition. Indeed, in many ways, machine learning is simply an algorithmic approach to the well-known (albeit somewhat artful) inductive process of cross-case theory building that researchers typically use. Yet, the approaches also complement each other—that is, theory building from cases can effectively identify constructs and causal mechanisms; while machine learning can provide quantitative assessments as well more accurate representations of equifinal paths and configurations, such as by using decision trees and other techniques that rely on regularization.

In summary, Welch et al. (2011) raise the critical issue of context that is particularly germane in international business research. Their article wisely points to the importance of case methods to shed light on context.

The article also reminds readers to use cases for more than just theory building, consider causality, and appreciate the pluralism of methods. Although its typology creates some very inaccurate distinctions and three of the 2 × 2 cells seem better understood as one, the article moves international business research ahead.

Finally, the use of cases for theory building (and beyond) has become richer and more diverse since 2011. The discourse around inductive methods is becoming less about pursuing minor (even incorrect) distinctions, and more broadly about similarities, genuine differences, and useful contingencies. Pluralism among researchers is indeed alive and well even as machine learning arrives.

References

Battilana, J., and S. Dorado. 2010. Building sustainable hybrid organizations: The case of commercial microfinance organizations. *Academy of Management Journal* 53: 1419–1440.

Bechky, B.A. 2011. Making organizational theory work: Institutions, occupations, and negotiated orders. *Organization Science* 22: 1157–1167.

Bingham, C.B., and K.M. Eisenhardt. 2011. Rational heuristics: The "simple rules" that strategists learn from process experience. *Strategic Management Journal* 32 (13): 1437–1464.

Eisenhardt, K.M. 1989. Building theories from case study research. *Academy of Management Review* 14: 532–550.

Eisenhardt, K.M., and M.E. Graebner. 2007. Theory building from cases: Opportunities and challenges. *Academy of Management Journal* 50: 25–32.

Eisenhardt, K.M., M.E. Graebner, and S. Sonenshein. 2016. Grand challenges and inductive methods: Rigor without rigor mortis. *Academy of Management Journal* 59: 1113–1123.

Gehman, J., V.L. Glaser, K.M. Eisenhardt, D. Gioia, A. Langley, and K.G. Corley. 2018. Finding theory–method fit: A comparison of three qualitative approaches to theory building. *Journal of Management Inquiry* 27 (3): 284–300.

Gioia, D.A., K.G. Corley, and A.L. Hamilton. 2013. Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods* 16: 15–31.

Glaser, B.G., and A.L. Strauss. 1967. *The discovery of grounded theory*. Chicago: Aldine.

Hallen, B.L., and K.M. Eisenhardt. 2012. Catalyzing strategies and efficient tie formation: How entrepreneurial firms obtain investment ties. *Academy of Management Journal* 55: 35–70.

Kaplan, S., J. Milde, and R.S. Cowan. 2016. Symbiotic practices in boundary spanning: Bridging the cognitive and political divides in interdisciplinary research. *Academy of Management Journal* 60 (4): 1387–1414.

Pache, A., and F. Santos. 2013. Inside the hybrid organization: Selective coupling as a response to competing institutional logics. *Academy of Management Journal* 56: 972–1001.

Tidhar, R., and K.M. Eisenhardt. 2019. Get rich or die trying…Unpacking revenue model choice using machine learning and multiple-case theory building. *Working paper.* Stanford University.

Van Maanen, J. 1988. *Tales of the field: On writing ethnography*. Chicago: University of Chicago Press.

Walsh, I., J.A. Holton, L. Bailyn, W. Fernandez, N. Levina, and B. Glaser. 2015. Rejoinder: Moving the management field forward. *Organizational Research Methods* 18: 620–628.

Welch, C., R. Piekkari, E. Plakoyiannaki, and E. Paavilainen-Mäntymäki. 2011. Theorising from case studies: Towards a pluralist future for international business research. *Journal of International Business Studies* 42 (5): 740–762.

Yin, R.K. 1984. *Case study research*. Beverly Hills: Sage Publications.

———. 2009. *Case study research*. Vol. 4. Thousand Oaks: Sage.

11

Theorizing from Cases: Further Reflections

Eriikka Paavilainen-Mäntymäki, Rebecca Piekkari, Emmanuella Plakoyiannaki, and Catherine Welch

Recently, social scientists have been rethinking how to generate theory from qualitative data, and we hope our 2011 *JIBS* article has contributed to this broader trend. In our paper we addressed the limitations of inductive theory building and advocated the use of other approaches, or

E. Paavilainen-Mäntymäki
Turku School of Economics, University of Turku, Turku, Finland
e-mail: eriikka.paavilainen-mantymaki@utu.fi

R. Piekkari (✉)
School of Business, Aalto University, Espoo, Finland
e-mail: rebecca.piekkari@aalto.fi

E. Plakoyiannaki
Leeds University Business School, University of Leeds, Leeds, UK
e-mail: E.Plakoyiannaki@leeds.ac.uk

C. Welch
The University of Sydney Business School, The University of Sydney, Sydney, NSW, Australia
e-mail: catherine.welch@sydney.edu.au

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_11

methods, of theorizing to enhance the explanatory power of qualitative research by harnessing the potential of contextualization.

We see three ways in which the themes of our 2011 paper have been further developed by others since its publication. We will canvas them briefly.

First, there is increasing consensus about the limits of induction in explaining how theory is generated from case data. Critics argue that induction, on its own, cannot explain theory generation (Dubois and Gadde 2002, 2014; Klag and Langley 2013; Van Maanen et al. 2007). Instead, contemporary writings are converging on a view of theorizing more as an abductive, non-linear process where theory and data are successively reinterpreted. While induction insists on theory-free observation, abduction emphasizes a theoretical starting point. Only a solid grounding in existing theory when entering the field allows the researcher to appreciate the limits of existing explanations, uncover deep structures, and generate novel theoretical insights (Dubois and Gadde 2014). The abductive approach is gaining traction because it offers qualitative researchers a vocabulary to articulate how they iterate between theory and data.

Second, calls for diverse voices and traditions in theorizing have intensified. There is increasing interest in alternative approaches, such as Qualitative Comparative Analysis (QCA), that provide more holistic explanations, dissolving the dichotomy between qualitative and quantitative research. *JIBS* has published a few papers applying QCA (see Crilly, Chap. 14, this volume), but this is not the only option. Cornelissen (2017a, b) highlights the potential of narrative or process-based theorizing and thick description as rich methods or 'styles' of theorizing, which deserve more attention. These theorizing styles will require IB researchers to embrace rarely used methodologies, such as longitudinal single cases, allowing them to uncover the causal complexity of social behaviour. Greater acceptance and use of historical methods and retrospective data would also help researchers to trace causal mechanisms over time and develop process explanations addressing how social change emerges and evolves (see Welch, Chap. 13, this volume).

Third, we have seen studies appearing in *JIBS* that offer both strong contextualization and strong explanation (the upper right corner of our

2 × 2 typology). This means researchers wishing to reconcile explanation and contextualization now have more concrete examples to follow. One such example is Geary and Aguzzoli (2016), who skilfully use the extended case method approach to bridge the micro and macro in their work. They explain how their findings would simply not have been possible without placing the phenomenon—the influence of institutional factors—in its context 'not only of the micro-politics of the firm but also the wider macro-political terrain' (p. 987). The trend of combining contextualization with causal explanation is also visible in strategy and management research more generally (see e.g., Sminia 2016). We hope that more such examples will follow in *JIBS*, realizing the possibilities we set out in the 2011 paper in future research.

References

Cornelissen, J. 2017a. Editor's comments: Developing propositions, a process model, or a typology? Addressing the challenges of writing theory without a boilerplate. *Academy of Management Review* 42 (1): 1–9.

———. 2017b. Preserving theoretical divergence in management research: Why the explanatory potential of qualitative research should be harnessed rather than suppressed. *Journal of Management Studies* 54 (3): 368–383.

Dubois, A., and L.-E. Gadde. 2002. Systematic combining: An abductive approach to case research. *Journal of Business Research* 55 (7): 553–560.

———. 2014. "Systematic Combining" – A decade later. *Journal of Business Research* 67 (6): 1277–1284.

Geary, J., and R. Aguzzoli. 2016. Miners, politics and institutional caryatids: Accounting for the transfer of HRM practices in the Brazilian multinational enterprise. *Journal of International Business Studies* 47 (8): 968–996.

Klag, M., and A. Langley. 2013. Approaching the conceptual leap in qualitative research. *International Journal of Management Reviews* 15 (2): 149–166.

Sminia, H. 2016. Contextualized explanation in strategy research. In *Research methods for strategic management*, ed. M.C. Cinici and G.B. Dagnino, 78–97. Milton Park: Routledge.

Van Maanen, J., J.B. Sørensen, and T.R. Mitchell. 2007. The interplay between theory and method. *Academy of Management Review* 32 (4): 1145–1154.

Part VI

Longitudinal Qualitative Research

12

Bridging History and Reductionism: A Key Role for Longitudinal Qualitative Research

Robert A. Burgelman

Introduction

There is ample evidence that the legitimacy and usefulness of qualitative research are no longer questioned by most prominent scholars in administrative and organization science. Special issues of leading academic journals devoted to qualitative research and sustained support for professional development workshops on qualitative research methods at the Academy of Management meetings, for instance, support this encouraging observation. It also confirms the emergence over the last several decades of a canon for rigorous qualitative research (e.g., Burgelman 1985; Eisenhardt 1989; Glaser and Strauss 1967; Jick 1979; Langley 1999; Pettigrew 1979; Van de Ven 1992; Yin 1984).

While the increase in acceptance of qualitative research can be documented in administrative and organization science in general, this is perhaps less the case in the academy of international business. If so, the field

R. A. Burgelman (✉)
Stanford University, Stanford, CA, USA
e-mail: profrab@stanford.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_12

is missing out on the potentially important scholarly contributions that this type of research can make. Hence the initiative of the guest editors of this Special Issue to focus attention on qualitative research is both timely and inspiring.

This essay examines the bridging role that longitudinal qualitative research could possibly play between the historian's narrative approach and the social scientist's reductionist approach in theory development of complex social systems. To that end, I discuss the use of grounded theorizing in qualitative research, suggest how modern historical methods can strengthen the longitudinal dimension of qualitative research and focus it on complex systems, and propose that developing novel conceptual frameworks helps qualitative research play its potential bridging role in theory development. Throughout, I use examples from my own studies, as well as some from other scholars, to provide insight into the sorts of research questions that qualitative research could potentially address in the field of international business.

The Use of Grounded Theorizing in Qualitative Research

Qualitative researchers in administrative and organization science often use the discovery of "grounded theory" (Glaser and Strauss 1967) as part of their research methods.[1] A recent study of crossbusiness collaboration in multibusiness organizations (Martin and Eisenhardt 2010), for instance, is representative. Key aspects of this methodology are:

1. avoiding theoretical preconceptions;
2. constant comparison, joint coding and analysis, and theoretical sampling;
3. capitalizing on quantitative data; and
4. distinguishing substantive from formal grounded theory.

Avoiding Theoretical Preconceptions

While no scholar can nor should approach phenomena with a mind devoid of previous theoretical and empirical knowledge, Glaser and Strauss (1967: 37) nevertheless encourage researchers, "… at first, literally to ignore the literature of theory and fact on the area under study, in order to assure that the emergence of categories will not be contaminated by concepts more suited to different areas." This somewhat surprising recommendation at first seems to emphasize the negative; yet it is actually a positive affirmation of the importance of openness to new substantive (empirical) phenomena that gives the qualitative researcher confidence that new theory generation is possible. For instance, the dearth of literature about internal corporate venturing (ICV) – a relatively understudied substantive area in administrative and organization science – made it possible for me to follow Glaser and Strauss's guideline (Burgelman 1980).

Once the analytic core of the emergent grounded theory of a substantive area has taken sufficient shape, however, the researcher should seek to establish linkages – from the inside out, so to speak – with the existing literature to find further corroboration and/or meaningful differences, and thereby establish the clear and distinct contribution to knowledge generated by the new theory. These linking efforts will in the first instance advance the existing substantive grounded theory about the empirical phenomenon: for example, ICV (in my case) or cross-business collaboration (Martin and Eisenhardt 2010). They can also, however, contribute to broader conceptual areas through what Glaser and Strauss call "formal" grounded theorizing (see further below): for instance, evolutionary organization theory, organizational ecology and complexity theory (in my case), or information processing, transaction cost economics, social networking and complexity theories in the case of Martin and Eisenhardt (2010). Grounded theorizing thus leaves the door open for an eclectic approach, but as the researcher moves from developing substantive to formal grounded theory, being able to demonstrate clear linkages to existing intellectual traditions is reassuring, and affirms the researcher's scholarly identity in an established field.

Constant Comparison, Joint Coding and Analysis, and Theoretical Sampling

Several research activities form the core of grounded theorizing: identifying a social phenomenon and starting data collection; constantly comparing data about different instances of the phenomenon (both success and failure cases where appropriate); in the process of comparing, continuously coding the data to arrive at novel categories and their properties by way of writing brief memos that serve conceptualization; letting the emerging theoretical insights into these categories determine the search for additional instances of the phenomenon (theoretical sampling); and continuing the sampling process (again, both success and failure cases where appropriate) until additional instances no longer add further insight (saturation). Field notes and conceptual memos form the basis for generating the grounded theory about the phenomenon under investigation.

The ICV research, for instance, involved studying multiple cases of succeeding and failing internal corporate ventures in different stages of development within one very large diversified company. It traced the history of each of these ventures, and followed their further development in real time over a period of almost a year. Continuously comparing the data collected about each of these ventures informed ongoing iterative conceptualization efforts, provided insight into patterns of "success breeding success" and "failure breeding failure," and resulted in the creation of a new set of categories for the interlocking key activities (technical and need linking, product championing, strategic forcing, strategic building, organizational championing, delineating, retroactive rationalizing, selecting, structuring) of executives of different levels in the organization that became building blocks for constructing the process model of ICV (Burgelman 1983b).

Critically important for the qualitative researcher is to be alert to the fact that the phenomenon of interest and its preliminary delimitation often will have already emerged during the first interviews and data-coding activities (Glaser 1992: 21). For instance, in the case of research about Intel's exit from its core dynamic random access memory (DRAM)

business, the initial expectation was that with Gordon Moore and Andy Grove at the helm Intel would probably be the poster case of top management driven strategy-making. At some point during the first interview, however, Andy Grove surprisingly said that some middle-level technical managers had already made technology decisions that "limited the decision space of top management." When he agreed that it would be possible to interview these managers, it seemed clear that pursuing the DRAM exit could be a potentially interesting research project (Burgelman 1994).

It is also important to note that comparative case analysis can be performed at *multiple* levels of analysis. For instance, in researching the role of strategy-making in Intel's evolution, the comparative analysis involved both the organizational and the intra-organizational levels of analysis. At the organizational level, the analysis used the dimension of time to compare Intel's strategic Epoch I (the "memory company") and strategic Epoch II (the "microprocessor company") (Burgelman 2002a). At the intra-organizational level of analysis, the research compared Intel's failure in several new businesses in relation to the single-minded focus on its highly successful core microprocessor business, which produced insights into co-evolutionary lockin as a previously little-noticed source of strategic inertia (Burgelman 2002b).

Capitalizing on Quantitative Data

While qualitative research is naturally oriented to collecting rich qualitative data, it is important to keep in mind that *quantitative* data can also play an important role in grounded theorizing (Glaser 1992). For instance, as a panelist at the 1989 Strategic Management Society Conference, Andy Grove presented a set of graphs, one of which showed the percentage decline of memory products and the percentage increase of logic products in Intel's total revenues over time. It seemed immediately clear that Grove's graph was a manifestation of what was inchoately identified in the ICV study as a firm's "internal selection environment," and also provided additional data to further develop the concept of "intraorganizational ecology of strategy-making" (Burgelman 1991).

Distinguishing Substantive from Formal Grounded Theory

As noted earlier, Glaser and Strauss (1967: 33–34) make an important distinction between substantive and formal grounded theory, and consider both as "middle range": falling between minor working hypotheses and grand theories. They view *substantive* theory as "a strategic link in the formulation and generation of grounded formal theory. We believe that although formal theory can be generated directly from data, it is most desirable, and usually necessary, to start the formal theory from a substantive one" (1967: 79). Experience as a reviewer for academic journals indicates that qualitative researchers often seem to miss the importance of this distinction. Too often they attempt to formulate formal theory without first developing a substantive theory grounded in research of one particular substantive phenomenon. As a result, as Glaser and Strauss (1967: 81) warned, "When the theory is very abstract, it becomes hard to see how it came from the data of the study, since the formal theory now renders the data without a substantive theory intervening."

Attempts to relate substantive grounded theory of a particular phenomenon to substantive grounded theories of different but related phenomena provides the basis for generating *formal* grounded theory, which uses more general concepts to capture the more general phenomenon of which the particular ones are distinct manifestations. In my ICV-related research this involved two major steps. The first step was the creation of a *substantive* grounded theory of the ICV process. This involved resolving the anomaly that all the newly found categories of key activities associated with ICV could not be mapped onto Bower's (1970) "process model" of strategic capital investment, a substantive grounded theory (though not named as such). Resolving this anomaly required extending the received process model to encompass "strategic context determination." Strategic context determination was the part of the corporate strategy-making process that became activated by ICV project-level initiatives that were trying to change the existing corporate strategy going forward. This extension turned the process model into a general tool for studying substantive areas of strategic change and learning (Bower and Gilbert 2005; Mintzberg et al. 1998).

The second step was the realization that the ICV research findings also produced an anomaly in relation to Chandler's (1962) proposition that "structure follows strategy": it was found that the creation of a new venture division was, at least in part, a corporate-level structural response of the company having a number of new venture initiatives dispersed in different divisions *before* top management had articulated a deliberate corporate-level diversification strategy. This finding led to postulating the existence of *autonomous* strategic initiatives (not driven by the existing corporate strategy) in parallel with *induced* strategic initiatives (driven by the existing corporate strategy), and resulted in developing an evolutionary framework of the strategy-making process (Burgelman 1983a). This general framework contributed to the conceptual area of strategic management, and constitutes *formal* grounded theory in Glaser and Strauss's (1967) terms.

By way of conclusion, it is useful to keep in mind, as Glaser and Strauss point out, that: "Our strategy of comparative analysis for generating theory puts a high emphasis on *theory as a process*; that is theory as an ever-developing entity, not as a perfected product" (1967: 32, italics in original). Hence the methodology of grounded theorizing offers the opportunity for exerting disciplined creativity and enjoying the associated intellectual pleasure of discovery, but at the same time reminds the researcher that his/her theory generation effort is only a step along the road toward additional, cumulative knowledge development that will lead to future modification and reformulation.

The Use of Historical Methods in Qualitative Research

Grounded theorizing strongly emphasizes the comparative dimension of qualitative research, and, while sometimes involving a quasi-longitudinal approach (as in my ICV research), often relies mostly on cross-sectional comparative analysis of cases without much explicit concern for the longitudinal dimension.[2] Historical methods, on the other hand, are inherently concerned with longitudinal development, and involve recon-

structing the unfolding of individual and collective action patterns leading up to relatively unique events.

The study of history from ancient times to today encompasses a wide variety of perspectives and approaches, and it is not possible here to provide a synoptical overview of historical methodologies (e.g., Burrow 2007).[3] Nevertheless, just as Glaser and Strauss (1967) provided useful methodological guidance to help qualitative researchers develop more powerful grounded theoretical contributions, Gaddis's (2002) explication of modern historians' ways of approaching the study of the past, and Ferguson's (1998) examination of the role of counterfactual analysis and the link of historical analysis and theories of complexity and chaos, provide useful guidance for developing the *longitudinal* dimension of qualitative research, and for directing such research toward the study of *complex* social systems. In light of this, key aspects of modern historical methods are:

1. the adoption of an ecological view of reality;
2. a focus on the intersection of continuities and contingencies, and the role of context;
3. exploring links with complexity and chaos theories;
4. disciplined use of counterfactual analysis; and
5. particular generalization, general particularization, and overdetermination.

Ecological View of Reality

In contrast to a reductionist approach, which isolates dependent and independent variables, and establishes and estimates relationships between these, historical methods adopt an *ecological* approach, which examines how individual components interact to become systems whose nature cannot be completely understood from looking only at the sum of the parts.[4] In view of this approach, history scholars assume the *interdependency* of variables, and seek to trace their interconnections through time (Gaddis 2002: 53–55).

The ICV study, for instance, without consciously planning to do so, adopted an ecological approach by documenting how the *simultaneous as well as sequential* activities of actors situated at *multiple* levels in a complex organizational system – a new venture division in a large corporation – helped shape the dynamics of this system over time. Having decided, after the first interviews, to focus the research on ICV strategy-making, Bower's (1970) process model of strategic capital investment was used, as noted earlier, as a methodological tool for capturing multilevel simultaneity and sequentiality in ICV strategy-making (Burgelman 1983b). Subsequent truly longitudinal (1988–2001) qualitative research about the role of strategy-making in Intel's evolution (e.g., Burgelman 2002a) suggested that a large, complex organization such as Intel can be viewed as an ecological system within which induced and autonomous strategic initiatives compete for the organization's resources in patterned ways that are consistent with the variation-selection-retention paradigm of evolutionary organization theory (Aldrich 1979; Campbell 1960; Weick 1979). This research found that organizational adaptation depends in important ways on the extent to which the organization's internal selection environment governing the competition between induced and autonomous strategic initiatives reflects more or less accurately the pressures of the changing external selection environment. These findings suggested that the intraorganizational ecology of strategy-making could be fruitfully viewed as an additional level in a nested hierarchy of ecological systems (Aldrich 1999; Hannan and Freeman 1989).

Intersection of Continuities and Contingencies, and the Role of Context

In reconstructing the unfolding of individual and collective action patterns leading up to relatively unique events, historians are keen to identify "continuities" – that is, "patterns that extend beyond time" – and "contingencies" – that is, "phenomena that do not form a pattern" (Gaddis 2002: 30–31); they happen fortuitously, and are virtually impossible to predict.

The longitudinal research of the role of strategy-making in Intel's evolution, for instance, indicated that intersections of patterns of ongoing strategic activities (e.g., continuous development of technological capabilities) with fortuitous events (e.g., the microprocessor design win for the IBM PC without having to grant IBM exclusivity) played an important role in shaping the opportunities and threats that the firm faced throughout its evolution. This led to proposing that "strategic recognition," which refers to the capacity of senior and top management to see the strategic implications of a confluence of forces that is already happening before most others do, is a key complement to strategic planning in strategy-making as adaptive organizational capability.

Also important is the relationship between necessary and sufficient causes and the role of context. Gaddis states: "I would go so far as to define the word 'context' as the dependency of sufficient causes upon necessary causes; … For while context does not directly *cause* what happens, it can certainly determine consequences" (2002: 97, italics in original). For instance, research about "crossboundary disruption" – how a company from a different industry can enter successfully into another industry, and radically change its dynamics and equilibrium (Burgelman and Grove 2007b) – examined the role of industry context – that is, the structure and dynamics of the music industry in the late 1990s vs the structure and dynamics of the wireless communications industry in the mid-2000s – in determining the likely consequences of cross-boundary disruption efforts. This led to the prediction that Apple's chances to replicate its enormous success as a cross-boundary disruptor in the music industry might be more limited in the wireless communications industry (by late 2010, the success of Google's Android would seem to support this).

Links with Complexity and Chaos Theories

Ferguson (1998: 72) suggests a correspondence between history-as-science and science-as-history, because "… many modern developments in the natural sciences have been fundamentally concerned with changes over time." He also suggests that chaos theory, a modern development in the physical sciences, offers a useful middle position between idealism

and determinism, because "chaos (…) means unpredictable outcomes even when successive events are causally linked" (1998: 79). Similarly, Gaddis highlights correspondences between history, which focuses on complex systems with many interdependent variables, and ideas about complexity and chaos from the physical sciences, such as "path dependence" (how a small event at the beginning of a process makes a big difference at the end of it), "self-similarity across scale" (patterns remain the same, regardless of the scale at which they are observed), and "criticality" (the possibility of self-organization through an abrupt transition from one phase to another). He points out that historians detect criticality in retrospect, when they trace the rise and fall of empires, the beginnings and endings of wars, the diffusion of ideas and technologies, and so on. Forecasting criticality, however, is problematic (Gaddis 2002: 71–89, *passim*).

While a focus on complex (nonlinear) causation puts limits on the predictive capacity of longitudinal qualitative research, it does not have to give up on it completely. For instance, by identifying the necessary, though not sufficient, conditions for venture success within the corporate context, the ICV process model predicts that a particular venture project is likely to fail if one or more of the key strategic activities (which can be observed in real time) are not performed effectively. Also, as noted earlier in relation to the cross-boundary disruptor study, understanding "context" may help in predicting outcomes.

Changes in context may then serve as a criterion for establishing the appropriate *time horizon* for longitudinal qualitative research of complex systems. Changes in context are usually caused by changes of the "rules of the game" (normative, technological, economic, and cognitive rules, among others), reinforced over time through positive feedback (Burgelman and Grove 2007a). For backward-looking longitudinal studies a potentially useful guideline is to trace events back in time to where the context significantly changed, and to look for changes in rules that caused the change. Forward-looking longitudinal studies could use "critical event horizons" – the time, known in advance, that it will normally take for a particular event to be completed and/or manifest its impact (for instance, the length of time it takes to develop and bring to market a new microprocessor) – to estimate the minimum necessary length of real-time observation.

The concept of a "strategic inflection point" (period of crisis) signalled by "strategic dissonance" (Burgelman and Grove 1996) may be useful as an indicator of contextual change. Backward-looking longitudinal research revealed that there was a period of about 3 years (1981/82–1985) during which Intel was in a period of crisis – indicated by increasingly intense strategic dissonance among the top executives – because its core DRAM business had completely faltered as a result of changes in the industry context that had turned DRAM into a commodity business, and it was not clear what the company's new strategic direction for survival and profitable growth could be. It took until late 1985, when the success of the IBM PC became clearly established, for top management to be able to decide to put its destiny for the foreseeable future on that new strategic growth vector (Burgelman 1991, 1994).

Disciplined Use of Counterfactual Analysis

Ferguson posits a double rationale for counterfactual analysis by historians:

> Firstly, it is a *logical* necessity when asking questions about causation to pose "but for" questions, and to try to imagine what would have happened if our supposed cause had been absent. … Secondly, to do this is a *historical* necessity when attempting to understand how the past "actually was" … as we must attach equal importance to all the possibilities which contemporaries contemplated before the fact, and greater importance to these than to an outcome which they did not anticipate. (Ferguson 1998: 87, italics in original)

A limited dose of disciplined counterfactual analysis can be helpful in longitudinal qualitative research of complex systems.[5] Related to the study of Intel's exit from its core DRAM business, for instance, one Stanford colleague asked "What if Intel had tried to match the changing conditions in the DRAM industry?" Since there had been voices raised within Intel to do this, this "what if" question forced me to think why top management had chosen not to do so.[6] One insight gained from considering this counterfactual question was that the "maximize margin-

per-wafer-start" rule Intel used to allocate scarce manufacturing capacity, which consistently selected against the commoditizing DRAM products, actually reflected something deep about the company's culture. It produced the insight that Intel as a leading-edge high-technology company pursued a differentiation strategy, and expected to obtain premium prices for its products. Consequently, internal demands for competing more effectively in a commodity business were inconsistent with this culture, and especially so in light of Intel having available a major new business opportunity with its microprocessor products.[7] Changing the resource allocation rule would quite likely have caused confusion about corporate identity in the organization (Burgelman 1994).

Particular Generalization, General Particularization, and Overdetermination

Rejecting the notion that historians do not wish to use theory, Gaddis points out that

> We do, however, normally *embed our generalizations within our narratives.* In seeking to show how past processes have produced present structures, we draw upon whatever theories we can find that will help us accomplish that task. (2002: 62, italics in original)

Hence "particular generalization" is what historians do; they "generalize for particular purposes." On the other hand, "general particularization" is what most social scientists typically do: "… *embedding narratives within generalizations."* Social scientists' "… principal objective is to confirm or refute a hypothesis, and they subordinate narration to that task" (2002: 63). In other words, historians make sense out of extensive narratives by giving them some limited general applicability (they generalize the particular); and social scientists use limited narratives to illustrate presumably general theories (they particularize the general).

An example of particular generalization is Chandler's (1962) earlier-mentioned proposition that "structure follows strategy," which was embedded in his historical narratives about DuPont, Sears, General

Motors, and Standard Oil of New Jersey. Even though further research by other scholars modified and reformulated this generalization, it proved seminal in stimulating the development of the strategic management field. A modest example of general particularization, on the other hand, is the development of a theoretical framework for deriving different forms of "strategic integration" involving collaboration between individual businesses in the multibusiness corporation (Burgelman and Doz 2001). "Exampling" – seeking out supporting cases (Glaser and Strauss 1967: 5) – is used to illustrate the theoretical forms of strategic integration:

1. minimal;
2. reach-driven;
3. scope-driven;
4. complex; and
5. over-ambitious.

Such general particularization runs the risk, as noted earlier, that it becomes hard to see how it derived from the data. On the other hand, it can be potentially useful to assess the findings of substantive grounded theorizing. For instance, the framework of strategic integration seems useful to further assess the proposition that bottom-up driven cross-business collaborations are more likely to be successful than top-down driven ones (Martin and Eisenhardt 2010).[8]

Another important difference between the historian's approach to theorizing and that of social science is, as Gaddis points out, the former's preference for "parsimony in consequences, but not causes" (i.e., multiple identified causes must converge upon a particular consequence). In other words, historians accept *overdetermination* of singular events, whereas social scientists consider over-determined events inadequately explained (Gaddis 2002: 105). Interestingly, Allison's analysis of the "Cuban Missile Crisis," one of the most celebrated qualitative research studies (Allison and Zelikow 1999), manifests such overdetermination: that is, the same event is explained by three completely different theories, each of which nevertheless is able to highlight clear and distinct insights into the origin, unfolding, and resolution of the crisis. This suggests that the challenge for

qualitative researchers is to try to shed meaningful light on events in spite of the overdetermination resulting from the large number of interdependent variables in play in the complex social systems under study: for instance, in the case of the Cuban missile crisis, the various branches of the US government, the various branches of the US armed forces, the various branches of the USSR government, the Cuban revolutionary government, the different routines of each of these organizations, their asynchronic actions, and so on.

By way of conclusion, qualitative researchers who are interested primarily in developing narratives (or case studies) can improve the quality of their output, and respond more effectively to the sorts of criticisms that have traditionally been directed at their work – for instance the issue of overdetermination – by adopting the methods elucidated by modern historians such as Gaddis and Ferguson. At the same time, they need to keep in mind that their narratives and particular generalizations are also subject to Glaser and Strauss's (1967) earlier mentioned *caveat* about grounded theory development: that is, their efforts are only a step along the road toward additional, cumulative knowledge development that will lead to future modification and reformulation.[9]

Most longitudinal qualitative researchers, however, as social scientists, will probably want to go beyond writing narratives and particular generalization. While taking into account historical methods to strengthen the longitudinal dimension of their research, they can use substantive and formal grounded theorizing along the lines summarized earlier in this essay. Substantive grounded theorizing uses fairly rudimentary concepts and categories that can be clearly linked to the data to construct a rudimentary parsimonious conceptualization, which sets it apart from historical narrative, because it is suggestive of the substantive phenomenon, independent of the complete narratives from which it was derived (e.g., the ICV process model can be meaningfully discussed without reference to any of the particular cases from which it was inducted).

On the other hand, formal grounded theorizing can also seek to establish links – again from the inside out – with reductionist empirical social science. For instance, as noted earlier, the formal grounded theory of the intraorganizational ecology of strategy-making could be linked to the literature on organizational ecology. And finding parallels between the

intraorganizational ecology of strategy-making and high theory of complex adaptive systems (e.g., Prigogine 1980, 1996) and adaptation at the edge of chaos (Gould 2002; Kauffman 1993) bolstered confidence in its potential adaptive role in organizational evolution (Burgelman and Grove 2007a; Burgelman forthcoming). Similarly, the longitudinal qualitative research of Intel's evolution, with the help of grounded theorizing, discovered and conceptualized the phenomenon of coevolutionary lock-in and was able to relate it to existing theory about path dependency (Burgelman 2010).

The Bridging Role of Longitudinal Qualitative Research in Theory Development: Generating Conceptual Frameworks

The preceding discussion suggests that longitudinal qualitative research using the methodology of grounded theorizing can be usefully situated on a spectrum of theory development between the historian's particular generalization and the reductionist's general particularization. Note that this spectrum does not imply a merit ordering: that is, theory development within each type along the spectrum can be of higher or lower quality, as measured, for instance, by its originality, scientific utility and practical utility (Corley and Gioia 2011).[10] The theory development spectrum is shown in Fig. 12.1.

At the history end of the spectrum of theory development, particular generalization involves carrying out case studies of *particular, concrete,*

HISTORY: PARTICULAR GENERALIZATION	↔	LONGITUDINAL QUALITATIVE RESEARCH	↔	REDUCTIONISM: GENERAL PARTICULARIZATION
Particular Concrete Experiential		Specific Substantive Suggestive		General Abstract Non-experiential
Narratives		Conceptual frameworks (box-and-arrow charts)		Statistical and mathematical models

Fig. 12.1 The bridging role of longitudinal qualitative research in theory development

and *experiential* social phenomena characterized by complex, nonlinear causation. Mostly natural language is used to construct a coherent and complete representative and explanatory *narrative.* This type of theory development should meet the standards of rigor associated with advanced historical methods. In particular, researchers should be concerned about replicability as indicated by the degree of consensus among scholars, at a given moment in time, about the reality fit of their narratives (Gaddis 2002: 107).

At the reductionist end of the spectrum of theory development, general particularization involves statistically based models (e.g., the application of the Lotka–Volterra model of competitive interactions in population growth used in organizational ecology) or axiom-based mathematical models (e.g., the Nash equilibrium model used in microeconomics). While general particularization involves using case examples as illustrations, the value of this type of theory as perceived by like- minded scholars does not significantly depend on this. These types of theories are *general*, *abstract*, and *non-experiential.*

In between the historian and the reductionist types of theory development, longitudinal qualitative research of complex social systems, with the help of substantive and formal grounded theorizing, seeks to go beyond particular generalizations by creating *conceptual frameworks*: boxes-and-arrow charts that show how the complex system hangs together, and its operative logic. Initial conceptual frameworks (substantive grounded theorizing) employ rudimentary categories and concepts that are still closely linked to the data provided in the narratives (e.g., the ICV process model). Further developed conceptual frameworks (formal grounded theorizing) encompass more general categories and concepts that link to a broader conceptual area (e.g., the intraorganizational ecology of strategy-making). Such conceptual frameworks are *specific* (representative of a class of phenomena), *substantive* (capturing the essential/material part underlying the phenomenon), and *suggestive* (evoking the phenomenon indirectly). They provide deeper and more general insight into phenomena than is possible with the natural language of narratives.

The logic underlying the spectrum of theory development presented in Fig. 12.1 indicates the potentially important bridging role that longitudinal qualitative research can play between the historian's particular gener-

alizations and the reductionist's general particularizations. Hence a test for assessing the value of such research may be, on the one hand, the extent to which its conceptual frameworks help better understand key aspects of the phenomena captured in narratives[11] and, on the other hand, the extent to which they also provide useful stepping stones toward the development of statistical and mathematical models.[12]

Implications and Conclusion

As is well known in the international business field, there are significant challenges associated with doing cross-national and cross-cultural research. A behavioral concept such as "political participation," for instance, has different ways of expressing itself in different nations (e.g., Przeworski and Teune 1970). Cognitive and perceptual processes are significantly different in different cultures: Westerners, for instance, are likely to overlook the influence of context on the behavior of objects and people, whereas Chinese think that one cannot understand the part without understanding the whole (Nisbett 2003).

Interestingly, these sorts of challenges would seem to create great opportunities for the potentially important theory-bridging role that longitudinal qualitative research as discussed in this essay can play in international business studies. This can be illustrated, for instance, by comparative studies of the strategy-making processes of Japanese and US companies (Burgelman 1988; Kagono et al. 1985). Based on this comparative qualitative research, I suggested that Japanese companies may be more ready to accept new ideas from lower-level executives as equally valuable as new ideas of higher-level executives because everybody knows and accepts their position in the social order, which made it seem likely that Japanese companies would stimulate less opportunistic strategic behavior and be more able to deal with variety. At the same time, however, the range of that variety also seemed more limited, because it could not threaten the existing social order. New strategic initiatives outside that range would likely be spun off, because the idea of strategy-making as an opportunity structure for personal advancement might be unacceptable in Japanese companies. This sort of insight generated by qualitative

research could potentially inspire further research, for instance sociological research about the functioning of status hierarchies in companies located in different cultures.

A more recent example (Immelt et al. 2009) concerns innovation driven by autonomous strategic actions of middle-level executives within GE, who discovered new needs in developing countries and were able to develop whole categories of new products tailored to these needs (e.g., a $1000 handheld electrocardiogram device and a $15,000 portable, PC-based ultrasound machine) and subsequently find applications for these products in the developed markets ("reverse innovation"). Carefully constructing narratives of several of these globally based autonomous strategic initiatives and the processes through which they were able to sustain or failed to sustain internal support within the complex global GE organization could generate new substantive and formal grounded theory about the functioning of the strategy-making process related to innovation in the rapidly changing global corporate context.

In conclusion, this essay suggests that qualitative researchers in international business, like their colleagues in other areas of administrative and organization science, may find it helpful to adopt the methodology of grounded theorizing and to be mindful of the guidance that the methods of modern historical research can offer to strengthen the longitudinal dimension of their studies. To capitalize fully on the bridging role in theory development that qualitative research can play between history and reductionism, international business scholars may want to seek out areas of study that require an ecological view of reality and are characterized by complexity and nonlinear causation. As they adopt a longitudinal dimension to their research, they will more readily be able to document the intersection of continuities and contingencies, and the effects of context on outcomes. Generating novel and imaginative conceptual frameworks would quite likely help stimulate other international business scholars' research and theory development in the reductionist mode.

Acknowledgements Thanks to Special Issue co-editor Mary Yoko Brannen and three anonymous reviewers for helpful suggestions to clarify the paper's main arguments, and to improve its readability.

Notes

1. Glaser and Strauss's original treatise (1967) remains the best source for researchers to familiarize themselves with the methodology of grounded theory development. Later elaborations (e.g., Corbin and Strauss 2008; Glaser 1992; Strauss 1987) provide additional details and procedures, but also show the emergence of a split between Glaserian and Straussian interpretations of the methodology. Glaser accuses Strauss of introducing a degree of "conceptual forcing" on data in lieu of the "emergence of concepts" from data, and restricting grounded theorizing too much to qualitative data at the expense of what can be contributed by quantitative data. Close reading of both authors' later works suggests that the substance of the split is somewhat overblown. The more authentic interpretation, nevertheless, in my view is that of Glaser (1992).
2. "Time," "longitudinal," "history" are not indexed in Glaser and Strauss (1967). Strauss (1987) indexes "trajectory" as a time-dependent notion, but does not explicitly discuss the longitudinal aspect of qualitative analysis. Similarly, Corbin and Strauss (2008) do not explicitly discuss the longitudinal aspects of qualitative research.
3. Burrow (2007: xiii, italics in original) points out that "A *histor* in Homer was someone who passed judgment based on the facts as a result of investigation, so the link between history and inquest is a very old one."
4. One of the reviewers suggested "holistic" as perhaps a better term than "ecological." While both terms can be used effectively to indicate a system with emergent properties, I use "ecological" because it refers explicitly to the relationship between systems and their environments.
5. For a thorough discussion of the uses of counterfactual analysis – both counterfactual history and causal modeling – in strategic management, see Durand and Vaara (2009).
6. A potentially interesting implication of this is that counterfactual analysis plays not only an important upstream role of theory generation in what Reichenbach (1951) calls the "context of discovery," but also a downstream role of theory falsification in the "context of justification" by offering tests of consistency and coherence. I thank one of the reviewers for suggesting this connection to scientific philosophy.
7. Another interesting counterfactual, which I did not pursue, would be to think about what the company could have done if these new opportunities had not been available.

8. For instance, the strategic integration framework (Burgelman and Doz 2001) raises the possibility that bottom-up driven cross-business collaboration was more likely to be successful than top-down driven cross-business collaboration in multibusiness firms (Martin and Eisenhardt 2010), because the latter might have been of the "overambitious" type of strategic integration. Informed by this possibility, potentially interesting further research might try to establish whether, in general, top-down driven cross-boundary collaborations are more likely to be overambitious than bottom-up driven cross-boundary collaboration (a potential alternative explanation) and/or under what conditions this might be the case.
9. A recent example is the major revision of the strategy-making of Czar Alexander and the Russian top military command during Napoleon's disastrous Russian campaign, based on newly available archival data from Russia. See Lieven (2009).
10. It would perhaps have been preferable for Corley and Gioia (2011) to develop their framework further by considering scientific utility and practical utility as two different dimensions (they too quickly pass on doing this, in my view).
11. For instance, Chandler, in an interview in 2002 related to the celebration of the 40th anniversary of the publication of *Strategy and Structure*, mentioned my book *Strategy is Destiny* as an example of research that continued the tradition that he had started. See Rodrigues (2002).
12. The model of induced and autonomous strategic behavior, for instance, seems to have served as one of the stepping stones for some scholars' development of a mathematical model in the economics of the firm (Rotemberg and Saloner 2000).

References

Aldrich, H. 1979. *Organizations and environments*. Englewood Cliffs: Prentice Hall.

———. 1999. *Organizations evolving*. London: Sage.

Allison, G., and P. Zelikow. 1999. *Essence of decision: The Cuban missile crisis*. 2nd ed. New York: Addison Wesley Longman.

Bower, J.L. 1970. *Managing the resource allocation process*. Boston: Harvard Business School Press.

Bower, J.L., and C.G. Gilbert. 2005. *From resource allocation to strategy*. Oxford: Oxford University Press.

Burgelman, R.A. 1980. Managing innovating systems: A study of the process of internal corporate venturing. Unpublished PhD dissertation, Graduate School of Business, Columbia University, New York.

———. 1983a. A model of the interaction of strategic behavior, corporate context and the concept of strategy. *Academy of Management Review* 8 (1): 61–70.

———. 1983b. A process model of internal corporate venturing in the diversified major firm. *Administrative Science Quarterly* 28 (2): 223–244.

———. 1985. Applying the methodology of grounded theorizing in strategic management: Recent findings and their implications. In *Advances in strategic management*, ed. R. Lamb and P. Shrivastava, vol. 3, 83–99. Greenwich: JAI Press.

———. 1988. A comparative evolutionary perspective on strategy making: Advantages and limitations of the Japanese approach. In *Innovation and management: International comparisons*, ed. K. Urabe, J. Child, and T. Kagono, 63–80. Berlin: De Gruyter.

———. 1991. Intraorganizational ecology of strategy making and organizational adaptation: Theory and field research. *Organization Science* 2 (3): 239–262.

———. 1994. Fading memories: A process theory of strategic business exit in dynamic environments. *Administrative Science Quarterly* 39 (1): 24–56.

———. 2002a. *Strategy is destiny: How strategy-making shapes a company's future*. New York: Free Press.

———. 2002b. Strategy as vector and the inertia of coevolutionary lock-in. *Administrative Science Quarterly* 47 (2): 325–357.

———. 2010. Strategic consequences of coevolutionary lock-in: Insights from a longitudinal process study. In *The hidden dynamics of path dependence: Institutions and organizations*, ed. G. Schreyoegg and J. Sydow, 233–248. Basingstoke: Palgrave Macmillan.

———. forthcoming. How Prigogine's ideas of "self-organization" and "becoming" inform the role of strategy-making in organizational evolution. In *Les grands inspirateurs de la théorie des organizations*, ed. O. Germain. Paris: Editions Management et Société.

Burgelman, R.A., and Y.L. Doz. 2001. The power of strategic integration. *Sloan Management Review* 42 (3): 28–38.

Burgelman, R.A., and A.S. Grove. 1996. Strategic dissonance. *California Management Review* 38 (2): 8–28.

———. 2007a. Let chaos reign, then rein in chaos – Repeatedly: Managing strategic dynamics for corporate longevity. *Strategic Management Journal* 28 (10): 965–979.

———. 2007b. Cross-boundary disruptors: Powerful interindustry entrepreneurial change agents. *Strategic Entrepreneurship Journal* 1 (3–4): 315–327.

Burrow, J. 2007. *A history of histories*. New York: Knopf.

Campbell, D.T. 1960. Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review* 67 (6): 380–400.

Chandler, A.D. 1962. *Strategy and structure*. Cambridge, MA: MIT Press.

Corbin, J., and A. Strauss. 2008. *Basics of qualitative research*. Los Angeles: Sage.

Corley, K.G., and D.A. Gioia. 2011. Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review* 36 (1): 12–32.

Durand, R., and E. Vaara. 2009. Causation, counterfactuals, and competitive advantage. *Strategic Management Journal* 30 (12): 1245–1264.

Eisenhardt, K.M. 1989. Building theories from case study research. *Academy of Management Review* 14 (4): 532–550.

Ferguson, N. 1998. *Virtual history*. New York: Basic Books.

Gaddis, J.L. 2002. *The landscape of history: How historians map the past*. New York: Oxford University Press.

Glaser, B.G. 1992. *Basics of grounded theory analysis*. Mill Valley: Sociology Press.

Glaser, B.G., and A.L. Strauss. 1967. *The discovery of grounded theory*. Chicago: Aldine.

Gould, S.J. 2002. *The structure of evolutionary theory*. Cambridge, MA: Harvard University Press.

Hannan, M.T., and J.H. Freeman. 1989. *Organizational ecology*. Cambridge, MA: Harvard University Press.

Immelt, J.R., V. Govindarajan, and C. Trimble. 2009. How GE is disrupting itself. *Harvard Business Review* 87 (10): 56–65.

Jick, T.D. 1979. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly* 24 (4): 602–611.

Kagono, T., I. Nonaka, K. Sakakibara, and A. Okumura. 1985. *Strategic vs evolutionary management: A US-Japan comparison of strategy and organization*. Amsterdam: North Holland.

Kauffman, S.A. 1993. *The origins of order: Self-organization and selection in evolution*. Oxford: Oxford University Press.

Langley, A. 1999. Strategies for theorizing from process data. *Academy of Management Review* 24 (4): 691–710.

Lieven, D. 2009. *Russia against Napoleon: The true story of the campaigns of war and peace*. New York: Viking.

Martin, J., and K.M. Eisenhardt. 2010. Rewiring: Cross-business unit collaboration in multibusiness organizations. *Academy of Management Journal* 53 (2): 265–301.

Mintzberg, H., B. Ahlstrand, and J. Lampel. 1998. *Strategy safari*. New York: Free Press.

Nisbett, R.E. 2003. *The geography of thought: How Asian and Westerners think differently… and why*. New York: Free Press.

Pettigrew, A. 1979. On studying organizational cultures. *Administrative Science Quarterly* 24 (4): 570–581.

Prigogine, I. 1980. *From being to becoming: Time and complexity in the physical sciences*. New York: WH Freeman.

———. 1996. *The end of certainty: Time, chaos, and the new laws of nature*. New York: Free Press.

Przeworski, A., and H. Teune. 1970. *The logic of comparative social inquiry*. New York: Wiley.

Reichenbach, H. 1951. *The rise of scientific philosophy*. Berkeley: University of California Press.

Rodrigues, J.N. 2002. An interview with Alfred Chandler. *Business Strategy Review* 13: 21.

Rotemberg, J.J., and G.S. Saloner. 2000. Visionaries, managers, and strategic direction. *RAND Journal of Economics* 31 (4): 693–716.

Strauss, A. 1987. *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

Van de Ven, A.H. 1992. Suggestions for studying strategy process: A research note. *Strategic Management Journal* 13 (Summer Special Issue): 169–188.

Weick, K.E. 1979. *The social psychology of organizing*. Reading: Addison-Wesley.

Yin, R.K. 1984. *Case study research*, Applied social research methods series. Vol. 5. Beverly Hills: Sage.

13

Bridging History and Reductionism: A Commentary

Catherine Welch

Burgelman's (2011) *JIBS* Perspectives article invites international business (IB) researchers to consider conducting longitudinal qualitative research that is informed by historical methods. He recognizes that doing such research entails much more than collecting data spanning multiple time periods. In addition, longitudinal qualitative research requires a fundamental rethinking of traditional assumptions about what social theories should look like.

In this commentary, I will draw out what I regard as being the implications of Burgelman's vision, which I shall argue are perhaps more far reaching than even he acknowledges. In doing so, I extend Burgelman's insights by relating them to more recent discussions on process theorizing. I will conclude that although few in the IB field have responded to

Burgelman, R. A. 2011. Bridging history and reductionism: A key role for longitudinal qualitative research. *Journal of International Business Studies* 42 (5): 591–601.

C. Welch (✉)
The University of Sydney Business School, The University of Sydney, Sydney, NSW, Australia
e-mail: catherine.welch@sydney.edu.au

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_13

Burgelman's call to date, recent methodological developments in process research provide a solid foundation for researchers to do so in the future.

Burgelman draws on his own considerable experience to explain the nature and strengths of longitudinal qualitative research. His sustained commitment to conducting this kind of research over many years is highly distinctive. Few researchers have been able to follow his example of building up such a detailed account of strategic change in a single organization (Intel) over an extended period in order to gain novel theoretical insights from immersion in the chosen research site. He has been able to use his intimate knowledge of Intel—derived from many years of real-time fieldwork, retrospective data, and the active participation of the organization's CEO (Andy Grove)—to pioneer an evolutionary approach to corporate strategy (e.g., Burgelman 1983).

Burgelman's research has also contributed to the development of what has become known as the 'process school' of strategy, which can be traced back to Joseph Bower at the Harvard Business School. The Harvard strategy process school paved the way for what is now a flourishing research agenda in strategy and management (e.g., Langley et al. 2013). The school has also had a strong influence on international business through the work of two of Bower's students, Doz and Prahalad (e.g., 1993), who have sought to develop a process approach to the study of the multinational enterprise.

In seeking to understand Burgelman's argument, the main task for the reader is to unpack what he means by 'longitudinal qualitative research.' For him, conducting qualitative research goes beyond collecting data that cover multiple time periods; such an endeavour is necessary but not sufficient. Instead, the research agenda Burgelman advocates requires us to rethink the research questions we pose, the analytical techniques we use, the nature of the theories we develop, and the way we view our role as social theorists. In sum, Burgelman advocates an alternative way of conducting social science to the standard types of case studies found in international business journals. (For a discussion of the latter, see Piekkari et al. 2009.)

In Table 13.1, I summarize the contrast between this 'standard' approach to qualitative research and the alternative offered by longitudinal qualitative research. I now consider the differences in more detail,

Table 13.1 Standard and longitudinal qualitative research compared

Standard qualitative research	Longitudinal qualitative research
Research questions seek to identify antecedents of particular outcomes	Research questions ask how outcomes are produced, i.e., mechanisms linking causes and effects
Cross-sectional research design	Backward-looking (retrospective) and forward-looking (real-time) research
Thematic analysis	Event-based not just thematic analysis, drawing on historical methods, e.g., periodization, narrative analysis, counterfactual analysis, rival explanations
Proposition-based style of theorizing	Process-based style of theorizing
Reductionist paradigm	Ecological paradigm
Maintaining division between social sciences and history	Historicizing social sciences

referring to relevant insights from recent developments in process and historical methodology.

Posing research questions that keep change and evolution at the forefront is an important element of a longitudinal qualitative study. Burgelman's own empirical work is an example of the types of questions that require a longitudinal research design. Some examples are: How do large firms engage in corporate entrepreneurship (Burgelman 1983)? How do firms develop strategies that enhance their likelihood of surviving in a dynamic environment (Burgelman 1991)? How do firms strategically exit from existing businesses (Burgelman 1994)? Such questions not only require the researcher to identify the antecedents of outcomes; they also require tracing the processes that produce these outcomes (Table 13.1). Undertaking a longitudinal study means tracing the causal mechanisms that lead to outcomes of interest.

Understanding the causal chain of events also requires going beyond cross-sectional research designs. Burgelman (2011) makes it clear that longitudinal research encompasses both retrospective (backward-looking) and real-time (forward-looking) data (Table 13.1). While reviewers are often quick to point out the limitations of retrospective data, such data also have considerable strengths that should not be overlooked. For

example, interviewees may be prepared to be less constrained in discussing past events than is possible when events are unfolding, given pressures to maintain a positive image of the organization.

The weaknesses of retrospective data can, moreover, be ameliorated by a well-constructed research design. Historians have well-established approaches for critically evaluating and triangulating retrospective data (Kipping et al. 2013); these approaches can equally be employed by IB scholars in constructing their own research designs (Buckley 2016). Retrospective data have an essential role to play in IB, given that many of the processes we study (such as institutional change or the internationalization of the firm) unfold over years or even decades, which makes their real-time study impractical. Without the inclusion of retrospective data, IB scholars may not be capturing the phenomenon they claim to be studying, leading to poor construct validity and erroneous conclusions.

Longitudinal qualitative research also requires alternative techniques for analyzing data (Table 13.1). This is one of the aspects of process research that has received greater attention from management scholars in recent years. While Burgelman (2011) advocates standard grounded theory techniques, which include coding data into themes, he also recognizes the need to supplement these coding techniques with methods that are more widely used by historians. Historical research provides guidance on how to analyze temporal interconnections, and not just thematic patterns in datasets. The many analytical techniques for studying processes have gained increasing recognition from management researchers conducting process research (see e.g., Langley 1999). Examples of these techniques include: sequence analysis, narrative analysis, periodization, comparative analysis (i.e., comparing across different time periods but also comparing rival explanations), using historical episodes as natural experiments, and—as Burgelman (2011) discusses in some detail—conducting thought experiments such as counterfactual analysis (i.e., asking 'what if?'). The value of these techniques has also been acknowledged by IB scholars (e.g, Buckley 2016; Jones and Khanna 2006), even though the techniques are still rarely used in empirical IB research.

The theoretical outcome of a longitudinal qualitative study is potentially very different from the standard form of theorizing familiar to IB researchers. Delineating these contrasting 'styles' of theorizing is a

development in management research (Delbridge and Fiss 2013), which took place subsequent to the publication of Burgelman's paper. Cornelissen (2017) labels the standard approach to theorizing as proposition-based theorizing, that is, formal statements setting out the relationships between constructs. Longitudinal research, on the other hand, lends itself to producing what Cornelissen (2017) terms a process or narrative-based style of theorizing (Table 13.1), one that proposes in abstract terms the causal chain of events—the process—by which an outcome is produced.

A narrative or process-based approach to theorizing is still rarely found in qualitative IB research, which remains dominated by the propositional style. This dominance should not come as a surprise, given that proposition-based theorizing underlies the dominant methodological authority for qualitative IB researchers: the case study roadmap proposed by Kathleen Eisenhardt (1989). Recognition that the propositional style is not the only legitimate form of theorizing should provide more openings for IB researchers to make a theoretical contribution by conducting longitudinal studies.

Above all, taking on the challenge of longitudinal research entails a paradigm shift: from reductionist to ecological forms of explanation (Table 13.1). Burgelman's (2011) discussion of these two forms of explanation has been sourced from John Lewis Gaddis (2002), a prominent historian of the Cold War who has been able to bridge the divide between history and political science—a divide which also persists in management, despite many calls to overcome it. (For a recent discussion of the relationship between history and management, see Greenwood and Bernardi 2014.)

The differences between reductionist and ecological approaches can be regarded as paradigmatic in that they can be traced to contrasting assumptions about the world and how to obtain knowledge of it. The reductionist paradigm is based on simple causation, linear relationships, the quest for universalities, parsimonious explanations, and the possibility of prediction. Accordingly, independent and dependent variables can be isolated and the relationships between them measured and tested. Reductionist theorizing still dominates IB research, as well as many other social sciences. From the viewpoint of the reductionist paradigm, history falls

short of producing the robust theories required by the social sciences, thus perpetuating the division between the two disciplines.

The ecological approach to scientific explanation is found not just in history but also in evolutionary sciences such as palaeontology and evolutionary biology (Gaddis 2002). This paradigm takes a systemic view that regards variables as interdependent, causality as complex and context dependent, and trajectories as non-linear, even chaotic. Indeterminacy rather than predictability is the property of such a system. Historical methods are well suited to developing explanations of the social world of this kind. Accordingly, an ecological approach breaks down the traditional divide between history and the social sciences in a way that historicizes social science.

Given the paradigmatic shift it entails, it is perhaps not surprising that Burgelman's (2011) call for a different kind of qualitative research in IB is yet to be realized—although some modest advances have been made (Perchard et al. 2017). Burgelman himself is careful to emphasize the potential for the two paradigms to be mutually supportive. He positions longitudinal qualitative research as forming a 'bridge' between historical and reductionist research, and a 'stepping stone' to formal mathematical models.

My reading of Burgelman (2011) has characterized his approach somewhat differently, emphasizing the contrast it represents to standard approaches to conducting qualitative research. Rather than forming a bridge to reductionist approaches, he outlines an alternative to them. In highlighting the differences between the two approaches, I am not suggesting the abandonment of reductionist explanations, predictive theories, linear causality, and a propositional style of theorizing. Rather, I am arguing that by building on the pioneering work of scholars such as Burgelman, as well as recent insights into process research, IB researchers have the opportunity to diversify and enrich the methods we use and the theories we develop. In a world that is rapidly evolving, expanding our theoretical arsenal so that we can better account for process and change is a necessary step for the field to take.

References

Buckley, P.J. 2016. Historical research approaches to the analysis of internationalisation. *Management International Review* 56 (6): 879–900.

Burgelman, R.A. 1983. Corporate entrepreneurship and strategic management: Insights from a process study. *Management Science* 29 (12): 1349–1364.

———. 1991. Intraorganizational ecology of strategy making and organizational adaptation: Theory and field research. *Organization Science* 2 (3): 239–262.

———. 1994. Fading memories: A process theory of strategic business exit in dynamic environments. *Administrative Science Quarterly* 39 (1): 24–56.

———. 2011. Bridging history and reductionism: A key role for longitudinal qualitative research. *Journal of International Business Studies* 42 (5): 591–601.

Cornelissen, J.P. 2017. Editor's comments: Developing propositions, a process model, or a typology? Addressing the challenges of writing theory without a boilerplate. *Academy of Management Review* 42 (1): 1–9.

Delbridge, R., and P.C. Fiss. 2013. Editors' comments: Styles of theorizing and the social organization of knowledge. *Academy of Management Review* 38: 325–331.

Doz, Y.L., and C.K. Prahalad. 1993. Managing DMNCs: A search for a new paradigm. In *Organization theory and the multinational enterprise*, ed. S. Ghoshal and D.E. Westney, 24–50. Houndsmill, Basingstoke, Hampshire: Macmillan.

Eisenhardt, K.M. 1989. Building theories from case study research. *Academy of Management Review* 14 (4): 532–550.

Gaddis, J.L. 2002. *The landscape of history: How historians map the past*. Oxford: Oxford University Press.

Greenwood, A., and A. Bernardi. 2014. Understanding the rift, the (still) uneasy bedfellows of History and Organization Studies. *Organization* 21 (6): 907–932.

Jones, G., and T. Khanna. 2006. Bringing history (back) into international business. *Journal of International Business Studies* 37 (4): 453–468.

Kipping, M., R.D. Wadhwani, and M. Bucheli. 2013. Analyzing and interpreting historical sources: A basic methodology. In *Organizations in time: History, theory, methods*, ed. M. Bucheli and R.D. Wadhwani, 305–329. Oxford: Oxford University Press.

Langley, A. 1999. Strategies for theorizing from process data. *Academy of Management Review* 24 (4): 691–710.

Langley, A., C. Smallman, H. Tsoukas, and A. Van de Ven. 2013. Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal* 56 (1): 1–13.

Perchard, A., N.G. MacKenzie, S. Decker, and G. Favero. 2017. Clio in the business school: Historical approaches in strategy, international business and entrepreneurship. *Business History* 59 (6): 904–927.

Piekkari, R., C. Welch, and E. Paavilainen. 2009. The case study as disciplinary convention: Evidence from international business journals. *Organizational Research Methods* 12 (3): 567–589.

Part VII

Fuzzy-Set Qualitative Methods

14

Predicting Stakeholder Orientation in the Multinational Enterprise: A Mid-Range Theory

Donal Crilly

Introduction

The increasing reach of corporations across borders intensifies the debate about the function of the corporation in society (Walsh et al. 2006). In a largely unregulated global economic space, multinational enterprises (MNEs) potentially challenge the sovereignty of nation-states (Vernon 1998). The United Nations' Global Compact calls on corporations to use this influence to pursue both social and economic objectives (UNDP 2009). Yet corporate involvement in social issues is contentious. On one hand, proponents of shareholder primacy argue that managers maximize societal welfare by attending to shareholder value (Jensen 2002) and should concern themselves with social issues only where there are pressing economic or legal reasons. On the other hand, the function of the corporation might extend beyond wealth creation to encompass a social role premised on redressing social ills (Margolis and Walsh 2003).

D. Crilly (✉)
London Business School, London, UK
e-mail: dcrilly@london.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_14

The theories advanced to explain stakeholder orientation – that is, whether business leaders define the function of the corporation narrowly around maximizing shareholder value, or broadly around creating benefits for a wider range of stakeholders – emphasize adaptation to local environments. Resource dependency theory, concerned with the *external* control of organizations (Pfeffer and Salancik 1978), predicts that firms attend to the demands of stakeholders who control important inputs (Mitchell et al. 1997). Institutional theory explains interactions with stakeholders on the basis of legal, normative and taken-for-granted standards of conduct. For example, national systems influence stakeholder interactions (Ahmadjian and Robbins 2005; Fiss and Zajac 2004), perceived responsibilities (Williams and Aguilera 2008), and beliefs about the roles of economic actors (Witt and Redding 2009). As MNEs operate across environments with diverse resource pressures and institutional arrangements, both resource dependency and institutional theories provide a foundation for considering local stakeholder orientations.

Yet these theories do not apply deterministically in international contexts, because overseas subsidiaries are not subject only to local environmental pressures. Rather, subsidiaries also depend on their parent corporations for crucial resources such as capital, knowledge and reputation. Further, subsidiaries may enjoy discretion in complying with local norms, because local actors do not have identical expectations from foreign and domestic firms (Kostova et al. 2008). In such contexts, maintaining legitimacy might involve responses other than isomorphism with local standards. Yet, despite the importance of MNEs' overseas activities, with very few exceptions (Husted and Allen 2006) almost all related research departs from the corporate level of analysis.

In this paper I address this gap, to elaborate theory about the conditions under which overseas subsidiaries adopt a shareholder-centric approach vs a stakeholder-centric approach to management. In the first part of my study I draw on interview and documentary evidence, compiled across 52 overseas subsidiaries of 14 MNEs. This qualitative evidence allows an in-depth understanding of the mechanisms shaping stakeholder orientation across the MNE. I build on this analysis in the second part using fuzzy-set analysis. Fuzzy-set analysis, which combines features of case-based and quantitative research, allows

greater generalizability than conventional qualitative research. It is particularly appropriate when causality is complex (Fiss 2007).

I find two important explanations for subsidiaries' stakeholder orientations. First, asset-seeking subsidiaries (Dunning 1993), whose activities are centred on research and development, are especially attentive to rewarding non-shareholding stakeholders. However, in contrast to conventional resource dependency explanations focused on external constraints, subsidiaries also face internal constraints emanating from their membership of global corporations. Parent corporations' demands for earnings can prompt efficiency concerns that crowd out the voices of local stakeholders, and produce a focus on shareholder returns and the interests of internal stakeholders such as the corporate headquarters and other units of the MNE.

Second, subsidiaries can face both local and global institutional pressures for social engagement. Even in the absence of local pressures for social engagement, some subsidiaries face global expectations to contribute to the solution of societal problems. The visibility of their parental corporations makes subsidiaries of the very largest MNEs vulnerable to the scrutiny of transnational monitors, including campaign groups and social rating agencies. In contrast, subsidiaries of smaller corporations frequently pass under their radar. Hence, in contrast to conventional explanations rooted in institutional theory, perceived standards of appropriate conduct depend not only on local standards, but also on the expectations of stakeholders who can sanction MNEs globally.

Taken together, these findings have implications for our understanding of the MNE. Although scholars argue for the logic of a single corporate objective function (Jensen 2002), attention to the subsidiary level reveals diversity across subsidiaries of the same corporation. This diversity does not result only from differences across environments. The interview evidence reveals the parent corporation as a key internal stakeholder influencing the actions of its subsidiaries. In contrast to explanations of stakeholder orientation centred on exogenous influences, I show how theory must be nuanced to take account of how corporate and environmental influences interact to shape subsidiaries' stakeholder orientations.

Of particular relevance to this special issue, this paper represents one of the first uses of fuzzy-set analysis to assess causal relationships that have been induced from interview evidence. Based on a subsample of cases, the qualitative analysis clarifies the mechanisms at play. The fuzzy-set analysis explores the generalizability of these mechanisms, shedding light on the circumstances when competing explanations are likely to be valid. More generally, this approach offers one way to bridge the divide between qualitative and quantitative research. It thus complements recent research that adopts mixed methods to study international business phenomena (Brannen and Peterson 2009).

The balance of the paper is organized as follows. I first outline the construct of stakeholder orientation, illustrating the dominant perspectives used to explain how firms prioritize value creation for shareholders vs non-shareholding stakeholders. I then set out the context for the study. Building on qualitative evidence from interviews, I establish propositions that I investigate using fuzzy-set analysis. I conclude by discussing the implications for research and practice.

Stakeholder Orientation

Do business leaders identify as their primary objective the advancement of shareholders' interests, or do they recognize a direct responsibility for the welfare of a broader range of stakeholders, defined as "any group or individual who can affect or is affected by the achievement of an organization's purpose" (Freeman 1984: 53)? Although this question has a long history (Berle 1931; Dodd 1932), the welfare implications of different stances remain hotly contested (Jensen 2002; Walsh et al. 2006).

Stakeholder orientation denotes the stakeholders salient to management, and the objectives that managers aim to achieve through engaging with stakeholders (Berman et al. 1999). A property conception of the firm, premised on assigning residual profit to investors (Allen 1992), represents a narrow orientation. In contrast, a social entity conception, premised on fulfilling social purposes – for example, providing meaningful employment, and supporting communities – while ensuring *sufficient* returns to shareholders (Allen 1992), represents a broad

stakeholder orientation. Firms differ in their attention to stakeholders (Kacperczyk 2009), and may prioritize shareholders, stakeholders involved in the production function or, at the extreme, social actors with no direct implications for financial performance (e.g., Aguilera et al. 2007; Brickson 2005).

Proponents of the property conception of the firm emphasize managers' fiduciary duties to shareholders (Fama and Jensen 1983; Friedman 1970). By maximizing shareholder value, managers indirectly fulfil their responsibilities to other stakeholders (Jensen 2002). In contrast, advocates of stakeholder management argue that no one stakeholder has a clear-cut priority over other groups. Though the dominant concern in the stakeholder literature is the creation of economic value (Freeman 1984), proponents argue that nonshareholding stakeholders might have an interest in the firm's sustainability. For example, employees might invest their time in learning non-transferable skills specific to the enterprise at hand (Blair and Stout 1999), thereby calling the special status of shareholders into question.

Though social engagement can have a strategic imperative (Baron 2001) and be consistent with positive financial performance when the market values firms' social contributions (McWilliams and Siegel 2001), some scholars make a normative case for engagement. Their argument is that stakeholders deserve attention, regardless of their capacity to create wealth (Donaldson and Preston 1995). From this perspective, managers might consider their impacts on an even broader section of society and address societal problems without a business-case justification (Margolis and Walsh 2003).

Existing explanations of stakeholder orientation emphasize the constraints facing managers (Phillips et al. 2010) and differ along instrumental and normative dimensions (Berman et al. 1999). The instrumental case for attending to stakeholders is premised on stakeholders' control over resources (Pfeffer and Salancik 1978). MNEs pursue distinct goals in different territories, often centred on the acquisition of resources (Dunning 1993, 1998). Correspondingly, local resource pressures might shape firm-stakeholder relations. Institutional (Bansal and Roth 2000) drivers of stakeholder orientation are typically premised on an intrinsic (Berman et al. 1999) or normative (Donaldson and Preston 1995) case

rather than any direct consideration of economic benefit. Responsibilities towards stakeholders could be viewed as morally appropriate or become taken for granted. There has been considerable scrutiny, under the lens of institutional theory, of the influence of national systems on firms' interactions with stakeholders and understandings about the role of the firm in society (Ahmadjian and Robbins 2005; Fiss and Zajac 2004; Williams and Aguilera 2008).

Although much analysis relevant to stakeholder orientation departs from the organizational level (e.g., Maignan and Ralston 2002), the limited available evidence supports the existence of different policies and practices across subsidiaries (Husted and Allen 2006). Neither resource dependency theory nor institutional theory applies deterministically to subsidiaries. Resource dependency theory addresses the *external* control of organizations (Pfeffer and Salancik 1978). Subsidiaries also depend on their parent corporations for resources. This internal control over resources could constrain subsidiaries' discretion to address the concerns of local stakeholders. Further, although social responsibility practices are subject to local institutional forces (Husted and Allen 2006), MNEs have some discretion in complying with norms that apply to domestic firms. Rather, MNEs might be expected to "bring something distinctive to their host countries, which is valued and appreciated by local constituents" (Kostova et al. 2008: 999). In short, securing legitimacy might involve responses other than isomorphism with local standards.

Hence existing theories that predominantly address the organizational level of analysis and emphasize adaptation to the environment are less applicable at the subsidiary level. Within the international business arena, theory has to be refined to take account of how subsidiaries prioritize relations with shareholding and non-shareholding stakeholders.

Context of the Study

Investigating subsidiaries' objectives towards stakeholders involves a number of empirical challenges. Publicly available codes and mission statements may not reflect actual policies or behaviours (Weaver et al. 1999). Social performance data, which provide insight into attention to

stakeholders, are unavailable at the subsidiary level. To address these issues I relied on a data set compiled over three years and involving the collaboration of faculty members across six business schools. Granted access to the top management teams of 13 large MNEs (the mean revenue in 2007 was over $60 billion), the research team conducted 298 interviews with (1) executives in company headquarters, (2) heads of overseas subsidiaries, and (3) stakeholders worldwide.

Method

As my central unit of analysis is the overseas subsidiary, a feature of my study is a data set with an intermediate number of observations (52 subsidiaries). Each data point consists of an interview, supplemented by insights from documentary evidence and interviews with stakeholders and headquarter executives. Although the number of observations does not readily permit standard quantitative analysis, it does enable a systematic investigation of the relationships that emerge from the interviews. For this reason, I adopt a two-stage approach to the analysis.

In the first stage, I build on interview evidence to identify features associated with subsidiaries' prioritizing of shareholders over non-shareholding stakeholders, and vice versa. Following Eisenhardt's (1989) recommendation of focusing on extremes, I attend in particular to those subsidiaries pursuing either (1) a shareholder-centric approach with no explicit aim of creating value for non-shareholding stakeholders, or (2) a broad stakeholder-centric approach premised on serving a range of non-shareholding stakeholders. Each case involved the development of a case study, leading to a within-case analysis. The within-case analysis was important for developing constructs and identifying relationships. A subsequent cross-case analysis refined the causal relationships.

In the second stage, I use fuzzy-set analysis to assess all 52 overseas subsidiaries systematically. Fuzzy-set analysis, which has appeared in sociology (Bail 2008; Vaisey 2007) and management journals (Fiss 2007, 2011; Pajunen 2008; Schneider et al. 2010), uses set-theoretic logic to identify relationships between causal conditions and outcomes. This method has advantages when causation is complex, and when different conditions produce identical results (Fiss 2007; Grandori and Furnari 2008).

A feature of my two-stage approach is iteration. The interview evidence serves to identify the causal conditions that I analyse in the second stage. I return to the interviews to interpret the output.

Sample Selection

MNEs were chosen from the natural resources, chemicals, high-technology, industrials, and foods sectors. These sectors encompass diverse patterns of foreign direct investment (FDI): access to physical resources (natural resources), market access (foods and industrials), and the development of assets (chemicals, industrials, and high technology). Corporations in each industry were matched on the basis of headquarter location (either northern Europe or the Anglo-Saxon region). Table 14.1 provides an overview of the MNEs in the study.

Interviews with Headquarter Executives The data set contains 148 semi-structured interviews with senior executives from 13 MNEs. Ninety-six interviewees exercised corporate functions in the headquarters. Informants included CEOs and the directors responsible for the main functions (finance, strategy, marketing, human resources). The average age of the interviewees was 48 years. Interviews, lasting between 50 and 125 min (78 min on average), were recorded and transcribed, producing over 2600 pages of text. The interviews addressed definitions of responsibilities to shareholders and non-shareholding stakeholders, actions (if any) to meet stakeholders' demands, and rationales for these. Appendix 1 provides an extract of the interview protocol.

Interviews with Country Managers Interviews were conducted across 52 country or regional heads based in subsidiaries in North America, Europe, Asia, and Africa. Interviewees exercised general oversight; the most frequent titles were country director, president or manager. The average age of the interviewees was 49.5 years. The same interview protocol (as above) was used, but each interviewee answered questions from the perspective of his or her subsidiary or country unit. The average duration was 72 min. The interviews were transcribed, producing approximately 1450 pages of text.

Table 14.1 Cases

Company	Industry	HQ location	Employee numbers	Executive interviews (HQ and subsidiaries)	Subsidiaries interviewed	Example titles
Compound	Chemicals/ pharma	N. Europe	60,000	15	Russia, China, USA, UK, Italy, Sweden, Germany	President, Asia
Molecule	Chemicals/ pharma	N. Europe	8000	10	USA, Brazil	President, N. America
Substance	Chemicals/ pharma	N. Europe	30,000	10	China, Poland	Country Head, Poland
Antibiotic	Chemicals/ pharma	N. Europe	70,000	10	France	Country Head, France
Drug	Chemicals/ pharma	Anglo-Saxon	80,000	9	Czech Republic, Italy	Managing Director, Czech Republic
Extractor	Natural resources	Anglo-Saxon	65,000	10	USA	President, US-based subsidiary
Explorer	Natural resources	Anglo-Saxon	45,000	7	USA, South Africa	President, US-based product division
Excavator	Natural resources	Anglo-Saxon	>100,000	9	West Africa, France, China, USA, Germany	Regional Director, West Africa
Engineer	Industrials	N. Europe	23,000	14	Italy, Germany, USA, Poland	Regional Manager, E. Europe
Automation	Industrials	N. Europe	>100,000	8	USA	President, US-based product division
Taste	Food	N. Europe	65,000	14	France, Belgium, Italy, Indonesia, Russia, Africa, USA	CEO, Russia

(*continued*)

Table 14.1 (continued)

Company	Industry	HQ location	Employee numbers	Executive interviews (HQ and subsidiaries)	Subsidiaries interviewed	Example titles
Cuisine	Food	N. Europe	>100,000	13	Italy, Poland	Country Chairman, Italy
Venture	High-Tech	Anglo-Saxon	>100,000	12	India, France, Switzerland, Spain, Sweden, Russia	Managing Director, India
Techno	High-Tech	Anglo-Saxon	90,000	7	Belgium, Turkey, France	Chairman, Middle East

Employee numbers have been adjusted by up to 30% (±) to protect the identity of the companies

In six cases, two executives were interviewed for the same firm in the same country, because the executives represented autonomous units with distinct remits and activities

Though the chemicals/pharma sector includes one MNE from outside northern Europe, the subsidiaries report to a northern European centre

Though interviewees may wish to appear more 'other oriented', the dominant norm in business is shareholder primacy (Dore 2008), and executives might be reluctant to underplay their fiduciary duties. Additionally, interviewees were assured of anonymity to reduce potential social desirability bias. Further, the research design included a large number of interviews with stakeholders to triangulate information about corporate and subsidiary behaviours.

Interviews with Stakeholders For the purposes of triangulation, on average 12 stakeholders were interviewed per firm (150 interviews in total). Interviews, lasting between 35 and 120 min, were recorded and transcribed, producing over 3000 pages of transcription. Most stakeholders were industry-structure and socio-political stakeholders (regulators, unions, governments, and nongovernmental organizations). Stakeholders were distributed across the MNEs' home countries and overseas. Eighteen interviews were conducted with stakeholders located in the country of the subsidiary discussed in the interview. In other cases, stakeholders operating across countries were able to provide information about firms' activities in different regions. Stakeholders described their interactions with the firm (or subsidiary), evaluating how it met their expectations and comparing its performance with that of its peers. An extract of the interview protocol is shown in Appendix 2.

Explaining Stakeholder Orientation

In this section I outline the diversity in stakeholder orientation across subsidiaries, showing that this diversity cannot be adequately explained by their firm or location. I then identify some of the conditions with greater predictive power.

Diversity Across Stakeholder Orientations

The interview evidence reveals diversity across, and within, MNEs. Many subsidiaries follow a shareholder-centric model of governance, whereas

others accentuate value creation for employees, contractual partners, and communities. Global stakeholders confirmed the frequent divergence in the conduct of headquarters and overseas subsidiaries:

> EXPLORER has dozens of business units. They vary quite a lot, depending on whether they've been brought in from outside, whether they're from a different culture, whether they are a new or old operation, whether they are in the US, whether they're in some part of Africa. The parent company has a clear policy. But if you look at the business units, you really have to look at them case by case.

This diversity was reflected in descriptions of responsibilities towards stakeholders. Country managers frequently expressed opinions at odds from those of the headquarter staff, and lacked awareness of the practices implemented in the headquarters. The US country manager of a Dutch chemicals firm confessed, "I must say that I don't know what is going on in Europe." Further, many country managers recognized some discretion in setting policy *vis-à-vis* local stakeholders:

> As a site manager, you have a large degree of individual freedom within the various locations around the world, but also a large degree of responsibility, and the responsibility is to the business, is to the people at work, the employees or associates that work for you.

Reflecting shareholder primacy, the head of the US-based subsidiary of EXPLORER, a natural resources company, stated: "You cannot turn a multinational corporation into something that it is not. It is an economic enterprise. The primary responsibility for multinationals is shareholder profitability." Likewise, the French country manager of EXCAVATOR, a competitor, explained how prioritizing shareholders influenced local objectives:

> Total shareholder return is a good indicator for performance, so we give a lot of consideration to shareholders. … Once you satisfy the shareholder, it's nice to have loyal customers, but I don't think we are brave enough to have a growth strategy, which is not going to return good financial performance. … When you have to choose between growth and profitability, most of the time profitability is the winner.

However, within the same country and industry, the US-based country head of EXTRACTOR was unable to allocate priority to any single stakeholder group over others, and expressed the appropriateness of broad social engagement:

> Any business in the resources sector needs to be accountable for making the best use of raw materials or resources. Multinationals have a special responsibility to bring the "best" to all parts.

Country managers emphasized the requirement to adapt policy to local conditions. A corporate manager of EXCAVATOR, with experience in multiple subsidiaries, observed: "In different geographical locations, you tend to get different perspectives. Our stakeholders are different in different parts of the world." Similarly, the manager of CUISINE Russia stated: "There are some corporate responsibilities that don't make any sense in local environments." There was widespread consensus that stakeholder relations had to be determined locally.

Yet the case analysis does not reveal a simple mapping between local environments and stakeholder orientations. Given that subsidiaries within the same locations, MNEs and industries differ in their stakeholder orientation, the question remains: how do we explain the divergence across these subsidiaries in their priorities? In the following section I attend to the two dimensions that most clearly distinguish subsidiaries pursuing a shareholder-centric governance model from those pursuing a model centred on value creation for a broader range of stakeholders.

The External or Internal Control of Resources?

Prior research suggests an instrumental case for stakeholder engagement (Berman et al. 1999), premised on stakeholders' control over resources (Pfeffer and Salancik 1978). I find that strategic objectives linked to efficiency-seeking prompt narrower stakeholder orientations than those based around knowledge development, which require ongoing interaction with local actors. Further, although resource dependency is essentially concerned with the control of resources by *external* stakeholders, subsidiaries also depend on their corporate parents for resources. This dependence on internal resources can crowd out the voices of local stakeholders.

Country managers, when justifying their attention to shareholders or non-shareholding stakeholders, frequently mentioned their strategic objectives. In an acknowledgement of stakeholders' interests in the sustainability of the enterprise, managers, whose subsidiaries required skilled human capital, accentuated the role of non-shareholding stakeholders in ensuring subsidiary performance. Strategic asset seeking forms of FDI are premised on enhancing the value of existing assets and extending these assets (Dunning 1993; Kuemmerle 1999). This objective might heighten the requirement for interaction with local actors, because developing assets might require closer partnerships based on learning (Dyer and Singh 1998; Freeman 1984; Post et al. 2002).

I find extensive evidence that subsidiaries engaged in research and development activities emphasize value creation for local partners. The head of VENTURE India, active in technical development, commented, "Working with the ecosystem definitely brings a big impact for the company. It's about global innovation." While underlining the importance of shareholders, he accentuated the roles of clients, employees and partners in developing technologies.

When stakeholders were required to make investments specific to the activities of the subsidiary, managers readily identified a connection between employee motivation, supplier motivation, and competitive advantage. For example, suppliers invested in plant machinery used in collaboration with MNEs, and employees invested their time to retrain and gain specific skills. In this vein, the German country head of COMPOUND, tying innovation to value creation for all players in the value chain, remarked:

> Innovation should really play a major part. We can only sell new things if we bring an additional value. … And what we've learnt is that the key is about getting more involved in the whole value chain that we are serving here.

Country heads of subsidiaries engaged in innovation, such as VENTURE France, repeatedly emphasized value creation for stakeholders, including employees and customers, and even social stakeholders.

> It's our customers, our suppliers, the people whom we depend on to provide us with capabilities so that we can deliver the end solution to the customer. It is the people in the country, in the community, where we do business.

In contrast, asset-exploiting forms of FDI, premised on taking advantage of capabilities that arise outside the host country, encompass accessing markets, resources, and efficient production (Dunning 1993). Taken together, asset-exploiting forms of FDI appear to prompt less regard for nonshareholding stakeholders. Country managers countered that the objectives of accessing resources and efficiency could be fulfilled by arm's length transactions. In many instances, managers considered stakeholders to act opportunistically and prioritized contractual relationships over looser forms of engagement. As human capital was rarely core to these subsidiaries' competitiveness, managers could easily acquire labour. This notion was summarized by the US manager of EXPLORER:

> The biggest risk is not having access to the resource. Frankly, even if our employees are important to our business, we can always hire more. It's always harder to find a supply of the minerals we need than it is people. Our shareholders are the biggest stakeholders. … They lent us their money to use; we have a responsibility to let them know how we're using it and to get a decent return.

Stakeholders confirmed this pattern. For example, a stakeholder assessing social performance in emerging markets remarked about a high-tech company's manufacturing in Latin America:

> Companies are more and more outsourcing their activities to third parties and using people who work on temporary contracts. If you outsource an activity, it's no longer your activity, which means that the company itself is getting smaller.

Securing access to natural resources may require support from governments and communities in the form of licenses to operate. Although some of these stakeholders exert power (Frooman 1999), subsidiaries'

responses may be largely symbolic in the absence of any sanctioning mechanism (Stevens et al. 2005).

In contrast to their strategic asset-seeking peers, market-seeking subsidiaries make a clear distinction between shareholding and non-shareholding stakeholders. One stakeholder, reporting about a food firm's sales activities in Asia, noted that here, too, employee relationships were being replaced by market contracts:

> They don't consider the promotion workers as employees. They use a language which distinguishes them from paid workers with rights, and they call them promotions staff. All the salaries for their promotion staff come out of a promotions budget, not from a salary line.

Crucially, country managers also referred to the constraints that they faced *internally* as they competed for resources from corporate parents, and tried to meet earnings expectations. Parent corporations can be important suppliers to, and customers of, their subsidiaries. From the perspective of the subsidiary, the corporate parent represented a further stakeholder to consider. The interests of the corporate parent were often deemed equivalent to those of shareholders.

> The shareholder is more equivalent to our head office rather than a shareholder in any meaningful sense, because I make profits which I pay to the head office. Then obviously the profits move to the shareholders, but I see the head office and shareholders as one step, because I don't have any direct shareholder relationship in Italy.
>
> Our stakeholders are group management … and the centralized functions of the groups. Then it's group sales companies which are our customers.

Subsidiaries differ in their dependence on their parent organizations (Kostova and Roth 2002). Corporate slack is particularly relevant to the corporation's control of subsidiary resources. Slack represents "resources funnelled into the satisfaction of individual and sub-group [vs organizational] objectives" (Cyert and March 1963).

By reducing goal conflict in the MNE, its presence could facilitate adaptation to local environments. In this vein, the Chairman of CUISINE mentioned that he was liberal with funds for local stakeholder engagement projects. He approved subordinates' budgets for corporate responsibility programs without assessing the expected returns, and generally left local heads to devise policy locally, depending on the needs of their territories.

In contrast, a lack of corporate slack could lead to greater control over subsidiaries, centralized decision-making, and the dominance of efficiency concerns (Singh 1986). Some country managers perceived tight corporate control in the form of stringent pressures from the top for earnings. This pressure directed attention to the interests of internal stakeholders, such as group companies, and shareholders. The manager of TASTE, a competitor of CUISINE, noted:

> We have to meet ambitious ratios, and we follow them closely. We have five ratios where we compare ourselves (against other units). Of course, there is sometimes conflict. […] We had a meeting last week to inform our people that we have to reduce staff by 25%. Of course, you say that employees are the most valuable assets in a company, but we don't believe that. If you want to stay in business, sometimes you have to take harsh measures to survive.

Excessive parental control could restrict the embeddedness of the subsidiary in the local environment (Mudambi and Navarra 2004). Hence, in contrast to local resource pressures, in particular those for skilled human capital, and which create ties to local actors and embed the subsidiary in its local environment, the internal control of resources through the corporate parent can prompt efficiency concerns and crowd out the voice of local stakeholders.

Table 14.2 provides additional qualitative evidence of these dimensions, tying narrow and broad stakeholder orientations to: (1) external constraints, shaped by subsidiaries' strategic priorities and, in particular, their resulting needs for human capital; and (2) internal constraints, shaped by corporate-level pressures for efficiency.

Table 14.2 External and internal resource constraints

	External resource constraints	Internal resource constraints
Prompting high social engagement	Asset-seeking rationales requiring specialized human capital and local knowledge "Many of the things we do on research are leading-edge technology processes … in cooperation with leading institutions… Shareholders, employees, customers, and society are all very, very important. Without a good relationship with all four groups, we'd go out of business." "Employees are the center of Venture, and so are the clients … Shareholders are more distant."	Parental support "We've been able to call on our core community relations team based in London. We are able to call on those resources whenever we need them." "I've never as a chairman thought of CSR budget proposals as 'I need to make money on this.' As for the budget, I don't think I've ever once said to one of the directors who had a budget in a branch, 'Give me the payback proposal.'"
Prompting low social engagement	Asset-exploiting rationales favouring arm's length transactions "Multinationals have a responsibility to use resources – whether people, commodities, or energy sources – in an efficient way. Waste costs in terms of bottom-line performance." "Shareholders are the most important stakeholder. You have society, the customer, the employee, blah, blah, blah, but the shareholder is stronger economically."	Earnings and efficiency pressures from corporate parent "You can never say you're doing enough. It's the financial aspect that limits that." "Before they became a DRUG subsidiary, they were one of the biggest national companies. Responsibility was instilled in the DNA. (Now) the company is very busy internationally, which gives little time to be present in this country. They could be more involved. They could develop a domestic focus again." (stakeholder interview)

Global or Local Pressures for Legitimacy?

Pressures for legitimacy, rather than material resources, represent a second dimension that explains diversity in stakeholder orientation. Institutional theorists highlight the importance of "social acceptability and credibility" (Scott 2001: 58). Acceptability and credibility are often premised on compliance with local regulations, norms and taken-for-granted ways of doing business (Husted and Allen 2006). Yet pressures for global legitimacy can apply when subsidiaries are vulnerable to stakeholders active in multiple countries. In such cases, maintaining legitimacy goes beyond compliance with domestic expectations to encompass meeting the expectations of global stakeholders.

In many contexts, maintaining legitimacy does involve behaviours similar to those of local firms. Subsidiary managers frequently adapted policies to meet local understandings about the role of the firm. Managers perceived pressures for social engagement when local stakeholders could sanction non-compliance. For example, where civil-society associations and governments were influential – for example, by awarding licenses to operate – managers were careful to implement policy to reflect the demands of a broad section of stakeholders and avoid obvious malpractice.

> We have a lot of complaints from the people living around the plant. You have to respond to them. If you don't, you won't have the chance to survive because they have political power and can make things happen.

However, adopting policies identical to those of domestic firms was not an automatic response. In particular, subsidiaries located in northern Europe – typically considered to be home to a stakeholder model of governance (Hall and Soskice 2001) – often reported less engagement with social stakeholders than domestic firms:

> We also do these things (making social contributions) in VENTURE Denmark, but because of our tax system we don't do it as much. Such a big portion of the employee's tax is already handled by the government, so they would support those areas. … In consumers' perceptions of socially responsible firms, we would rank lower than national companies such as Maersk and Danfoss.

Crucially, some subsidiary managers mentioned pressures for social engagement that did not emanate locally, and which required subsidiaries to follow transnational standards. An EXTRACTOR executive stated, "The rise in social responsibility is the result of the world becoming a more global market place. EXTRACTOR and other companies work in countries where standards of responsibility are not as high as in North America or Europe." MNEs can come under increased scrutiny from globally active stakeholders headquartered in North America and Europe, including social rating agencies and campaign groups. Consequently, a subsidiary's malpractice can have ramifications for other parts of the MNE. A common theme expressed by headquarter staff was the potential danger to corporate reputation from overlooking the interests of local stakeholders in countries that lacked formal sanctions. In contrast to firms' exploitation of poorly regulated territories to evade accountability at home (Vernon 1998), some MNEs face global scrutiny, and have much to lose if their subsidiaries are seen to engage in inappropriate conduct. Stakeholders remarked that misconduct, even in isolated parts of the globe, could not be concealed:

> The speed and efficiency with which communications now travel around the world mean that any major event in the world can be communicated in 10 min. The public at large is much more aware of what is going on. Companies have to take note.
>
> Information is circulating rapidly. People are more informed and more critical of corporations. Corporations could previously hide things. Now they have become much more transparent. People can mobilize in ways that can influence businesses.

Managers stressed that global standards were linked to an emergent sense of expectation that corporations use their influence towards social ends. For example, the United Nations' Global Compact calls on firms to go beyond compliance with national laws and to contribute to the solution of social and environmental problems. Managers declared that growing social expectations went hand in hand with the reach and power of corporations.

> An important reason for the increasing debate on CSR is the globalization process. The regulations of individual states can't ensure that multinational companies comply with the social and environmental standards in each country which they operate in.When a country lacks the capacity for regulation, or the ability to implement regulation, companies have a responsibility to ensure that the way they do their business has no negative impact on environment or society. … Legal compliance is not enough in many cases.

Important in explaining diversity across subsidiaries is that not all subsidiaries perceived equal pressure for social engagement. In part, this can be explained by MNEs' different levels of visibility. In particular, global stakeholders were much better informed about the infringements of large MNEs' subsidiaries, and could name specific instances of human rights abuses overseas. Likewise, managers of the largest MNEs perceived themselves to be under greater scrutiny than those of smaller MNEs. For example, a director of EXCAVATOR, the largest MNE in the sample, expressed a fear of finding his firm on the front page of the *Financial Times* because of inappropriate conduct in a remote territory. In contrast, the manager of COMPOUND's China operations perceived a lower risk that local malpractice could have ramifications on other parts of the corporation: "We have a very low profile. We are not a well-known company here."

Managers report that the size and influence of the corporation prompt increased expectations for societal engagement. This finding is consistent with arguments that large firms could be vulnerable to public pressure (Rowley and Moldoveanu 2003; Siegel and Vitaliano 2007), but also extends prior theory by taking into consideration how constituent parts of the corporation are vulnerable even in the absence of local pressure groups. Ironically, as large firms are better resourced to contribute to economic development, managers reported that size and visibility can become liabilities:

> Because governments are losing governance over societies, and companies are becoming more resourceful, companies have more and more power. Society accepts this under the condition that companies are willing to take on responsibility. It's impossible to get power without taking responsibility.

> Corporations have a role in influencing standards within society. Governments do not cross borders, so you can have an environment where the lowest common denominator sets the future. … More than ever before, large corporations have a cross-border role of setting higher standards of business, corporate and social principles.

Table 14.3 provides additional qualitative evidence for local and global pressures for legitimacy, and provides an overview of the drivers of these pressures.

In summary, the induction suggests two dimensions that influence stakeholder orientation at the subsidiary level. First, subsidiary-level strategic orientations premised on research and development direct attention towards a broad range of stakeholders – including employees, suppliers, and government bodies – perceived as crucial to the long-term competitiveness of the firm. However, in contrast to conventional resource dependency explanations, subsidiary managers are also attentive to the internal control of resources by corporate parents.

Second, local institutional pressures drive subsidiaries to comply with local standards in the presence of effective sanctions from governments and civil society. However, in contrast to conventional institutional explanations centred on adaptation to local contexts, the subsidiaries of large, highly visible MNEs are subject to scrutiny from global stakeholders. In such cases, legitimacy could involve exceeding compliance with local standards.

Fuzzy-Set Analysis

I now advance to a more systematic consideration of these findings. To do so, I employ fuzzy-set analysis, a variant of qualitative comparative analysis (QCA). This approach, lying between conventional qualitative and quantitative analyses, combines the complexity of case analysis with a degree of generalizability through formal analysis. In this part of the study I use the entire sample of 52 overseas subsidiaries, consistent with sample sizes in prior fuzzy-set studies (Pajunen 2008). This sample meets the criteria outlined by Rihoux and Ragin (2008): variance in outcomes

Table 14.3 Local and global pressures for legitimacy

	Local pressures	Global pressures
Features	Ensuring the local license to operate	Concern with the global standing of the corporation
	"Compliance has become more important. The GDP of China has grown at around two digits over the past ten years. When living standards get higher, people care more about the environment and their long-term health. There seems to be a social requirement from the community to understand the risks they are exposed to from neighboring factories."	"A reason for the debate on CSR is the globalization process. Regulations of individual states can't assure that multinational companies comply with social and environmental standards in each country where they operate. So the importance of CSR has increased, representing a way of monitoring and evaluating companies and the impacts of their activities."
	"As environmental problems increase in China, the local authorities realize this and are now catching up with legislation."	"Take the Dow Jones Sustainability Index … You always have to be very careful about how the outside world is judging you, and you also have to be very careful about what measures, yard sticks and beliefs the outside world uses to measure you."
Drivers	Influence of local government and civil society	Size and visibility of parent
	"In Russia the role of the government is much more important than in the US. That's why our dialogue with the government is important. We have to be connected to the governmental institutions. It is the cost of doing business."	"NGOs use events and issues to drive their global message. They look for who is the headline news. For example, if EXCAVATOR caused pollution in a sanctuary for a special marine species, it would be headline news. If a very small company did the same, it might not even catch the press."
	"In Africa, host governments and local communities have an ability to invite anyone in. If EXTRACTOR has better policies in education and health care, it has a strategic advantage."	"Globalization results in the increased scrutiny of multinationals. Large corporations are seen to be returning profits to a few people, and society wants to share in their profits. Communities also want to be compensated for the impacts the company has on them."

(divergence in stakeholder orientation across subsidiaries), commonality and variance in subsidiaries' characteristics, and an intermediate-*N* design that allows familiarity with the individual cases.

Analysis Approach

QCA uses Boolean algebra to identify the causal conditions associated with an outcome.[1] The core idea is that a subsidiary is the member of multiple sets (e.g., the set of R&D-intensive subsidiaries, and the set of subsidiaries of large MNEs). QCA provides techniques to identify patterns between set memberships and outcomes (e.g., a broad stakeholder orientation). QCA techniques allow for equifinality (Fiss 2007), and are therefore appropriate in the present study, because the prior analysis suggests multiple causes of a broad stakeholder orientation.

Building on fuzzy-set theory (Zadeh 1965), fuzzyset analysis departs from standard QCA in that it recognizes that set membership is not always binary. For example, it may be meaningful to distinguish subsidiaries along degrees of research intensity. Many subsidiaries conduct some research to adapt products to local demands. Similarly, few subsidiaries are focused on maximizing returns to shareholders without giving any consideration to the welfare of other stakeholders, and few ignore shareholder returns completely. Rather, most subsidiaries lie between these two extremes.

The first step in performing a fuzzy-set analysis is to calibrate set membership. Unlike quantitative approaches that treat all variance as equally important (Ragin 2008b), the aim of calibration is to identify meaningful groupings of cases. This requires substantive knowledge of the cases at hand or theoretical knowledge (Rihoux and Ragin 2008). Fuzzy-set analysis frequently uses four-level or six-level scales (Ragin 2006a). For example, levels used could be 0, 0.33, 0.67 and 1, where 0 represents non-membership of a set, 1 represents complete membership, and 0.33 and 0.67 represent intermediate levels. Underlying the choice of the number of levels is that different levels of membership should "reflect agreed upon standards" (Ragin 2008a: 80). Correspondingly, my calibration is informed by the existing literature and the qualitative analysis. I provide details below.

The second step involves the construction of a truth table to identify combinations of causal conditions associated with the outcome. The truth table lists all logically possible combinations. The present study involves seven conditions, producing 2^7 causal combinations. To identify the relevant combinations, I delete combinations not associated with any of the subsidiaries in the data set, maintaining only those associated with at least one observation. Setting a frequency threshold of one observation is acceptable when the aim is to build theory from a relatively small sample (Ragin 2006a).

I then specify a consistency threshold. Consistency measures the degree to which a combination of causal conditions is reliably associated with the outcome of a broad stakeholder orientation. Consistency thresholds of at least 0.75 (Ragin 2006a) and up to 0.95 (Epstein et al. 2008) are recommended, but should not be applied mechanistically. One approach is to choose a threshold that corresponds to a gap observed in the distribution of consistency scores (Schneider et al. 2010). Following that approach, I apply a threshold of 0.927.

The next step involves an algorithm to simplify the causal combinations and to arrive at a more parsimonious understanding of the drivers of stakeholder orientation. I employ the truth table algorithm (cf. Ragin 2008a), which generates a range of possible solutions. The reason for this range is that most data sets, especially those of an intermediate size, do not contain instances of all logically possible causal configurations. The truth table algorithm uses counterfactual analysis to speculate about the most plausible outcomes of the combinations that do not exist in the data set.[2] The logically simplest solution is the parsimonious solution, which contains only those conditions considered core and takes advantage of all possible simplifying assumptions. This parsimonious solution is contained within an intermediate solution. The intermediate solution is more conservative, because it only takes advantage of the most plausible simplifying assumptions (Ragin 2008a). It contains not only core conditions, but also peripheral conditions.

Outcome

I study one outcome: stakeholder orientation at the subsidiary level. Based on a coding of the interviews with subsidiary heads, I identify four levels of stakeholder orientation that differ in concern for shareholding

and non-shareholding stakeholders, and which reflect distinct perspectives in the literature. Interviews were coded by at least two researchers on the basis of a clear scheme, and discrepancies were resolved by a panel of three researchers. Below, I outline these levels, and provide illustrative quotations from the interviews.

- 0 (absence of membership): The subsidiary seeks to maximize shareholder value without recognizing a requirement to take account of stakeholder demands. Example: "I haven't seen much evidence that [stakeholder engagement] translates into an advantage for a company. An enterprise is not a philanthropic association. Companies must work to deliver a profit to the shareholders."
- 0.33 (partial membership): The subsidiary seeks to maximize shareholder value under the constraint of satisfying stakeholders. Example: "Our primary objective is that we deliver shareholder value and consistently higher returns on the invested capital, but this is a result of developing relationships with clients, employees, suppliers and communities."
- 0.67 (partial membership): The subsidiary seeks to create value for a range of stakeholders involved in the production function. Example: "We have a responsibility towards our owners, customers, suppliers, and also our colleagues around the world."
- 1 (full membership): The subsidiary seeks to create value for a range of stakeholders, including those that are not involved in the production function. Example: "Companies take up a lot of space and use a lot of resources. Corporations need to look at ways to give back, to create resources for employees, for the environment, for society."

Causal Conditions

I derive the causal conditions from the qualitative evidence presented above. The first dimension that appears important in predicting stakeholder orientation involves external and internal resource constraints. The interviews suggest that these are shaped by (1) subsidiaries' strategic

orientations, and (2) parent corporations' provision of resources. The second dimension involves local and global pressures for legitimacy. The interview evidence suggests that local pressures are shaped by (3) norms concerning firms' roles in society, and (4) the capacity of governments to sanction non-compliance. Exposure to global pressures for engagement is shaped by (5) the visibility of parental corporations on the global stage, and potentially (6) home-country norms, as stakeholders could have an interest to monitor overseas subsidiaries of national firms. Further, (7) host-country development affects legitimacy pressures, because subsidiary managers in developing countries observed particular scrutiny from global stakeholders. Seven causal conditions are appropriate for a sample of 52 observations (Ragin 2006a).

Strategic Orientation I measure strategic orientation by the subsidiary's R&D intensity, and rely on Dunning's (1993) typology of FDI motives to inform my calibration. I reserve full membership (i.e., a score of 1) to subsidiaries primarily engaged in the development of strategic assets. Non-membership (0) applies to subsidiaries involved only in seeking efficiency, seeking markets, or extracting resources. I allocate partial membership to those subsidiaries that, besides other functions, have a responsibility for developing assets for the MNE (0.67). I allocate a lower degree of membership (0.33) to subsidiaries engaged in research and development for local adaptation, or engaged in local process innovation. The data are sourced from country heads' descriptions of their business activities, and triangulated using documentary evidence.

Slack Subsidiaries with broad strategic mandates, including activities requiring substantial local knowledge, are likely to have greater autonomy within the MNE (Edwards et al. 2002). Where corporate resources are freely available, subsidiaries have discretion to engage with local actors. In contrast, earnings pressure exerted by the corporate centre can crowd out the voices of local non-financial stakeholders. A particular dimension associated with efficiency concerns and centralized decision-making is the lack of slack (Singh 1986). Consequently, I take account of slack at the corporate level for the year 2007 using the ratio of working capital

and marketable securities to sales. This ratio represents the level of uncommitted liquid resources (Singh 1986). Based on a clustering of the ratios, MNEs with ratios below 0.07 are coded as entirely out of the set of corporations with slack. Ratios between 0.07 and 0.2 are scored 0.33. Ratios between 0.2 and 0.3 are scored 0.67. MNEs with ratios above 0.3 are coded as full members of the set.

Local Government Influence Based on the interview evidence, governments are a particularly important stakeholder facing MNEs. They influence local standards and grant licenses to operate. Where governments are strong, firms could engage in corporate social responsibility as a form of selfregulation to preclude governmental action. In the interviews, managers identified and ranked the stakeholders that had the greatest impact on the long-term performance of their subsidiaries. Where managers rank government among the top two stakeholders, the subsidiary is accorded full membership (1). In contrast, where the government is not recognized as relevant, the subsidiary is coded as a non-member (0). Partial membership is accorded where the government is ranked elsewhere (0.67 if third or fourth place, 0.33 if less than fourth place).

Collectivism (Host/Home) Expectations for social engagement could differ across countries. In particular, individualist societies might be less likely to call for a balance of attention between shareholders and non-shareholding stakeholders than collectivist societies that emphasize the welfare of multiple stakeholders (Brickson 2005). To capture this dimension, I rely on Hofstede's (1984) measures of collectivism. Host countries scoring above 0.7 on Hofstede's measure, largely east Asian and Latin American societies, are considered full members in the set of collectivist societies. Those scoring between 0.3 and 0.5, including Brazil, India, and the Arab world, are considered to have a high degree of membership at the 0.67 level. Those between 0.5 and 0.8 receive 0.33. These are predominantly continental European countries, many of which are characterized by stakeholder-based management (Hall and Soskice 2001), albeit with movement towards shareholder value orientations (Fiss and Zajac 2006).

Those above 0.8 were coded as fully out of the set of collectivist societies. These are predominantly Anglo-Saxon countries, where firms are assumed to prioritize shareholder value (Hall and Soskice 2001).

The interview analysis also indicates that subsidiaries' legitimacy concerns could also depend on global pressures for social engagement. Correspondingly, I take account of the collectivism of MNEs' home countries, because stakeholders in home countries might pay particular attention to the overseas subsidiaries of MNEs headquartered there.

Corporate Size Corporate size could act as a driver of social engagement (McWilliams and Siegel 2001). The interview analysis suggests that size makes the corporation more visible, increasing its scrutiny by civil society. Subsidiaries of large MNEs appear vulnerable to pressure from globally active stakeholders. Although all MNEs in the sample have annual revenues above $1 billion, they vary in size considerably. Consequently, I create a four level condition to reflect global revenue based on a clustering of the corporations' revenues (2007). Corporations with revenue below $10 billion are scored 0, between $10 billion and $20 billion are scored 0.33, between $20 billion and $100 billion are scored 0.67, and above $100 billion are scored 1.

Host-Country Development Regulation in developed countries is likely to be tighter than in emerging economies (Johnson et al. 1997). However, the interview evidence also suggests that managers in emerging markets are directly exposed to social needs, and identify greater potential for social contribution as well as pressures to observe global standards. The Human Development Index, a measure used by the United Nations Development Programme (UNDP) takes account of (1) education, (2) health, and (3) purchasing power. The UNDP identifies four distinct levels of development: low, medium, high, and very high. I adopt the UNDP's categorization scheme, with countries classified as very highly developed coded as full members of the set of developed countries.

I provide an overview of the seven causal conditions in Table 14.4.

Table 14.4 Overview of causal conditions

Dimension	Relevance	Operationalization
External and internal resource constraints	Needs for local human capital and knowledge	Subsidiary strategic orientation
	Parental largesse with resources	Corporate slack
Local pressures for legitimacy	Local norms re firms' roles	Host-country collectivism
	Capacity to sanction non-compliance	Local government influence
	Thoroughness of local regulation	Host-country development
Global pressures for legitimacy	Global visibility of corporation	Corporate size
	Social expectations in MNE's home country	Home-country collectivism

Results

I provide the calibration table in Table 14.5. The calibration provides an overview of the condition and outcome values for each case, and is the input for the fuzzy-set analysis using the truth table algorithm described above.

To identify the causal conditions associated with a broad stakeholder orientation, I followed the steps above using the fs/QCA software program (Ragin et al. 2006). In doing so, I specified some theoretically grounded assumptions to guide the analysis. Existing research establishes that collectivist values are likely to be associated with a broad stakeholder orientation (Brickson 2005), and hence I performed the analysis specifying that the presence, but not the absence, of home-or hostcountry collectivism might prompt a broad stakeholder orientation. Theory surrounding the other causal conditions did not allow additional assumptions to guide the analysis. Although government influence is, on balance, likely to prompt a broad stakeholder orientation, the regulatory vacuum in some territories does not free corporations from scrutiny (Scherer and Palazzo 2011), and I discuss this in greater detail below.

The analysis suggests six configurations of conditions that predict a broad stakeholder orientation (see Table 14.6). I report below the core conditions in the parsimonious solution. Table 14.6 also shows peripheral conditions that feature in the intermediate solution.

Table 14.5 Calibration table for scope of stakeholder orientation

Case			Outcome	Causal conditions						
Industry	Firm	Subsidiary region	Scope of stakeholder orientation	R&D intensity	Global revenue	Home-country collectivism	Host-country collectivism	Developed host country	Local government importance	Corporate slack
Chem/pharma	Compound	E. Europe	0.33	0	0.33	0.33	0.67	0.67	0	0.67
Chem/pharma	Compound	Asia	1	0.33	0.33	0.33	1	0.33	0	0.67
Chem/pharma	Compound	N. America	0.67	0.67	0.33	0.33	0	1	0	0.67
Chem/pharma	Compound	W. Europe	0.33	0	0.33	0.33	0.33	1	0	0.67
Chem/pharma	Compound	W. Europe	0.67	1	0.33	0.33	0.33	1	0	0.67
Chem/pharma	Compound	W. Europe	0.67	0.67	0.33	0.33	0	1	0.67	0.67
Chem/pharma	Compound	Asia	0.67	0.33	0.33	0.33	1	0.33	0	0.67
Chem/pharma	Compound	W. Europe	0.67	0.33	0.33	0.33	0.33	1	0	0.67
Chem/pharma	Molecule	N. America	0.67	1	0	0.33	0	1	0.33	0.67
Chem/pharma	Molecule	S. America	0.67	0.67	0	0.33	0.67	0.67	0.33	0.67
Chem/pharma	Substance	Asia	0.67	0.33	0.33	0.33	1	0.33	0.33	1
Chem/pharma	Substance	E. Europe	0.33	0	0.33	0.33	0.33	0.67	1	1
Chem/pharma	Antibiotic	W. Europe	0.33	0.33	0.67	0	0.33	1	0.67	1
Chem/pharma	Drug	E. Europe	0.67	0.67	0.67	0	0.33	0.67	0.67	0
Chem/pharma	Drug	W. Europe	0.67	0.33	0.67	0	0.33	1	0.33	0
Resources	Extractor	N. America	0.67	0.33	0.33	0	0	1	0	0.33
Resources	Extractor	N. America	0.33	0	0.33	0	0	1	0	0.33
Resources	Explorer	N. America	0	0	0.67	0	0	1	0.33	0
Resources	Explorer	Africa	1	0	0.67	0	0.33	0.33	1	0
Resources	Excavator	Africa	0.67	0	1	0	0.67	0	0.33	0
Resources	Excavator	W. Europe	0	0	1	0	0.33	1	0.67	0
Resources	Excavator	Asia	0.67	0	1	0	1	0.33	1	0
Resources	Excavator	N. America	0.33	0.33	1	0	0	1	1	0
Resources	Excavator	W. Europe	0.33	0.33	1	0	0.33	1	1	0
Industrials	Engineer	W. Europe	0.67	0	0	0.33	0.33	1	0	0.33
Industrials	Engineer	W. Europe	0.33	0	0	0.33	0.33	1	0	0.33
Industrials	Engineer	N. America	0.67	0.33	0	0.33	0	1	0	0.33

(*continued*)

Table 14.5 (continued)

Case			Outcome	Causal conditions						
Industry	Firm	Subsidiary region	Scope of stakeholder orientation	R&D intensity	Global revenue	Home-country collectivism	Host-country collectivism	Developed host country	Local government importance	Corporate slack
Industrials	Engineer	E. Europe	0.33	0	0	0.33	0	0.67	0	0.33
Industrials	Engineer	W. Europe	0.33	0	0	0.33	0.33	1	0	0.33
Industrials	Automation	N. America	0.67	0.67	0.67	0.33	0	1	0	1
Industrials	Automation	N. America	0.67	0.67	0.67	0.33	0	1	0.67	1
Food	Taste	W. Europe	0.33	0	0.67	0.33	0.33	1	0	0
Food	Taste	E. Europe	0.67	0	0.67	0.33	0.33	0.67	0.33	0
Food	Cuisine	W. Europe	0.33	0	0.33	0.33	0.33	1	0.33	0
Food	Cuisine	W. Europe	0.67	0	0.33	0.33	0.33	1	0.33	0
Food	Cuisine	W. Europe	0.33	0	0.33	0.33	0.33	1	0.67	0
Food	Cuisine	Asia	0.67	0	0.33	0.33	1	0.33	1	0
Food	Cuisine	E. Europe	0	0	0.33	0.33	0.67	0.33	0	0
Food	Cuisine	Africa	0.33	0	0.33	0.33	0.67	0.33	1	0
Food	Cuisine	N. America	0	0	0.33	0.33	0	1	0	0
Food	Cuisine	N. America	0.67	0	0.33	0.33	0	1	0.33	0
High-tech	Venture	Asia	1	0.33	1	0	0.67	0.33	0.33	0
High-tech	Venture	W. Europe	0.67	0.33	1	0	0.33	1	0.33	0
High-tech	Venture	W. Europe	0.33	0.33	1	0	0.33	1	0.33	0
High-tech	Venture	W. Europe	0.67	0	1	0	0.33	1	0	0
High-tech	Venture	W. Europe	0.33	0.67	1	0	0.33	1	0.33	0
High-tech	Venture	E. Europe	0.33	0.33	1	0	0.67	0.67	1	0
High-tech	Venture	W. Europe	0.67	0	1	0	0.33	1	0	0
High-tech	Techno	W. Europe	0.67	0.67	0.67	0	0.33	1	1	1
High-tech	Techno	Asia	1	0	0.67	0	0.67	0.67	0.33	1
High-tech	Techno	W. Europe	0.67	0.67	0.67	0	0.33	1	0.67	1
High-tech	Techno	W. Europe	0.33	0.33	0.67	0	0.33	1	0	1

Table 14.6 Configurations of causal conditions leading to broad/narrow stakeholder orientations

		Configurations for broad stakeholder orientation						Configurations for narrow stakeholder orientation		
	Condition	1	2	3	4	5	6	1	2	3
Local	R&D intensity	●	●	⊗			●	●	⊗	⊗
	Host-country collectivism					●	●	⊗	⊗	⊗
	Government influence	●		●	⊗			⊗	●	⊗
	Developed host country	●	●	⊗	●	●	●	●	●	⊗
Global	Slack		●		●	●	●	⊗	⊗	●
	Global revenue	●		●	⊗		⊗	●		●
	Home-country collectivism							⊗	⊗	⊗
	Consistency	1.00	1.00	1.00	0.934	0.949	0.949	0.917	0.929	0.917
	Coverage	0.231	0.317	0.121	0.352	0.230	0.230	0.148	0.296	0.148
	Unique coverage	0.085	0.025	0.084	0.037	0.012	0.037	0.027	0.216	0.027
		Solution consistency: 0.965 Solution coverage: 0.693				Solution consistency: 0.923 Solution coverage: 0.484				

Key: ● A core causal condition (present); ● peripheral causal condition (present); ⊗ core causal condition (absent); ⊗ peripheral causal condition (absent)

Note: This format of presenting the results from the fuzzy-set analysis is based on Ragin and Fiss (2008)

The Boolean equation linked to a broad stakeholder orientation (prioritizing value creation for a broad range of stakeholders) is:

$$\begin{aligned}\text{Broad orientation} = {} & \text{R \& Dintensity} * \text{government} \\ & + \text{R \& Dintensity} * \text{slack} \\ & + \text{size} * \sim \text{developed} * \text{government} \\ & + \sim \text{size} * \sim \text{government} * \text{slack} \\ & + \text{collectivist host} * \text{develped} \\ & + \text{collectivist host} * \text{slack}\end{aligned} \tag{14.1}$$

Each line represents a configuration of conditions associated with a broad stakeholder orientation. The star (*) represents the Boolean logic term AND. The plus sign (+) represents the Boolean term OR. The tilde (~) represents the Boolean logic term NOT.

In general, the results support the contention that attributes of the local environment interact with subsidiary strategy and corporate-level forces to shape stakeholder orientation. The first two configurations (R&D intensity * government + R&D intensity * slack) imply that subsidiaries requiring specialized human capital are likely to have a broad stakeholder orientation when they are subject to local government influence, or their parent corporations have substantial slack. These are consistent with the interview analysis. In many territories, governments facilitate access to public research institutions. A good illustration is MOLECULE Brazil's research program in agricultural technologies, a context in which both local legislation and public research institutions play important roles in facilitating the development of new solutions. The crucial role of corporate slack – associated with low corporate pressures on subsidiaries for earnings in the short term – in driving attention to local stakeholders underlines the combined effect of subsidiary needs for specialized resources and corporate policy governing resources within the MNE.

The third configuration (size * ~ developed * government) also highlights the interaction between environmental forces and the attributes of the corporate parent. This configuration reflects concerns about

legitimacy rather than material resources. Subsidiaries located in emerging economies are likely to have a broad stakeholder function, and especially so when they belong to large corporations, and where public authorities play an important role. Consistent with the qualitative evidence presented above, the visibility of the largest MNEs makes their subsidiaries vulnerable to stakeholder pressure. This pressure is particularly high in emerging economies, where there could be expectations to respond to societal challenges – for example, facilitating access to education and health care – and where global stakeholders exist to monitor corporate compliance.

The fourth configuration (~ size * ~ government * slack), however, implies that neither corporate size nor local government influence is a necessary condition. In the presence of corporate slack and the absence of these conditions, subsidiaries are likely to adopt a broad stakeholder orientation. This adds evidence that slack, associated with less concern for efficiency, prompts attention to a broad range of stakeholders – including those not directly involved in the production function. Although the link between parental largesse with resources and a broad stakeholder orientation is reasonably intuitive, one puzzling condition in this combination is that the *absence* of government influence in association with slack contributes to a broad stakeholder orientation. How should we interpret this? Revisiting the qualitative evidence, managers from some northern European firms stated that the absence of effective governments in some economies forced their subsidiaries to become more attentive to stakeholder concerns. For example, the lack of social welfare provision meant that subsidiaries stepped in to fulfil the roles played by governments in their home countries. A TASTE head office manager explained a pattern across the firm's subsidiaries:

> Where the needs of the workforce, families and communities were particularly depressing, there was an important role to play beyond the factory gates. (Otherwise) there would be no consumer population there. … The TASTE subsidiaries who are best at playing this role are those in the emerging part of the world, because they are confronted by countries where the government doesn't work as it works in the rest of the world.

Consequently, the absence of strong governments and regulatory oversight does not necessarily allow firms a free hand to neglect stakeholder concerns. Conceivably, governments play contradictory roles. Where governments actively support research and development, their influence can prompt a broad stakeholder orientation in research-active subsidiaries (configuration 1). Yet, where governments do not even provide the basic social services or infrastructure expected by businesses, subsidiaries face the task of supporting the local communities from where their employees and future customers are likely to come.

The final configurations (collectivist host * developed + collectivist host * slack) reflect conventional institutional explanations centred on adaptation to local contexts. Subsidiaries located in collectivist societies are characterized by broad stakeholder orientations. However, the configurations reveal important boundary conditions. First, collectivist norms matter most when the subsidiary is located in a developed country. As stated above, a reason for this effect could be the general absence of formal sanctions against non-compliance in emerging markets (Johnson et al. 1997). Taken together with the third configuration, this suggests distinct causal explanations for stakeholder orientation for subsidiaries in developed and less developed countries. In developed countries, local norms of collectivism matter. In less developed countries, the parent corporation's visibility on the world stage is a more important driver of a broad stakeholder orientation. Second, subsidiaries whose parent firms have substantial slack are also likely to comply with collectivist host-country norms. This finding is consistent with the interview evidence that earnings pressures exerted by the parent could prompt a shareholder orientation, whereas subsidiary autonomy could facilitate its embeddedness in local contexts, and hence its adaptation to local ways of conducting business.

I report two measures of fit in Table 14.6: consistency and coverage. The consistency score measures how well the solution corresponds to the data (Ragin 2006b). This score is calculated for each configuration separately and, subsequently, for the solution as a whole. The measure of consistency can range from 0 to 1, which implies a perfect consistency between the theoretical relationship and the actual data. The scores

reported here – 0.96 for the solution as a whole (0.93 for the parsimonious solution), and between 0.94 and 1.0 for each configuration – suggest the presence of clear set-theoretic relationships.

Solution coverage, in the present study 0.69, measures the empirical importance of the solution as a whole (Ragin 2006b). The raw coverage measures the explanatory power of an individual configuration. However, any single observation might be explained by multiple configurations, and hence I also provide a measure of each configuration's unique contribution to the explanation of a broad stakeholder orientation. An analysis of the coverage figures suggests that the third configuration (size ∗ ~ developed ∗ government) is relatively distinct, because its unique coverage is especially high. This uniqueness, explained by the low number of R&D-intensive subsidiaries operating in undeveloped countries, suggests different underlying causes of a broad stakeholder orientation in developed and less developed economies.

Being able to explain the causes of a broad stakeholder orientation does not necessarily help us to understand when subsidiaries pursue a narrow stakeholder orientation centred around attention to shareholders. Regression analysis assumes causal symmetry: that is, the absence of the conditions leading to a broad stakeholder orientation should predict a narrow stakeholder orientation. However, the interplay between causal conditions at different levels of analysis can make this assumption difficult to justify. In contrast, fuzzy-set analysis is also suitable for investigating causal asymmetry (Fiss 2011; Ragin 2008a). I therefore replicate the analysis to identify the combinations of causal conditions associated with a narrow stakeholder orientation centred around attention to shareholders.

The Boolean equation linked to a narrow stakeholder orientation is:

$$\begin{aligned} \text{Narro} = {} & \text{R \& Dintensity}* \sim \text{government}* \sim \text{slack} \\ & + \sim \text{R \& Dintensity}* \sim \text{collectivist host} \\ & *\text{developed} * \text{government influence}* \sim \text{slack} \\ & + \sim \text{R \& Dintensity} * \text{size}* \sim \text{collectivist host} \\ & * \sim \text{government influence} * \text{slack} \end{aligned} \tag{14.2}$$

The first configuration is crucial, because it demonstrates that a high research and development intensity is not sufficient for a broad stakeholder orientation. In the absence of organizational slack and local government influence, a high research and development intensity can even be associated with a narrow stakeholder orientation. Close parental control and pressure for earnings potentially displace attention to local stakeholders, especially where these are not instrumental in facilitating access to local research institutions, or where they cannot sanction firms.

The second and third configurations imply that in non-collectivist host countries – predominantly Anglo-Saxon countries – subsidiaries do not hold to a broad stakeholder orientation under many conditions. In line with the insights from the interview analysis, many of these conditions relate to attributes of corporate parents. In particular, the absence of slack drives a narrow orientation, consistent with the idea of internal resource constraints. Large size is also associated with a narrow orientation. Hence the relationship between size and stakeholder orientation is a complex one. Paradoxically, in emerging economies, subsidiaries of large MNEs might face substantial pressure for social engagement because they face global pressures for legitimacy, whereas subsidiaries of smaller MNEs might be less vulnerable to this pressure. In contrast, in highly developed countries there are conceivably fewer expectations on MNEs to contribute to the solution of societal problems – as these fall to governments to solve – and larger MNEs might have greater discretion over their conduct.

The coverage (0.484) suggests the existence of other causal paths, not captured in the present analysis. Hence the dimensions that emerged in the induction are more powerful in explaining a broad stakeholder orientation than a narrow, shareholder-centric orientation.

The interactions between causal conditions in the present study support the usage of configurational and set-theoretic methods. These methods complement the prior inductive analysis by showing the empirical importance of the underlying causal conditions. They also demonstrate the value of considering not only net effects (e.g., requirements for specialized human capital), but combined effects (e.g., the interaction of requirements for specialized human capital and being part of a MNE with substantial slack vs being part of an MNE with a tight control over resources). These combined effects allow a deeper understanding of how local, global, and corporate forces interact to shape stakeholder orientations.

Robustness Tests

I conducted a number of robustness checks to understand the stability of the solutions. In particular, following the suggestion of Epstein et al. (2008), I replicated the analysis with a reduced consistency threshold of 0.85. Six combinations of conditions remain in the parsimonious solution, predicting a broad stakeholder orientation; however, the solution consistency is considerably lower (0.75). The configurations are similar to those in the solution presented above, but they are less precise, as might be expected when applying a lower consistency threshold (Schneider and Wagemann 2006). For example, R&D intensity appears sufficient by itself to explain a broad orientation, and this finding arguably sacrifices much of the nuance of the analysis above.

Discussion

The aim of this research was to understand the conditions that lead subsidiary managers to define their objectives narrowly in terms of maximizing shareholder value or, more broadly, in terms of attending to the welfare of a broad range of stakeholders. Existing explanations, giving weight to features of the environment such as local resource pressures and institutional arrangements, are incomplete. Subsidiaries are also subject to internal resource constraints, and the visibility of their parent corporations can subject them to a higher degree of scrutiny than domestic firms. The midrange theory developed in this paper combines conventional explanations focused on environmental factors, and an internal-stakeholder perspective that acknowledges the crucial roles of the parent corporation as resource provider and owner.

Towards a Mid-Range Theory of Stakeholder Orientation

The role of the corporation in society is emerging as an issue at the forefront of management research (Pfeffer 2009). The size and influence of MNEs make this theme particularly salient to international business

scholars. In explaining stakeholder orientation, research has prioritized influences in the external environment. A mid-range theory that links resource pressures, institutional pressures and organizational attributes makes two contributions. First, it provides an explanation for the diversity of stakeholder orientation within the MNE and across MNEs. Second, it contributes to a descriptive theory of stakeholder engagement.

Conventional understandings of firm-stakeholder relationships emphasize the 'extended enterprise' (Post et al. 2002), which denotes the interdependencies between the firm and stakeholders in its resource-base, industry and socio-political environments. Yet centring explanations on the external environment overlooks the importance of the corporate parent, and the coexistence of local responsiveness and global integration in the MNE (Prahalad and Doz 1987). By attending to the subsidiary level, I draw attention to diversity. Some subsidiaries operate according to a shareholder-centric model, whereas others endeavour to fulfil the goals of non-shareholding stakeholders. Stakeholders confirm the variance across territories in the approaches of the same MNE. The identification of these differences has implications for our understanding of the firm. Despite the compelling logic for a sole corporate objective function (e.g., Jensen 2002), the reality is that the MNE is characterized by heterogeneity.

This diversity is not solely the result of environmental pressures. From the perspective of subsidiaries, many crucial stakeholders are internal – not only employees and investors, but also corporate parents and other group companies. Recognizing the corporation's role as an internal stakeholder – as an owner, customer, and supplier – has important implications for our understanding of how subsidiaries define their objectives *vis-à-vis* shareholders and non-shareholding stakeholders, and shifts the focus from concern with external conditions to mid-range theory that integrates both external and internal influences.

Resource dependency theory is concerned with external stakeholders' control over resources. Strategic orientations based on the development of knowledge often prompt a broad stakeholder orientation, whereas those based on efficiency appear to prompt concern for a narrow range of internal stakeholders, chiefly owners. The central idea is that, where human capital is crucial to the maintenance of competitive advantage,

firms often motivate talented employees by making them identify as insiders (Akerlof and Kranton 2005). Further, social contributions can serve as a further motivator of employees and other stakeholders. Hence privileging non-shareholding stakeholders may emanate from strategic asset-seeking strategies that require specific investments from motivated stakeholders. In contrast, efficiency-seeking objectives can often be achieved by arm's length contracting, which relegates employees and suppliers to actors removed from the firm.

Yet, from the perspective of the subsidiary, the corporate parent is also a crucial stakeholder. Subsidiaries depend on corporate parents for resources such as capital, technology, and expertise. This implies that we have to nuance our theories of resource dependency to take account of the balance between external and internal resource constraints. This shifts the focus in explaining stakeholder orientation to internal resource processes within the MNE. When subsidiaries have particular requirements for local knowledge, local adaptation could be important (Edwards et al. 2002). In contrast, the efficiency concerns of corporate parents and their pressures for earnings from subsidiaries potentially narrow subsidiaries' stakeholder orientations.

Likewise, corporate parents' ownership stakes in their subsidiaries make their reputations vulnerable to subsidiaries' misdeeds. In explaining stakeholder-related decisions, institutional theorists conventionally focus on isomorphism with local norms. However, in territories where regulation is poorly enforced, complying with local norms and standards might be less crucial than abiding by transnational standards enforced by global stakeholders. This is especially the case for the subsidiaries of the very largest MNEs, which are subject to scrutiny by civil society actors on the global stage. For example, human rights controversies in the Niger delta can cause substantial negative publicity, not only for Shell Nigeria, but for all Shell companies. Hence, although large MNEs might have discretion to comply with local standards (Kostova et al. 2008), their global reach makes them susceptible to global standards of conduct, even when these cannot be enforced locally.

A core value of this mid-range theory lies in its contribution to a descriptive stakeholder theory: when do subsidiaries prioritize different kinds of stakeholders? Stakeholder theory has strong normative and

instrumental pillars. Normative stakeholder theory seeks to explain managers' moral obligations to engage with stakeholders (Donaldson and Preston 1995). Instrumental stakeholder theory seeks to explain the conditions under which engagement maximizes shareholder value or corporate value (Hillman and Keim 2001; McWilliams and Siegel 2000). Less research investigates the conditions that prompt attention towards different stakeholders (Phillips et al. 2010). Such descriptive theory complements instrumental theory, because it allows scholars to derive mechanisms that direct managerial attention to different kinds of stakeholders. A core mechanism that emerges from the present study is the control exercised by the corporate parent. The autonomy that corporate parents allocate to their subsidiaries may broaden subsidiaries' scope of attention to stakeholders, because autonomy allows for greater interaction with, and resource allocation to, local stakeholders. In contrast, tighter control from corporate parents, linked to pressure for efficiency and earnings, prompts narrow stakeholder orientations in many contexts.

Linking subsidiaries' stakeholder orientations to financial performance represents an avenue for future research. Different orientations could have either positive or negative consequences for competitiveness. Although social engagement can contribute to firms' competitive advantage when it is backed by strategic thinking (Bagnoli and Watts 2003; Baron 2001; McWilliams and Siegel 2001), engagement aimed solely at addressing social ills often destroys corporate value (Hillman and Keim 2001). The present study suggests that some MNEs perceive pressures to make social contributions that appear to go beyond the interests of firms. The liability of foreignness (Hymer 1976) coupled with MNEs' requirements for legitimacy (Kostova and Zaheer 1999) might, however, suggest positive returns to broader engagement with stakeholders in foreign than in domestic locations. This would be consistent with the finding in the present study that, in some contexts, maintaining legitimacy involves going beyond local standards.

A particular challenge involved in linking stakeholder orientation to financial performance at the subsidiary level concerns the collection of data. Prior research has used social ratings to measure value creation for stakeholder groups (Kacperczyk 2009), but social ratings are not available at the subsidiary level. A substitute could be to rely on the judgments of

informed global stakeholders – in a manner similar to the present study. This use of stakeholder ratings is consistent with the perceptual nature of social performance (Wood 1991). As financial data are not systematically available for subsidiaries, market exit might serve as a reasonable proxy. Given the inconclusive, and contradictory, results of prior investigations of stakeholder engagement and firm performance, this research project could employ fuzzy-set analysis. This method would be appropriate for identifying the contingencies on which a positive relationship between stakeholder engagement and financial performance depend.

Implications for Conducting Research in International Business

Thus far, much international business research has used conventional variance-based techniques to analyse large-scale databases. These are useful for understanding the net effects of variables, but they are less appropriate for identifying interactions (Fiss 2007). In contrast, comparative research methods, such as the fuzzy-set analysis in this study, offer a novel approach to uncover causality, especially when there are multilevel influences at work (Greckhamer et al. 2008). Multilevel explanations are particularly important in international business, where organizational and environmental influences interact to shape decisions and performance.

Importantly, the correct application of set-theoretic methods is not mechanistic. Rather, the researcher requires substantive knowledge of individual cases to select causal conditions, calibrate cases, and interpret findings. For this reason, inductive analysis, resting on in-depth interviews and rich data, is a useful first step. Further, iteration between the results from set-theoretic analysis and case-based, qualitative evidence advances the interpretation of the results. Without an in-depth understanding of the cases at hand, set-theoretic analysis could become a shot in the dark as researchers cede to the temptation merely to quantify qualitative data (Pratt 2009).

Inductive research is particularly useful for theory elaboration. Multiple-case studies, typically consisting of between four and 10 cases, provide more robust theories than single case studies (Eisenhardt 1989). Yet relying on small samples limits the generalizability of the resultant theory. Trying

to analyse an intermediate number of cases using conventional case comparison would overstretch the capacities of the researcher. Further, the measures of fit that have been developed for set-theoretic methods, such as fuzzy-set analysis, allow researchers to identify the reliability of their causal arguments and the empirically most relevant conditions.

In the present study, I developed tentative propositions about the causal relationships based on the initial case analysis. Other sequences are possible. In the future, researchers might conduct a fuzzy-set analysis to uncover rare causal configurations, because such configurations could prove particularly fruitful for theory development.

In sum, by bringing the logic and empirical depth of qualitative analysis to intermediate and large numbers of cases, researchers potentially gain greater insight into causal processes. International business is characterized by causally complex phenomena (Kostova and Zaheer 1999). Mixed methods involving set-theoretic methods are particularly suited to address this complexity and, hence, potentially open up new topics for research to international business scholars.

Acknowledgements This research was funded by a grant provided by the 6th Framework Programme of the European Commission (Directorate-General for Research). I am indebted to the Social Innovation Center at INSEAD, and to Lourdes Casanova, Kai Hockerts, Mario Minoja, Peter Neergaard, Esben Pedersen, Francesco Perrini, Susan Schneider, Pamela Sloan, Antonio Tencati, and Maurizio Zollo for contributing to the data collection. I thank Subi Rangan, Javier Gimeno, *JIBS* editor Julian Birkinshaw, and two anonymous reviewers for their insightful comments and suggestions. All remaining errors and omissions are my responsibility.

Appendix 1: Extract of Interview Protocol for Managers

(1) What do you think the responsibilities of multinationals towards society are?
(2) Why do you think that is the case?
(3) What is the corporate responsibility of firms within your sector?

(4) Does corporate responsibility play any role in competing within this sector? Is it a strategic issue?
(5) Some see responsible behaviour as a pure cost factor. What do you think?

The Relevance of Corporate Responsibility

(6) How is corporate responsibility relevant for your day-to-day work? How important an issue is it for you? If so, provide examples of how you integrate these issues in your day-to-day work.

The Firm's Stakeholders

(7) Name the most relevant stakeholders for your company or business unit.
(8) How would you rank them on their impact on your company (business unit)?
(9) How would you rank them based on your company's (business unit's) impact on *their* well-being?

Consider each stakeholder one at a time *(repeat for each stakeholder):*

(10) What is your company's responsibility *vis-à-vis* stakeholder X?
(11) How does your company (unit) *try to meet* its responsibilities with X? Which initiatives have been taken?

Day-to-Day Management Processes

(12) Are socially motivated decisions or initiatives recognized in your performance evaluation?
(13) In your part of the organization, is there a requirement to include an evaluation of the social impact of investment plans in order to decide on project proposals and resource allocation?

Appendix 2: Extract of Interview Protocol for Stakeholders

(1) What is the corporate responsibility of firms within sector X?
(2) Which standards should a multinational follow, home and/or country based?
(3) The firm in comparison with others:

    (3.1) Benchmark how good firm X is in honouring its responsibilities.
    (3.2) How do you evaluate the firm's ability to honour its responsibilities?
    (3.3) Name the firm's most important competitors. Where would you position its competitors (or the industry average)?
    (3.4) On what basis do you decide who is better or worse?

The Company and You

(4) How important is firm X to your organizational goals?
(5) Tell us the specific story of your interaction with firm X.

    (a) What is your organization doing to impact the firm?
    (b) How often do you interact with the firm? When did you last meet?
    (c) Have you noticed any changes arising from this interaction?

The Internal Organization

(6) What are the strengths and weaknesses in the way firm X has organized to handle its responsibilities towards stakeholders?

Notes

1. QCA is based on a paradigm with different assumptions from conventional quantitative methodology (see Fiss 2007; Rihoux and Ragin 2008). To be consistent with this paradigm, I adopt the terminology of QCA researchers. Readers may consider causal conditions to resemble independent variables, and outcomes to resemble dependent variables. Please see Ragin (2008a) for a detailed introduction to the method.
2. For more information on counterfactual analysis, please consult Ragin (2008a). Fiss (2011) provides a helpful overview for management scholars.

References

Aguilera, R.V., D.E. Rupp, C.A. Williams, and J. Ganapathi. 2007. Putting the s back in corporate social responsibility: A multilevel theory of social change in organizations. *Academy of Management Review* 32 (3): 836–863.

Ahmadjian, C.L., and G.E. Robbins. 2005. A clash of capitalisms: Foreign shareholders and corporate restructuring in 1990s Japan. *American Sociological Review* 70 (3): 451–471.

Akerlof, G.A., and R.E. Kranton. 2005. Identity and the economics of organizations. *Journal of Economic Perspectives* 19 (1): 9–32.

Allen, W.T. 1992. Our schizophrenic conception of the business corporation. *Cardozo Law Review* 14 (2): 261–282.

Bagnoli, M., and S. Watts. 2003. Selling to socially responsible consumers: Competition and the private provision of public goods. *Journal of Economics and Management Strategy* 12 (3): 419–445.

Bail, C.A. 2008. The configuration of symbolic boundaries against immigrants in Europe. *American Sociological Review* 73 (1): 37–59.

Bansal, P., and K. Roth. 2000. Why companies go green: A model of ecological responsiveness. *Academy of Management Journal* 43 (4): 717–736.

Baron, D. 2001. Private politics, corporate social responsibility and integrated strategy. *Journal of Economics and Management Strategy* 10 (1): 7–45.

Berle, A.A. 1931. Corporate powers as powers in trust. *Harvard Law Review* 44 (7): 1049–1074.

Berman, S., A.C. Wicks, S. Kotha, and T. Jones. 1999. Does stakeholder orientation matter? The relationship between stakeholder management models

and firm financial performance. *Academy of Management Journal* 42 (5): 488–506.

Blair, M.M., and L. Stout. 1999. A team production theory of corporate law. *Virginia Law Review* 85 (2): 247–328.

Brannen, M.Y., and M.F. Peterson. 2009. Merging without alienating: Interventions promoting cross-cultural organizational integration and their limitations. *Journal of International Business Studies* 40 (3): 468–489.

Brickson, S.L. 2005. Organizational identity orientation: Forging a link between organizational identity and organizations' relations with stakeholders. *Administrative Science Quarterly* 50 (4): 576–609.

Cyert, R.M., and J.G. March. 1963. *A behavioral theory of the firm*. Englewood Cliffs: Prentice-Hall.

Dodd, E.M. 1932. For whom are corporate managers trustees? *Harvard Law Review* 45 (7): 1145–1163.

Donaldson, T., and L.E. Preston. 1995. The stakeholder theory of the corporation: Concepts, evidence, and implications. *Academy of Management Review* 20 (1): 65–91.

Dore, R. 2008. Financialization of the global economy. *Industrial and Corporate Change* 17 (6): 1097–1112.

Dunning, J.H. 1993. Re-evaluating the benefits of foreign direct investment. *Transnational Corporations* 3 (1): 23–52.

———. 1998. Location and the multinational enterprise. A neglected factor? *Journal of International Business Studies* 29 (1): 45–66.

Dyer, J.H., and H. Singh. 1998. The relational view: Cooperative strategy and sources of interorganizational competitive advantage. *Academy of Management Review* 23 (4): 660–697.

Edwards, R., A. Ahmad, and S. Moss. 2002. Subsidiary autonomy: The case of multinational subsidiaries in Malaysia. *Journal of International Business Studies* 33 (1): 183–191.

Eisenhardt, K.M. 1989. Building theories from case study research. *Academy of Management Review* 14 (4): 532–550.

Epstein, J., D. Duerr, L. Kenworthy, and C. Ragin. 2008. Comparative employment performance: A fuzzy-set analysis. In *Method and substance in macro-comparative analysis*, ed. L. Kenworth and A. Hicks, 67–90. Basingstoke: Palgrave Macmillan.

Fama, E.F., and M.C. Jensen. 1983. Separation of ownership and control. *Journal of Law and Economics* 26 (2): 301–325.

Fiss, P.C. 2007. A set-theoretic approach to organizational configurations. *Academy of Management Review* 32 (4): 1180–1198.

———. 2011. Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal* 54 (2): 393–420.

Fiss, P.C., and E.J. Zajac. 2004. The diffusion of ideas over contested terrain: The (non)adoption of a shareholder value orientation among German firms. *Administrative Science Quarterly* 49 (4): 501–534.

———. 2006. The symbolic management of strategic change: Sensegiving via framing and decoupling. *Academy of Management Journal* 49 (6): 1173–1193.

Freeman, R.E. 1984. *Strategic management: A stakeholder approach*. Boston, MA: Pitman.

Friedman, M. 1970. The social responsibility of business is to increase its profits. *The New York Times Magazine*, September 13: 122–126.

Frooman, J. 1999. Stakeholder influence strategies. *Academy of Management Review* 24 (2): 191–205.

Grandori, A., and S. Furnari. 2008. A chemistry of organization: Combinatory analysis and design. *Organization Studies* 29 (3): 459–485.

Greckhamer, T., V. Misangyi, H. Elms, and R. Lacey. 2008. Using qualitative comparative analysis in strategic management research: An examination of combinations of industry, corporate, and business-unit effects. *Organizational Research Methods* 11 (4): 695–726.

Hall, P.A., and D. Soskice. 2001. An introduction to varieties of capitalism. In *Varieties of capitalism: The institutional foundations of comparative advantage*, ed. P.A. Hall and D. Soskice, 1–68. Oxford: Oxford University Press.

Hillman, A.J., and G.D. Keim. 2001. Shareholder value, stakeholder management, and social issues: What's the bottom line? *Strategic Management Journal* 22 (2): 125–139.

Hofstede, G. 1984. *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage Publications.

Husted, B.W., and D.B. Allen. 2006. Corporate social responsibility in the multinational enterprise: Strategic and institutional approaches. *Journal of International Business Studies* 37 (6): 838–849.

Hymer, S.H. 1976. *The international operations of national firms: A study of foreign direct investment*. Cambridge, MA: MIT Press.

Jensen, M.C. 2002. Value maximization, stakeholder theory, and the corporate objective function. *Business Ethics Quarterly* 12 (2): 235–256.

Johnson, S., D. Kaufmann, and A. Shleifer. 1997. The unofficial economy in transition. *Brookings Papers on Economic Activity* 2: 159–239.

Kacperczyk, A. 2009. With greater power comes greater responsibility? Takeover protection and corporate attention to stakeholders. *Strategic Management Journal* 30 (3): 261–285.

Kostova, T., and K. Roth. 2002. Adoption of an organizational practice by subsidiaries of multinational corporations: Institutional and relational effects. *Academy of Management Journal* 45 (1): 215–233.

Kostova, T., and S. Zaheer. 1999. Organizational legitimacy under conditions of complexity: The case of the multinational enterprise. *Academy of Management Review* 24 (1): 64–81.

Kostova, T., K. Roth, and M.T. Dacin. 2008. Institutional theory in the study of multinational corporations: A critique and new directions. *Academy of Management Review* 33 (4): 994–1006.

Kuemmerle, W. 1999. The drivers of foreign direct investment into research and development: An empirical investigation. *Journal of International Business Studies* 30 (1): 1–24.

Maignan, I., and D.A. Ralston. 2002. Corporate social responsibility in Europe and the US: Insights from businesses' selfpresentations. *Journal of International Business Studies* 33 (3): 497–514.

Margolis, J.D., and J.P. Walsh. 2003. Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly* 48 (2): 268–305.

McWilliams, A., and D. Siegel. 2000. Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal* 21 (5): 603–609.

———. 2001. Corporate social responsibility: A theory of the firm perspective. *Academy of Management Review* 26 (1): 117–127.

Mitchell, R.K., B.R. Agle, and D.J. Wood. 1997. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. *Academy of Management Review* 22 (4): 853–886.

Mudambi, R., and P. Navarra. 2004. Is knowledge power? Knowledge flows, subsidiary power and rent-seeking within MNCs. *Journal of International Business Studies* 35 (5): 385–406.

Pajunen, K. 2008. Institutions and inflows of foreign direct investment: A fuzzy-set analysis. *Journal of International Business Studies* 39 (4): 652–669.

Pfeffer, J. 2009. Renaissance and renewal in management studies: Relevance regained. *European Management Review* 6 (3): 141–148.

Pfeffer, J., and G.R. Salancik. 1978. *The external control of organizations*. New York: Harper & Row.

Phillips, R.A., S.L. Berman, H. Elms, and M.E. Johnson-Cramer. 2010. Strategy, stakeholders, and managerial discretion. *Strategic Organization* 8 (2): 176–183.

Post, J.E., L.E. Preston, and S. Sachs. 2002. *Redefining the corporation: Stakeholder management and organizational wealth*. Palo Alto: Stanford University Press.

Prahalad, C.K., and Y. Doz. 1987. *The multinational mission*. New York: Free Press.

Pratt, M.G. 2009. For the lack of a boilerplate: Tips on writing up (and reviewing) qualitative research. *Academy of Management Journal* 52 (5): 856–862.

Ragin, C.C. 2006a. *User's guide to fuzzy-set/qualitative comparative analysis 2.0.* Tucson: University of Arizona Department of Sociology.

———. 2006b. Set relations in social research: Evaluating their consistency and coverage. *Political Analysis* 14 (3): 291–310.

———. 2008a. *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago: University of Chicago Press.

———. 2008b. Measurement versus calibration: A set-theoretic approach. In *The Oxford handbook of political methodology*, ed. J. Box Steffensmeier, H. Brady, and D. Collier, 174–198. Oxford: Oxford University Press.

Ragin, C.C., and P. Fiss. 2008. Net effects versus configurations: An empirical demonstration. In *Redesigning social inquiry: Fuzzy sets and beyond*, ed. C.C. Ragin, 190–212. Chicago: University of Chicago Press.

Ragin, C.C., K.A. Drass, and S. Davey. 2006. *Fuzzy-set/qualitative comparative analysis 2.0.* Tucson: University of Arizona Department of Sociology.

Rihoux, B., and C.C. Ragin, eds. 2008. *Configurational comparative methods: Qualitative Comparative Analysis (QCA) and related techniques*. London: Sage.

Rowley, T.J., and M. Moldoveanu. 2003. When will stakeholder groups act? An interest-and identity-based model of stakeholder group mobilization. *Academy of Management Review* 28 (2): 204–219.

Scherer, A.G., and G. Palazzo. 2011. The new political role of business in a globalized world: A review of a perspective on CSR and its implications for the firm, governance and democracy. *Journal of Management Studies* 48 (4): 899–931.

Schneider, C.Q., and C. Wagemann. 2006. Reducing complexity in qualitative comparative analysis (QCA): Remote and proximate factors and the consolidation of democracy. *European Journal of Political Research* 45 (5): 751–786.

Schneider, M., C. Schulze-Bentrop, and M. Paunescu. 2010. Mapping the institutional capital of high-tech firms: A fuzzy-set analysis of capitalist variety and export performance. *Journal of International Business Studies* 41 (2): 246–266.

Scott, W.R. 2001. *Institutions and organizations*. London: Sage.

Siegel, D.S., and D.F. Vitaliano. 2007. An empirical analysis of the strategic use of corporate social responsibility. *Journal of Economics and Management Strategy* 16 (3): 773–792.

Singh, J.S. 1986. Performance, slack, and risk taking in organizational decision making. *Academy of Management Journal* 29 (3): 562–585.

Stevens, J.M., H.K. Steensma, D.A. Harrison, and P.L. Cochran. 2005. Symbolic or substantive document? The influence of ethics codes on financial executives' decisions. *Strategic Management Journal* 26 (2): 181–195.

UNDP. 2009. *Human development report 2009*. New York: Palgrave Macmillan.

Vaisey, S. 2007. Structure, culture, and community: The search for belonging in 50 urban communities. *American Sociological Review* 72 (6): 851–873.

Vernon, R. 1998. *In the hurricane's eye: The troubled prospects of multinational enterprises*. Cambridge, MA: Harvard University Press.

Walsh, J.P., A.D. Meyer, and C.B. Schoonhoven. 2006. A future for organization theory: Living in and living with changing organizations. *Organization Science* 17 (5): 657–671.

Weaver, G.R., L.K. Trevino, and P.L. Cochran. 1999. Corporate ethics programs as control systems: Influences of executive commitment and environmental factors. *Academy of Management Journal* 42 (1): 41–57.

Williams, C.A., and R.V. Aguilera. 2008. Corporate social responsibility in a comparative perspective. In *The Oxford handbook of corporate social responsibility*, ed. A. Crane, A. McWilliams, D. Matten, J. Moon, and D.S. Siegel, 452–472. Oxford: Oxford University Press.

Witt, M.A., and G. Redding. 2009. Culture, meaning, and institutions: Executive rationale in Germany and Japan. *Journal of International Business Studies* 40 (5): 859–885.

Wood, D.J. 1991. Corporate social performance revisited. *Academy of Management Review* 16 (4): 691–718.

Zadeh, L.A. 1965. Fuzzy sets. *Information and Control* 8 (3): 338–353.

15

Foreign Subsidiary Stakeholder Orientation and FsQCA: A Commentary

Stav Fainshmidt

Introduction

Many international business (IB) phenomena are inherently configurational and thus call for methodologies that can address configurational patterns. Based on Boolean algebra (Ragin 2000), fuzzy-set Qualitative Comparative Analysis (fsQCA) is a set-theoretic technique that demonstrates how the membership of cases in causal conditions relates to their membership in an outcome of interest. It accommodates complex causality by examining causal conditions in concert rather than as independent net effects (holding other variables constant) and by permitting equifinality, whereby more than one configuration of causal conditions can lead to the same outcome. For those interested in fsQCA and best practices related to this approach, several primers are available (e.g., Fiss et al.

Crilly, D. 2011. Predicting stakeholder orientation in the multinational enterprise: A mid-range theory. *Journal of International Business Studies* 42 (5): 694–717.

S. Fainshmidt (✉)
Florida International University, Miami, FL, USA
e-mail: sfainshm@fiu.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_15

2013; Greckhamer et al. 2018; Misangyi et al. 2017; Ragin 2000, 2008; Schneider and Wagemann 2012).

Crilly (2011) is an excellent example of how IB scholars can use fsQCA to advance IB theory. Crilly leverages fsQCA to build theory regarding the stakeholder orientation of foreign subsidiaries in their host-country environment. Crilly's (2011) primary approach was to use qualitative data collected from 52 foreign subsidiaries. However, he brought more rigour to his study by analysing the qualitative data with fsQCA. This exercise proved fruitful in identifying configurations of external and internal conditions associated with the presence of a broad or narrow stakeholder orientation. One of the advantages of combining fsQCA with traditional qualitative analysis is that it is possible to straddle both qualitative and quantitative analyses in an iterative manner, which helps to pinpoint causal mechanisms as well as generalize and contextualize qualitative findings. Additionally, Crilly demonstrates the ability of fsQCA to analyse phenomena that span multiple levels of analysis. FsQCA is not a variance-based technique and thus is particularly useful when causality is complex and multiple explanations are plausible (Fiss 2007).

Whether one is employing mixed methods or not, fsQCA requires that the researcher make judgement calls throughout the analytical process. As Crilly notes (p.714), "the correct application of set-theoretic methods is not mechanistic. Rather, the researcher requires substantive knowledge of individual cases to select causal conditions, calibrate cases, and interpret findings." Hence, it is critical that the researcher provide ample transparency and explanation. Here, I comment on a few key judgement calls involved in utilizing fsQCA.

Calibration

Calibration entails defining set membership in causal and outcome conditions for the cases in the sample. Calibration is critical because it determines the extent to which each case exhibits a given condition using specific anchor points that give meaning to the data. A robust calibration reduces erroneous set-theoretic relations and strengthens the validity of

the results. Ideally, calibration decisions are based on theoretical criteria, but insights from the data collection process can be used when such criteria are unavailable. For example, when calibrating country-level collectivism, Crilly's approach aligns well with the vast research on national culture and cultural clusters. Other data provide natural anchors for calibration (e.g., UNDP country development classification) or are informed by interview insights (e.g., stakeholder orientation).

Frequency and Consistency Thresholds

Frequency thresholds indicate the minimum number of times a given configuration of causal conditions must be present in the sample for it to be considered relevant for analysis. Providing a substantive explanation of the choice of frequency and consistency thresholds is good practice for authors, whenever possible. A frequency of 1 is appropriate for small to medium sample sizes, but this number is usually higher for large sample sizes. For example, Crilly has a sample size of 52 (a medium sample size for fsQCA) and he specifies a frequency threshold of 1, meaning that a configuration is considered if it is exhibited by at least one case in the data. This decision is appropriate given the intimate knowledge Crilly has with the cases in the data (and the resulting reduction in the likelihood of error leading to miscalibration).

Broadly speaking, consistency measures the extent to which membership in one condition is a subset of membership in another condition. A higher consistency threshold raises the bar for how strongly configurations must be associated with the outcome for them to be considered as potentially causing that outcome. Crilly uses a high value based on his identification of an inflection point among consistency scores in the truth table. This is a commonly accepted technique, as it may indicate a natural distinction between two groups of configurations (consistent vs. inconsistent with the outcome). Another statistic that can help researchers utilizing fsQCA is the proportional reduction in inconsistency (PRI). PRI consistency can help correct for the potential contribution of paradoxical cases by giving "minor penalties for small inconsistencies and substantial penalties for large inconsistencies" (Van der Heijden, 2015, p. 584).

When dealing with fuzzy sets, using PRI can help identify paradoxical rows in the truth table and thus produce a more accurate solution.

One of the trade-offs in fsQCA is between consistency and coverage. Coverage (i.e., explanatory power) indicates the extent to which configurations explain the outcome. For instance, requiring very high consistency with the outcome might cause some configurations not to be considered, but these configurations might cover a unique portion of the outcome. Similarly, allowing for more relevant configurations to be considered in the solution might entail relaxing consistency requirements. In Crilly (2011), the high consistency threshold resulted in a solution containing configurations that are highly consistent with the outcome (e.g., 0.96). The solution coverage scores are appropriate as well (e.g., 0.69, 0.48), and all of the identified configurations contribute uniquely to explaining the outcome (i.e., they have non-zero unique coverage).

Additional Analyses

Given the importance of researcher choices in fsQCA, it is good practice to examine how sensitive results are to such choices. Crilly conducts a robustness test with a reduced consistency threshold, thus providing an assessment of the stability of the solution with regard to this important judgement call. Other analyses that researchers often conduct include, among others, varying the frequency threshold, evaluating the impact of alternative calibration approaches, and revisiting decisions related to counterfactuals. Conducting stability tests with higher or lower frequency and consistency thresholds can provide insights regarding patterns in the data, particularly for lower sample sizes. These analyses should be considered in light of the decisions made in the main analysis and the context of the study or data at hand. Crilly also conducts an analysis of the absence of the outcome of interest, thus highlighting the ability of fsQCA to accommodate asymmetry in the conditions associated with the outcome versus those associated with the negation of the outcome. For instance, a certain organizational structure might facilitate product innovation, but the absence of that structure may not necessarily inhibit product innovation. In some cases, the presence of this structure in a

configuration of complementary organizational conditions may inhibit product innovation, but in other cases the same structure may be irrelevant. This analysis is not necessary due to the set-theoretic nature of fsQCA, but it can be a useful exercise in some cases to further flesh out insights pertaining to the outcome (e.g., Crilly's insights regarding the complexity in the role of firm size). In other cases, it may not reveal much (e.g., Fiss 2011).

Counterfactuals

FsQCA operates with available data, but it can also account for counterfactuals, or configurations for which there is no empirical evidence and "therefore must be imagined" (Ragin 2008, p. 150). These configurations can help simplify the solution, specifically the complex solution that is based only on set-theoretic relations in the data, because counterfactuals add information otherwise unavailable in the analytical process. There are excellent works on counterfactuals and their usage in fsQCA (e.g., Ragin 2008; Ragin and Sonnett 2005; Soda and Furnari 2012). Technically, this is done by specifying the expected set-theoretic relation between each causal condition and the outcome, which is reflected in the intermediate solution. The parsimonious solution is then further simplified and contains only configurations that would likely not be minimized in the face of additional information, thus providing stronger evidence for a causal set-theoretic relationship between a configuration or condition (also considered "core") and the outcome. However, it requires the usage of "difficult" counterfactuals, namely, those that are more difficult to "imagine" based on existing knowledge. In Crilly's case, he specified that the presence of collectivism should prompt a broad stakeholder orientation. Importantly, he did so by drawing on established theory, and he clearly details these specifications in the article. As for the other conditions, he notes that existing theory did not provide a sufficient basis to specify the set-theoretic relations a priori.

Conclusions

Overall, Crilly's article makes notable contributions to IB theory, but it also reflects many of the best practices and usefulness in the application of fsQCA, thus providing an exemplar for rigorously applying fsQCA to advance IB research. There is much that can be done with configurational analysis in IB studies:

- "Researchers might conduct a fuzzy-set analysis to uncover rare causal configurations, because such configurations could prove particularly fruitful for theory development." (Crilly 2011, p. 714).
- FsQCA is useful not only for theory building but also for theory testing (e.g., Bell et al. 2014; Fainshmidt et al. 2019) and theory elaboration (e.g., Dwivedi et al. 2018).
- In recent years, there have been impressive advances in the application of fsQCA for large-N data (Greckhamer et al. 2013; Vis 2012) and various longitudinal designs (Aversa et al. 2015; Caren and Panofsky 2005; Ragin and Strand 2008).
- Scholars can use fsQCA to revisit important research questions in IB that have not been conclusively answered, such as explaining foreign subsidiary performance or foreign location choice, or to explore entirely new questions with configurational thinking. FsQCA has the advantage of allowing for multiple interplays among causal conditions, whereas variance-based techniques are not well suited for interactions including more than three variables.

In sum, many IB phenomena are configurational, so our empirical approaches should be as well. FsQCA is one tool that, when applied rigorously, can allow IB researchers to break new ground.

References

Aversa, P., S. Furnari, and S. Haefliger. 2015. Business model configurations and performance: A qualitative comparative analysis in formula one racing, 2005–2013. *Industrial and Corporate Change* 24 (3): 655–676.

Bell, R.G., I. Filatotchev, and R.V. Aguilera. 2014. Corporate governance and investors' perceptions of foreign IPO value: An institutional perspective. *Academy of Management Journal* 57 (1): 301–320.

Caren, N., and A. Panofsky. 2005. TQCA: A technique for adding temporality to qualitative comparative analysis. *Sociological Methods & Research* 34 (2): 147–172.

Crilly, D. 2011. Predicting stakeholder orientation in the multinational enterprise: A mid-range theory. *Journal of International Business Studies* 42 (5): 694–717.

Dwivedi, P., A. Joshi, and V.F. Misangyi. 2018. Gender-inclusive gatekeeping: How (mostly male) predecessors influence the success of female CEOs. *Academy of Management Journal* 61 (2): 379–404.

Fainshmidt, S., L. Wenger, A. Pezeshkan, and M.R. Mallon. 2019. When do dynamic capabilities lead to competitive advantage? The importance of strategic fit. *Journal of Management Studies*. https://doi.org/10.1111/joms.12415.

Fiss, P.C. 2007. A set-theoretic approach to organizational configurations. *Academy of Management Review* 32 (4): 1180–1198.

———. 2011. Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal* 54 (2): 393–420.

Fiss, P.C., B. Cambre, and A. Marx, eds. 2013. *Configurational theory and methods in organizational research*, Research in the Sociology of Organizations. Vol. 38. Bingley: Emerald Group Publishing Limited.

Greckhamer, T., V.F. Misangyi, and P.C. Fiss. 2013. Chapter 3: The two QCAs: From a small-N to a large-N set theoretic approach. In *Configurational theory and methods in organizational research*, Research in the Sociology of Organizations, ed. P.C. Fiss, B. Cambre, and A. Marx, vol. 38, 49–75. Bingley: Emerald Group Publishing Limited.

Greckhamer, T., S. Furnari, P.C. Fiss, and R.V. Aguilera. 2018. Studying configurations with qualitative comparative analysis: Best practices in strategy and organization research. *Strategic Organization* 16 (4): 482–495.

Misangyi, V.F., T. Greckhamer, S. Furnari, P.C. Fiss, D. Crilly, and R. Aguilera. 2017. Embracing causal complexity: The emergence of a neo-configurational perspective. *Journal of Management* 43 (1): 255–282.

Ragin, C.C. 2000. *Fuzzy-set Social Science*. Chicago: University of Chicago Press.

———. 2008. USER'S GUIDE TO Fuzzy-Set / Qualitative Comparative Analysis. http://www.u.arizona.edu/~cragin/fsQCA/download/fsQCAManual.pdf

Ragin, C.C., and J. Sonnett. 2005. Between complexity and parsimony: Limited diversity, counterfactual cases, and comparative analysis. In *Vergleichen in der Politikwissenschaft* [Comparing in political science], ed. S. Kropp and M. Minkenberg, 180–197. Wiesbaden: VS Verlag fur Sozialwissenschaften.

Ragin, C.C., and S.I. Strand. 2008. Using qualitative comparative analysis to study causal order: Comment on Caren and Panofsky (2005). *Sociological Methods & Research* 36 (4): 431–441.

Schneider, C.Q., and C. Wagemann. 2012. *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge: Cambridge University Press.

Soda, G., and S. Furnari. 2012. Exploring the topology of the plausible: Fs/QCA counterfactual analysis and the plausible fit of unobserved organizational configurations. *Strategic Organization* 10 (3): 285–296.

Van der Heijden, J. 2015. The role of government in voluntary environmental programmes: A fuzzy set qualitative comparative analysis. *Public Administration* 93 (3): 576–592.

Vis, B. 2012. The comparative advantages of fsQCA and regression analysis for moderately large-N analyses. *Sociological Methods & Research* 41 (1): 168–198.

Part VIII

Interaction Effects

16

Explaining Interaction Effects Within and Across Levels of Analysis

Ulf Andersson, Alvaro Cuervo-Cazurra, and Bo Bernhard Nielsen

Introduction

As editors, we are increasingly seeing papers with interaction effects – also known as multiplicative effects, product terms or moderation effects – that benefit from the powerful statistical analyses now available to scholars. Such research strategy has the potential to yield new theoretical

U. Andersson
Malardalen University, Vasteras, Sweden

BI Norwegian Business School, Oslo, Norway
e-mail: ulf.r.andersson@mdh.se

A. Cuervo-Cazurra (✉)
Northeastern University, Boston, MA, USA
e-mail: a.cuervocazurra@neu.edu

B. B. Nielsen
The University of Sydney Business School, The University of Sydney, Sydney, NSW, Australia

Copenhagen Business School, Frederiksberg, Denmark
e-mail: bo.nielsen@sydney.edu.au

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_16

insights that may advance the international business (IB) field. However, incorporating interaction effects is challenging because it identifies new and complex relationships that needs to be adequately explained. To help authors, in this editorial we provide suggestions for how best to explain the theoretical mechanisms[1] behind proposed interaction effects in order to clarify the theoretical contribution of their studies. We go beyond statistical explanations of interaction effects and their detection (see, e.g., Aguinis and Gottfredson 2010; Jaccard and Turrisi 2003; Shieh 2009), which, depending on how the variables are measured and the type of statistical method used, can be quite challenging.

We discuss two types of interaction effects: within and across levels of analysis. First, for interactions within levels of analysis, we suggest that authors first provide an explanation of the theoretical mechanisms that link the main independent variable to the dependent variable, and then explain how and why the interaction variable modifies these theoretical mechanisms. Additionally, we suggest that authors theoretically rule out the existence of a reverse interaction effect in which the independent variable is actually affecting the relationship between the moderator and dependent variable. Second, for interactions across levels of analysis, we suggest that authors first identify the level of analysis of the main relationship, then specify the cross-level nature of the moderating relationships, before clarifying the hierarchy and nature of theoretical nesting. In addition, we propose that authors theoretically explain the multilevel influences, separating the justification of the cross-level interaction effect from the explanation of the cross-level direct influences.

Explaining Interaction Effects

Interaction Effects

Generally, interaction is said to occur when the effect of an independent variable (X) on a dependent variable (Y) varies across levels of a moderating variable (Z). Identifying and specifying relevant and important interaction effects pertaining to relations between independent and dependent variables is at the heart of theory in social science (Cohen et al. 2003) and indicates

the maturity and sophistication of a field of inquiry (Aguinis et al. 2001). Interactions provide researchers with the ability to enrich our understanding of economic and social relationships by establishing the conditions under which such relationships apply, or are stronger or weaker. As such, interactions enable the extension of well-known relationships to contexts that the original research did not consider, and they also help provide more detailed predictions about the relationships, going beyond the simplistic argument "it depends". However, merely detecting a statistically significant effect of the interaction between independent and moderating variables on the dependent variable is not sufficient to be considered a contribution to the literature. The interaction effect has to be explained, and there must be theoretical arguments for why including this interaction results in better theory.

Research questions involving moderators typically address "when" or "under what conditions" an independent variable most strongly influences an outcome variable. More specifically, a moderator is a variable that alters the *nature* or *strength* of the relationship between an independent and an outcome variable (Baron and Kenny 1986). The distinction between circumstances where the *nature* of the relationship of X on Y varies as a function of Z (differential prediction) vs the *strength* of the relationship of X on Y varies as a function of Z (differential validity) is important for several reasons. First, only differential prediction is appropriately tested with moderated multiple regression, which is the statistical test typically employed in moderation studies (Carte and Russell 2003). Differential validity is typically tested via subgroup moderation: the sample is split into two or more groups based on the level of the moderator variable, and t-tests of the correlation coefficients and χ^2 tests are performed to assess the strength of the moderation effect and differences among groups. Second, the language and argumentation employed in moderation hypotheses is often inaccurate in relation to the actual tests performed. For instance, if a researcher asserts that "the strength of the multinationality–performance relationship depends on the level of product diversification", then he/she must report differences in strength of the multinationality–performance relationship (i.e. $r_{\text{multinationality–performance}}$) across levels of product diversification rather than the often-reported differences in the slope (nature) of the multinationality–performance relationship across levels of product diversification. Scholars must specify the

role of the moderation a priori and make sure that the language, theoretical argumentation, and ensuing empirical tests match.

The choice of the moderating variable should be based on a specific theory regarding why, or under what conditions, a given relationship may be significantly influenced for some types of firms, teams, or individuals rather than for others. This choice is important because it drives the specific type of interaction that needs to be explained. First, there are interactions between two continuous variables, which can take three typical patterns (Cohen et al. 2003: 285–286): (a) enhancing interactions, in which both the predictor and moderator affect the outcome variable in the same direction and together they have a stronger effect than a merely additive one; (b) buffering interactions, in which the moderator variable weakens the effect of the predictor variable on the outcome; and (c) antagonistic interactions, in which the predictor and moderator have the same effect on the outcome but the interaction is in the opposite direction. Second, there are interactions between a categorical variable and a continuous variable, which can take two different patterns: (a) existence interaction, when an independent variable is positively related to the dependent variable for one particular group but unrelated for another group; and (b) competing interactions, when an independent variable is positively related to the dependent variable for one particular group but it is negatively related for another.

The distinction between the different patterns of interaction has important implications for theory, as the selection of the particular type of interaction should be driven by the specific nature of the concepts analyzed rather than by the particular measurement of the variables used in the statistical analysis. Although all interaction types have the potential for advancing theory, the buffering and antagonistic interactions between continuous variables, and the competing interaction between a categorical variable and continuous variable, hold the greatest potential because they are more likely to challenge existing theory.

Challenges in Explaining Interactions

Regardless of the particular type of interaction proposed, the following are some of the common challenges we find in the explanation of interaction effects in many initial drafts of manuscripts that propose an interaction.

First, there is often no explanation of, or indeed theoretical justification for, the direct effect. Far too often, manuscripts simply start with an explanation of the interaction effect. One reason for this may be that the authors think that the novelty of the paper resides in the interaction effect, because the direct effect has been explained in detail and tested before. However, this approach is problematic because the theoretical mechanism explaining the baseline argument remains unspecified. As a result, it becomes unclear what baseline effect the interaction is supposed to modify. This is particularly problematic because many management and IB phenomena can be explained from many alternative theoretical perspectives; the different mechanisms that link the independent variable to the dependent variable may be rooted in different theories, which offer different logics even if they end up resulting in the same hypothesized relationship. For instance, the relationship between multinationality and performance can be explained from the theoretical standpoints of internalization theory (Buckley and Casson 1976), behavioural theory of internationalization (Johanson and Vahlne 1977), or resource-based view (Penrose 1959), among others. While related, each theory argues for different mechanisms explaining the performance consequences of multinationality, so the explanation of the moderating variable's influence will also differ depending on the theory considered. For example, if one considers product diversification as a moderator, the challenge is to decide which mechanism it changes: product diversification may alter the costs of multinationality, the risks of multinationality, or the benefits from leveraging firm resources across markets; the choice depends on the theory used to explain the baseline multinationality–performance relationship. Moreover, different theories may actually specify opposing main effects in certain cases; in such cases, failure to specify the nature and direction of the main relationship renders any theorizing or interpretation

of the interaction effects ambiguous at best. Such instances may be particularly problematic because even when the direct effect is not statistically significant while the interaction is, the direct relationships must be theoretically justified and described in order for the interaction to make sense.

Second, many manuscripts explain the direct effect of the moderating variable on the dependent variable rather than the impact of the moderating variable on the relationship between the independent and dependent variables. Since manuscripts proposing interaction effects often introduce a new variable, authors often start by defining the new variable and then provide a review of studies that have discussed its impact on the dependent variable; however, these are explanations of a direct effect rather than a moderation effect. In some cases authors end the explanation with a couple of statements along the lines of "since the independent variable has an impact on the dependent variable and the moderating variable also has an impact on the dependent variable, one can conclude that the moderating variable interacts with the independent variable to affect the dependent variable, leading to a hypothesis that argues for the interaction". However, such explanatory strategy does not actually provide an explanation of the interaction per se. A moderator or interaction variable may or may not have an effect on the dependent variable (Carte and Russell 2003); moreover, the independent and moderator variables should not be theoretically related as this would imply mediation (Baron and Kenny 1986). Thus, the arguments for a variable's moderating effect on the main relationship must be distinct from its direct effect on the dependent variable, and if there is a relationship between the moderator and dependent variable, the underlying theoretical mechanism linking them must differ from the theoretical mechanism that influences the main relationship.

Third, some papers face the challenge that although they discuss the relationship between independent and moderating variables on the dependent variable, they do not actually explain the direction of the relationship in the interaction effect. This is problematic especially when the direction of causality can theoretically go both ways. As a result of not specifying theoretically the direction of the relationship, it is unclear which is the main mechanism and which is the interaction effect. A

discussion of how the variables have been found to be related to each other in previous research, or indeed how the interaction among variables has been found to be statistically significant in previous studies, does not qualify as an explanation of causality. Most of the interaction effects are statistically analyzed by simply multiplying the independent and moderating variables and studying how this product term affects the dependent variable. While the empirical results are the same, the theoretical implications of the direction of causality are not equivalent, and conclusions and recommendations drawn from an inaccurate causality relationship may be erroneous.

Recommendations for Explaining Within-Level Interaction Effects

The typical figure explaining a within-level interaction effect appears in Fig. 16.1. This representation is commonly found in manuscripts in which the relationship between the independent variable of interest (*X*) and the dependent variable (*Y*) is argued to be modified by some other variable (*Z*). This simple representation then results in one or two hypotheses being discussed and formally presented in the paper. The first hypothesis tends to be the direct effect, predicting the impact of the independent variable of interest on the dependent variable. The second hypothesis is the moderation effect, predicting the strengthening or weakening of the direct effect under the moderating condition (see Cuervo-Cazurra and Dau 2009, as an illustration).

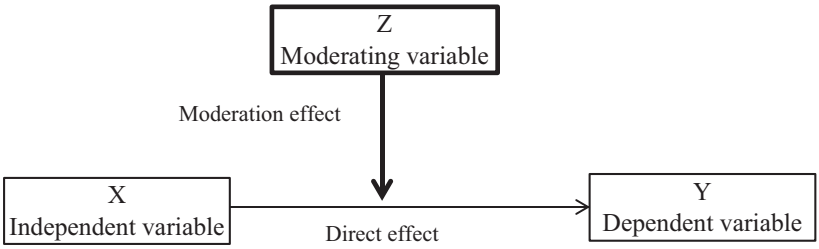

Fig. 16.1 Typical relationships in a within-level moderation model. (Note: The thicker line is the relationship of focus)

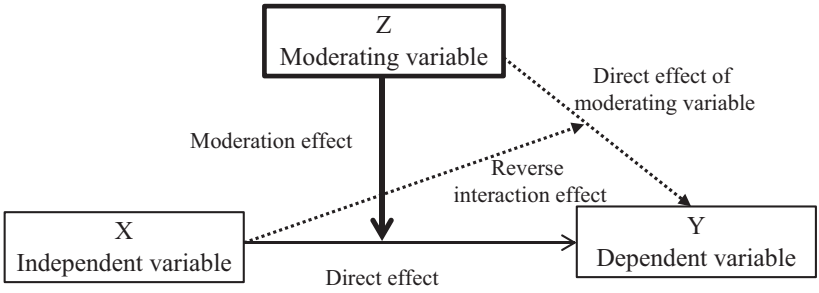

Fig. 16.2 Additional relationships in need of theoretical explanation when analysing a within-level moderation model. (Note: Dashed lines indicate alternative relationships to consider)

Although parsimony can be useful, authors need to be careful and avoid oversimplifying to the point of making simplistic arguments. Fig. 16.2 presents some of the potential additional relationships that may have an influence on the explanation of the moderating relationships and that need to be theoretically addressed: the direct effect of the moderating variable on the dependent variable, the reverse interaction (i.e., the independent variable becoming the moderating variable), and alternative explanations for the moderation effect. Although other relationships, such as mediation and additional exogenous variables, may also have an effect on the moderation, they are beyond the scope of this article.

We now provide a sequence of steps that authors can use to ensure that the interaction effects are clearly explained. In some cases, there is little empirical literature one can use to justify the proposed relationship. In such cases, it is even more important that the moderation effect is clearly explained and that the choice of moderating variables, as well as the proposed nature and effects of these on the direct relationship, is clearly guided by theory. This does not imply that the explanation should not include citations, but rather that the author needs to outline and articulate the underlying theoretical basis of the conceptual mechanisms to explain the interaction. We suggest the following steps to explain within-level moderation:

1. First, identify the theory or theories that are used to explain the direct and moderating effects. Clearly stating the theory used, the logic for using such theory, and an outline of the key arguments and assumptions of the theory not only helps the author clarify the theoretical approach used to build hypotheses, but also helps the reader understand how the author explains the arguments.
2. Second, apply the selected theory to the research question and explain the direct effect and the mechanisms behind it. The explanation of the mechanisms requires statement such as "variable *x* has a positive effect on variable *y*, because …". This does not imply a discussion of how the variables are related, but rather one of why they are related and why the causality goes in a particular direction. If the direct effect has been widely analyzed before and there is a consensus on the relationship from the theoretical standpoint, you may state that this is a well-known argument, and that the direct effect is merely a baseline hypothesis.
3. Third, provide a theoretical justification for the choice of moderator variable. The inclusion of moderating effects in the analysis must be driven by theory rather than by the existence of previous empirical studies that have discussed such interaction, or by the statistical significance of the interaction term in the statistical analysis. The moderating variable establishes conditions under which the direct effect varies, and thus its selection needs to be within the realm of the theory used. Even if you find a statistically significant interaction effect, this does not mean that the moderating variable is theoretically justified; you may be finding such effect because there is mediation or because there is a common determinant (for a discussion, see Frazier et al. 2004).
4. Fourth, explain the direct effect, if any, of the moderator variable on the dependent variable so that it is clear how this direct effect differs from the interaction effect. Although this may be a well-explained relationship, you still need to clarify the mechanisms that lead the moderating variable to affect the dependent variable. As with the direct effect, you may want to present a separate baseline hypothesis if it is relevant. These mechanisms need to differ not only from the mechanisms explaining the interaction effect, but also from the

mechanisms explaining the direct effect of the independent variable on the dependent variable.

5. Fifth, explain how the interaction changes the mechanisms that explain the direct relationship. Using theory, specify arguments such as "the impact of *X* on *Y* is strengthened when *Z* is present because *Z* changes the mechanism in this manner …" or "the influence of *X* on *Y* is reduced in the presence of *Z* because the mechanism is weakened in this way …". Conceptualized as a contingency hypothesis, moderation can be used to examine the boundaries and limitations of a theory (Boyd et al. 2012). In this way, moderation specifies the conditions under which a given theory applies (or not) and thus increases the precision of theoretical predictions (Edwards 2010). Again, make sure that the explanation of the interaction effect differs from the explanation of the direct effect as well as from the explanation of the impact of the moderating variable on the dependent variable.
6. Sixth, theoretically rule out the reverse interaction in which the independent variable *X* is moderating the relationship between the moderating variable *Z* and the dependent variable *Y*. This of course only becomes an issue if a theoretical rationale exists for linking *Z* to *Y*, which we discussed in point four above. The theoretical challenge is to argue that the moderation can only exist in one direction and not the other, for example, because the moderator operates at a different level of analysis or temporally precedes the relationship. Phrase the hypotheses and graph the interactions in a way that is consistent with the theoretically grounded direction of the moderating relationship (Aguinis et al. 2013).
7. Seventh, return to theory when interpreting the results and explain them from a theoretical viewpoint. Rather than state the usual "hypothesis *x* is supported because the coefficient of the interaction term is statistically significant", put far more emphasis on the substantive meaning of such results in terms of our theoretical understanding of the phenomenon under investigation. Specify whether the nature and/or strength of the focal relationship changed as a result of the inclusion of the interaction and how such results inform theory and research moving forward. Non-significant results from the inclusion of moderation effects may also provide useful insights.

Recommendations for Explaining Cross-Level Interaction Effects

Multilevel studies involve relationships between independent and dependent variables at different levels; thus cross-level relationships can be direct and/or moderating. Applying multilevel lenses requires both conceptual and analytical considerations (Snijders and Bosker 2012).

The cross-country nature of IB is particularly ripe for multilevel studies and cross-level interactions (Peterson et al. 2012). Multilevel theorizing provides ample opportunities for cross-fertilization of theories originating from different disciplines; at the same time it requires careful attention to the underlying assumptions of those theories. One typical use of cross-level interaction in IB is the analysis of the impact of country-level variables on firm-level behavior. A baseline hypothesis may be grounded in well-known IB theory, with the contribution to the literature coming in the form of a modification of the expected relationship based on insights from another theory that operates at a different level.

However, simply adding another moderating variable, even at a different level of theory, does not constitute a theoretical contribution per se. While borrowing concepts and variables from different disciplines may yield new insights, an in-depth appreciation for the underlying theory and rigorous integration with IB theory is paramount (Bello and Kostova 2012; Cheng et al. 2009). It is critical to avoid committing the error of "rebottling old wine in new bottles"; that is, selecting variables that have previously been studied in similar settings. Variables and concepts from other levels or disciplines are often part of a system of constructs that together make up a theory; separating (i.e., cherry picking) one or two of these constructs and utilizing them in an isolated fashion as moderators in IB studies violate the underlying coherence of the theory and constructs and may lead to flawed theorizing. Moreover, though drawing on concepts from theories at different levels holds much promise for advancing IB theory, and we certainly promote such theorizing, careful attention must be paid to the ability of theoretical constructs to traverse levels without losing their substantive meaning. Key constructs may be subject

to different meanings at different levels, in different cultures and environments, or between headquarters and subsidiaries.

The prevailing logic in management is that the larger context within which organizations are embedded exerts a greater downward influence than vice versa (Mathieu and Chen 2011). In cross-level research, accurately accounting for the nesting structure is critical because observations within higher level units are more similar than observations across those units; since lower level units share common characteristics and influences from the higher level units, they are not independent from each other. For instance, in IB, subsidiaries are nested within the multinational firm, which is in turn hierarchically nested within its home country. Lower level units may, however, belong simultaneously to multiple higher levels (i.e., industries and countries). To the extent that industries are not country specific but rather global in nature, the resulting structure is non-hierarchical or cross-classified, since each firm uniquely belongs to a combination of both home country and industry levels (for a discussion of different nesting structures in IB research, see Nielsen and Nielsen 2010). In IB, upward cross-level influences can be theorized (e.g., multinational firms influencing a host-country institutional environment via lobbying and non-market strategies); however, such influences are typically main rather than moderating effects. Clear specification of the levels of theory and variables helps conceptualize the nature and direction of cross-level relationships.

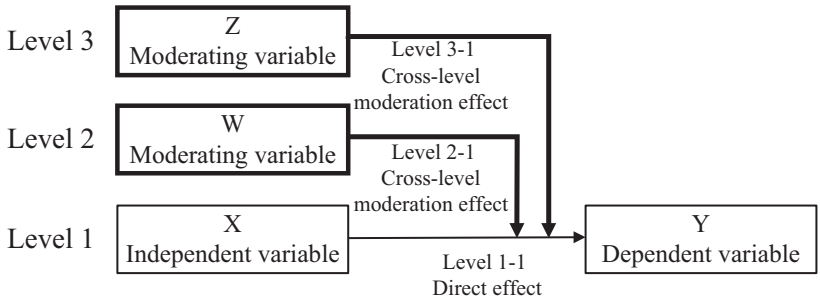

Fig. 16.3 Typical relationships in a cross-level moderation model. (Note: The thicker lines are the relationships of focus)

Figure 16.3 depicts a typical hierarchical nesting structure with three levels (e.g., subsidiaries within the multinational firm, nested within the home country institutional environment), allowing for simple Level 3–1 and Level 2–1 cross-level interactions.

Figure 16.3 is, however, an oversimplification of the reality surrounding cross-level interaction, as several additional relationships must be recognized and discussed. First, Levels 2 and 3 moderator variables (Z, *W)* may also exert direct (downward) influence on the dependent variable (*Y*). Such potential influences must be acknowledged and accounted for both theoretically and empirically. Moreover, the potential for reverse or symmetrical interaction effects should be ruled out. Multilevel modelling can help identify the directionality of the interaction effects in that it is logical that the contextual variable moderates the relationship between lower level variables (Aguinis et al. 2013). However, the theoretical rationale for directionality of interaction effects must still be specified, with constructs and measurement treated accordingly (Klein et al. 1994). Multilevel studies of this kind can develop and test hypotheses pertaining to three types of relationships: (a) lower level direct effects (Level 1–1); (b) cross-level direct effects (Levels 2–1 and 3–1); and (3) cross-level interaction effects (Levels 2–1 and 3–1). Manuscripts with cross-level interactions tend to discuss the direct effect first followed by the moderation of this effect by Levels 2 and 3 cross-level interactions. Fig. 16.4

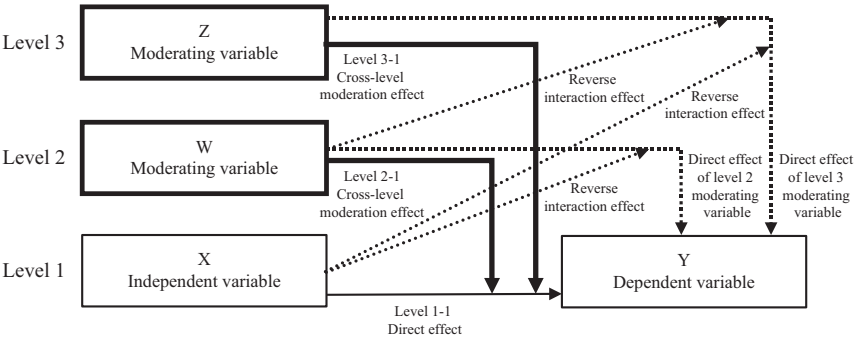

Fig. 16.4 Additional relationship in need of theoretical explanation when analyzing a cross-level moderation model. (Note: Dashed lines indicate alternative relationships to consider)

illustrates these additional relationships associated with cross-level interactions.

Similar to the previous discussion, we offer a set of concrete steps on how to develop theoretical insights for authors considering cross-level interactions:

1. First, specify the focal unit of analysis of the study. This is typically determined by the analytical level of the dependent variable (e.g., firm for studies analyzing firm performance; team for studies analyzing innovation in teams) and represents the level to which generalizations are made. This first step is important, as the unit of analysis determines the appropriate level of associated theoretical constructs and helps avoid misattribution of effects, commonly referred to as "fallacies of the wrong level" (Rousseau 1985: 5).
2. Second, specify the hierarchy and nature of theoretical nesting (e.g., individuals, teams, firm, industry, country, region, etc.). Nesting is important because lower level units share commonalities with higher level units. You need to determine the appropriate levels at which your phenomenon is operating and where you plan to draw the boundaries of the theoretical extension. For example, you may be interested in analyzing subsidiaries, nested within a multinational firm that is in turn headquartered in a particular country. In such a three-level hierarchical model, the multinational firm's headquarter and/or home country characteristics may act as moderators on the relationship between subsidiary characteristics and strategic choice (e.g., Goerzen et al. 2013).
3. Third, choose relevant independent variables from theories at each level of nesting and clearly specify their relationships with the dependent variable (upward or downward). In the selection of variables, take into account how the theory to which they adhere, or its extension, operates at the level of analysis of the focal unit. Explicit integration of theories that span different levels holds great potential for facilitating new theory generation; however, careful attention must be paid to how theoretical constructs operate across levels without losing their substantive meaning. For instance, trust at the institutional level

may mean something different than inter-organizational or interpersonal trust (Nielsen 2010).

4. Fourth, model the within-group variance by specifying the lower level (Level 1–1) direct effects. Identify the theoretical mechanisms explaining these effects and make arguments consistent with the level of theory (i.e., the dependent variable). Direct effects hypotheses at the lower level are typically specified in the same manner as regular hypotheses, using statements like "predictor *X* is positively/negatively associated with outcome *Y*", or "the influence of predictor *X* on outcome *Y* is positive/negative".
5. Fifth, choose relevant moderator variables from theories at higher (lower) levels. If the moderator variable has a relationship with the dependent variable, clearly distinguish the theoretical arguments for the main effect from those for the moderating effect.
6. Sixth, model the between-group variance in intercepts by specifying cross-level (Levels 2–1 and 3–1) direct effects. Identify the theoretical mechanisms explaining these effects and make arguments consistent with downward (upward) direct influences. Cross-level direct effects hypotheses are often specified using statements like: "industry competition negatively influences firm performance", or "host country governance quality is positively associated with non-equity entry mode".
7. Seventh, model the between-group variance in slopes by specifying cross-level (Levels 2–1 and 3–1) interactions. Identify the theoretical mechanisms that explain how and why the nature or strength of the lowest level relationships changes as a function of the higher level moderator. Cross-level interaction effects hypotheses are often specified using statements like "the relationship between firm international diversification and performance varies with home country institutional quality such that firms originating from countries with higher quality institutional environments are more likely to benefit from international diversification than firms originating from countries with lower quality institutional environments".
8. Eighth, rule out reverse interaction of the independent variable on the cross-level direct relationship between the moderator and the dependent variable. The statistical analysis software cannot detect the direction of relationships and theory must guide this choice. Cross-level

interactions typically involve contextual variables at higher levels and this often makes it easier to rule out reverse interaction from a logical standpoint; it is far more likely that industry- or country-level factors moderate the relationship between firm strategy and performance than it is that direct effects of industry or country characteristics vary with one firm's strategy or conduct. Depicting the research model in a figure helps clarify theoretical nesting and the nature and direction of relationships and interactions.

9. Ninth, return to theory when interpreting the results. Multilevel research offers the opportunity to extend theory by bridging or integrating theories from different domains. You need to explain how the cross-level effects (direct or interaction) change our understanding of the theoretical mechanisms that link concepts in a model. It is often useful to examine to what extent the cross-level interactions modify both the theory of the focal unit of analysis and theories at higher levels from which the moderators are drawn.

Conclusions

Interaction effects are increasingly being analyzed in research papers. This is especially the case in IB, because the cross-disciplinary nature of the phenomenon enables researchers to generate new insights by analyzing the boundary conditions of well-known relationships that have hitherto been explained in a domestic setting. Given the theoretical challenges posed by interaction effects, in this editorial we provided a sequence of steps that researchers can follow to explain interaction effects within and across levels of analysis. These steps should be viewed as tools that can be adapted and modified depending on the specific research question and nature of data, rather than strict steps that all submitted papers must follow. The objective of these suggestions is to create papers that provide deeper discussions and extensions of theory. This editorial complements other editorials that have discussed how to develop theory in IB (Bello and Kostova 2012; Cheng et al. 2009; Cuervo-Cazurra et al. 2013;

Thomas et al. 2011) and how to incorporate advanced statistical techniques (Chang et al. 2010; Peterson et al. 2012; Reeb et al. 2012).

Acknowledgements We thank Editor John Cantwell and two anonymous reviewers for useful suggestions for improvement. Alvaro Cuervo-Cazurra thanks the Patrick F. and Helen C. Walsh Research Professorship and the Robert Morrison Research Fellowship at Northeastern University for financial support. All errors are ours.

Note

1. We use the term mechanisms to denote underlying theoretical processes (or reasons) for certain proposed effects. This is different from the use of mechanisms to denote intervening (mediating) variables in a causal chain of relationships (see Baron and Kenny 1986).

References

Aguinis, H., and R.K. Gottfredson. 2010. Best-practice recommendations for estimating interaction effects using moderated multiple regression. *Journal of Organizational Behavior* 31 (60): 776–786.

Aguinis, H., R.J. Boik, and C.A. Pierce. 2001. A generalized solution for approximating the power to detect effects of categorical moderator variables using multiple regression. *Organizational Research Methods* 4 (4): 291–323.

Aguinis, H., R.K. Gottfredson, and S.A. Culpepper. 2013. Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management* 39 (6): 1490–1528.

Baron, R.M., and D.A. Kenny. 1986. The moderator–mediator distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173–1182.

Bello, D.C., and T. Kostova. 2012. From the editors: Conducting high impact international business research: The role of theory. *Journal of International Business Studies* 43 (6): 537–543.

Boyd, B., K.T. Haynes, M.A. Hitt, D.A. Bergh, and D.J. Ketchen. 2012. Contingency hypotheses in strategic management research: Use, disuse or misuse? *Journal of Management* 38 (1): 278–313.

Buckley, P.J., and M. Casson. 1976. *The future of the multinational enterprise*. London: Palgrave Macmillan.

Carte, T.A., and C.J. Russell. 2003. In pursuit of moderation: Nine common errors and their solutions. *MIS Quarterly* 27 (3): 479–501.

Chang, S.J., A. van Witteloostuijn, and L. Eden. 2010. From the editors: Common method variance in international business research. *Journal of International Business Studies* 41 (2): 178–184.

Cheng, J., W. Henisz, and K. Roth. 2009. From the editors: Advancing interdisciplinary research in the field of international business: Prospects, issues and challenges. *Journal of International Business Studies* 40 (7): 1070–1074.

Cohen, J., P. Cohen, S.G. West, and L.S. Aiken. 2003. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed. Mahwah: Erlbaum.

Cuervo-Cazurra, A., and L.A. Dau. 2009. Promarket reforms and firm profitability in developing countries. *Academy of Management Journal* 52 (6): 1348–1368.

Cuervo-Cazurra, A., P. Caligiuri, U. Andersson, and M.Y. Brannen. 2013. From the editors: How to write articles that are relevant to practice. *Journal of International Business Studies* 44 (4): 285–289.

Edwards, J.R. 2010. Reconsidering theoretical progress in organizational and management research. *Organizational Research Methods* 13 (4): 615–619.

Frazier, P.A., A.P. Tix, and K.E. Barron. 2004. Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology* 51 (1): 115–134.

Goerzen, A., C.G. Asmussen, and B.B. Nielsen. 2013. Global cities and multinational enterprise location strategy. *Journal of International Business Studies* 44 (5): 427–450.

Jaccard, J., and R. Turrisi. 2003. *Interaction effects in multiple regression*. 2nd ed. Thousand Oaks: Sage Publications.

Johanson, J., and J. Vahlne. 1977. The internationalization process of the firm: A model of knowledge development and increasing foreign commitments. *Journal of International Business Studies* 8 (1): 23–32.

Klein, K.J., F. Dansereau, and R.J. Hall. 1994. Levels issues in theory development, data collection, and analysis. *Academy of Management Review* 19 (2): 195–229.

Mathieu, J.E., and G. Chen. 2011. The etiology of the multilevel paradigm in management research. *Journal of Management* 37 (2): 610–641.

Nielsen, B.B. 2010. Multilevel issues in strategic alliance research. In *Researching strategic alliances: Emerging perspectives*, ed. T.K. Das, 57–76. New York: Age Publishing.

Nielsen, B.B., and S. Nielsen. 2010. Understanding the internationalization-performance relationship: A multilevel approach. In *Advances in international management*, ed. T. Devinney, T. Pedersen, and L. Tihanyi. New York: Emerald.

Penrose, E.T. 1959. *The theory of the growth of the firm*. Oxford: Oxford University Press.

Peterson, M.F., J.L. Arregle, and X. Martin. 2012. Multilevel models in international business research. *Journal of International Business Studies* 43 (5): 451–457.

Reeb, D., M. Sakakibara, and I.P. Mahmood. 2012. From the editors: Endogeneity in international business research. *Journal of International Business Studies* 43 (3): 211–218.

Rousseau, D.M. 1985. Issues of level on organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior* 7 (1): 1–37.

Shieh, G. 2009. Detecting interaction effects in moderated multiple regression with continuous variables: Power and sample size considerations. *Organizational Research Methods* 12 (3): 510–528.

Snijders, T., and R. Bosker. 2012. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. 2nd ed. Thousand Oaks: Sage.

Thomas, D.C., A. Cuervo-Cazurra, and M.Y. Brannen. 2011. From the editors: Explaining theoretical relationships in international business research: It's about the arrows, NOT the boxes. *Journal of International Business Studies* 42 (9): 1073–1078.

17

Explaining Interaction Effects: A Commentary

Jose Cortina

As many researchers have pointed out, interactions are essential to international business (IB) research. It is in fact rare for authors of IB papers to specify individual- or firm-level effects that *are not* moderated by something, usually a variable at a higher level of analysis. At the same time, explaining and justifying interactions is tricky, especially if one cannot lean upon qualitative differences between moderator groups. It was for this reason that Andersson et al. (2014) offered their editorial.

I was first exposed to this article when I was visiting the University of Sydney shortly after the article had appeared in print. I thought it would be influential, and it has been. In order to better understand its influence, I read many of the papers that have cited it. My reading has led me to three observations.

Andersson, U., A. Cuervo-Cazurra, and B.B. Nielsen. 2014. From the editors: Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies* 45 (9): 1063–1071.

J. Cortina (✉)
Virginia Commonwealth University, Richmond, VA, USA
e-mail: jcortina@vcu.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_17

My first observation is that the primary lessons taken from Andersson et al. (2014) appear to have been that (a) interaction models should begin with the effect of the predictor and, where appropriate, the effect of the moderator (e.g., Domingues et al. 2017; Hollender et al. 2017), and (b) one must rule out the alternative hypothesis that the 'predictor' moderates the relationship between the 'moderator' and the dependent variable (DV) (e.g., García-García et al. 2017; Gray et al. 2015).

Regarding (a), Andersson et al. (2014) noted that authors often leap straight into an explanation of moderation without any consideration of the relationship being moderated. As the authors explain, the reason for this may be that, 'authors think that the novelty of the paper resides in the interaction effect, because the direct effect has been explained in detail and tested before' (2014: 1065). Generally, an interaction exists if the relationship between a predictor and an outcome is moderated by a third variable. It seems reasonable, therefore, to begin by explaining the relationship between the predictor and the outcome. Only when this has been done would one then explain how this relationship varies as a function of the moderator. In addition, one must consider the possibility that the moderator also has an effect on the outcome.

I have often wondered if the confusion around this sort of thing is due to our reliance on so-called conceptual path diagrams. Figure 17.1 in Andersson et al. (2014) is the classic example. There is an arrow leading from the putative predictor to the outcome, and another arrow leading from the moderator to the first arrow. The idea is fairly straightforward—the coefficient connecting X to Y varies as a function of Z. Unfortunately, several things seem to get lost in this sort of diagram. Chief among them

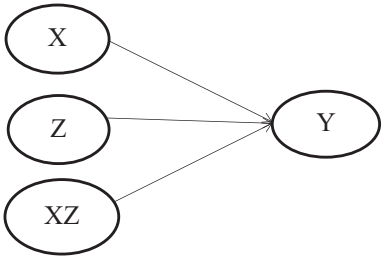

Fig. 17.1 Statistical version of Figure 1 in Andersson et al. (2014)

is the fact that there is no reason why the moderator cannot also have an effect on the DV. A path from the moderator to the DV is often omitted from conceptual diagrams, but it is usually included in so-called statistical diagrams such as the one here.

This is called a statistical diagram because it reflects the model that actually gets tested. The arrow from Z to Y actually has to be included because the coefficient for the product must reflect the rate of change in Y per unit increase in the product holding *both* of its components constant. If Z is not held constant, then the weight for the product is scale dependent. However, this arrow is usually omitted from the conceptual diagram. As a result, the conceptual diagram, which is intended to aid comprehension, actually leads us astray (see Holland et al. 2017, for other examples of diagram-related confusion). If reviewers and editors were in the habit of insisting on either diagrams such as Fig. 17.1 in Andersson et al. (2014) or statistical diagrams, then this first lesson from Andersson et al. (2014) might get absorbed more quickly.

A closer look at (b) led to my second observation. Andersson et al. (2014) suggest that one 'theoretically rule out the reverse interaction in which the independent variable X is moderating the relationship between the moderating variable Z and the dependent variable Y' (2014: 1067). Thus, we see authors ruling out this reverse interaction theoretically, and citing Andersson et al. (2014) along the way. Lesson (b) from Andersson et al. (2014) goes hand in hand with lesson (a). Generally, the logic of moderation arguments goes something like this. 'Previous research has shown that X causes Y. There is reason to believe that the degree to which this is true depends on the level of Z.' Then the reasons are given, and a hypothesis is offered such as, 'The positive relationship between X and Y is stronger when Z is high than when Z is low {synergistic interaction},' or 'The negative relationship between X and Y is weaker when Z is high than when Z is low {buffering interaction}.' This is as it should be. Previous research has shown that X affects Y, but there is reason to believe that this effect is conditional upon Z. Phenomenologically, this is different from saying that the effect of Z on Y is conditional upon X, which is the point being made by Andersson et al. (2014). But here is the catch.

Mathematically, there is no moderator-predictor distinction. B_{xz}, the partial regression coefficient for the product, is interpreted as the rate of

change in the X-Y slope per single point increase in Z. But it can just as easily (and correctly) be interpreted as the rate of change in the Z-Y slope per single point increase in X. For example, many IB papers examine the effect of a firm characteristic X on a firm-level outcome Y as moderated by a country-level variable Z. As a case in point, Lu et al. (2014) argued that firm experience with a given target market is positively related to FDI in that market, and that the degree to which market-supporting institutions are present in that country moderates this relationship. Mathematically, one could also say that the presence of market-supporting institutions affects firm FDI, and that this relationship is moderated by firm experience with that market. However, that is not the point of Lu et al. (2014). Instead, their point is that although previous research has shown experience to be related to FDI, experience is really useful only in countries with weak market-supporting institutions. In such countries, experience is needed to navigate the turbulent waters of investment. Where there are strong market-supporting institutions, this experience is not necessary. The lesson to be taken from Andersson et al. (2014) is that one should make clear why it makes sense to say that the effect of X on Y depends on the level Z *rather than* that the effect of Z on Y depends on the level of X. The lesson is *not* that there is a mathematical difference between these two.

My final observation has to do with next steps. As I mentioned, I first read Andersson et al. (2014) when I was visiting the University of Sydney. It was there that I met Bo Nielsen, who was on a sabbatical at the time. We got talking about his article, and then about how one might offer more concrete advice for authors who were building a case for interaction models. That conversation led to our editorial in *JIBS* (Cortina et al. 2015), which led to a *JOM* article (Cortina et al. in press), and several other articles under review. The point of all of these articles is that a certain type of interaction phenomenon, the restricted variance interaction, is quite common not only in IB, but in cross-level models more generally. It seems to me that what authors really need is a set of conceptual tools that help them to get from a general notion regarding Z moderating the X-Y relationship to a variable-specific justification for a particular interaction pattern. Restricted variance reasoning is one such tool, but there must be others. The Andersson et al. (2014) editorial will, I hope, be a

springboard for such articles. We are already starting to see them (e.g., Buckley et al. 2017), with (hopefully) more to come.

References

Andersson, U., A. Cuervo-Cazurra, and B.B. Nielsen. 2014. From the editors: Explaining interaction effects within and across levels of analysis. *Journal of International Business Studies* 45 (9): 1063–1071.

Buckley, P.J., J.P. Doh, and M.H. Benischke. 2017. Towards a renaissance in international business research? Big questions, grand challenges, and the future of IB scholarship. *Journal of International Business Studies* 48 (9): 1045–1064.

Cortina, J.M., T. Köhler, and B.B. Nielsen. 2015. Restriction of variance interaction effects and their importance for international business research. *Journal of International Business Studies* 46 (8): 879–885.

Cortina, J.M., T. Köhler, K.R. Keeler, and B.B. Nielsen. in press. Restricted variance interaction effects: What they are and why they are your friends. *Journal of Management*: 0149206318770735.

Domingues, J., V.A. Vieira, and R. Agnihotri. 2017. The interactive effects of goal orientation and leadership style on sales performance. *Marketing Letters* 28 (4): 637–649.

García-García, R., E. García-Canal, and M.F. Guillén. 2017. Rapid internationalization and long-term performance: The knowledge link. *Journal of World Business* 52 (1): 97–110.

Gray, S.J., T. Kang, Z. Lin, and Q. Tang. 2015. Earnings management in Europe post IFRS: Do cultural influences persist? *Management International Review* 55 (6): 827–856.

Holland, S.J., D.B. Shore, and J.M. Cortina. 2017. Review and recommendations for integrating mediation and moderation. *Organizational Research Methods* 20 (4): 686–720.

Hollender, L., F.B. Zapkau, and C. Schwens. 2017. SME foreign market entry mode choice and foreign venture performance: The moderating effect of international experience and product adaptation. *International Business Review* 26 (2): 250–263.

Lu, J., X. Liu, M. Wright, and I. Filatotchev. 2014. International experience and FDI location choices of Chinese firms: The moderating effects of home country government support and host country institutions. *Journal of International Business Studies* 45 (4): 428–449.

Part IX

Endogeneity

18

Endogeneity in International Business Research

David Reeb, Mariko Sakakibara, and Ishtiaq P. Mahmood

The Ideal Research Setting

Empirical research in international business (IB) is difficult. Our interests typically centre on whether some particular IB phenomenon causes a specific outcome or effect. We might, for instance, be interested in how expatriate postings influence future career opportunities. Or we might be seeking to understand how firm-level internationalization affects corporate decision-making. In an ideal research setting, to test such a cause and effect, we would examine the impact of firm internationalization on a particular outcome (such as profitability) by randomly assigning some firms to be multinational corporations (MNCs) and other firms to be domestic corporations (DCs). Experimentalists would characterize these as the treatment and control groups. Preferably, we would then observe

D. Reeb (✉) • I. P. Mahmood
National University of Singapore, Singapore, Singapore
e-mail: bizdmr@nus.edu.sg; bizipm@nus.edu.sg

M. Sakakibara
University of California at Los Angeles (UCLA), Los Angeles, CA, USA
e-mail: mariko.sakakibara@anderson.ucla.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_18

and compare the subsequent decision-making of the firms in the treatment and control groups over the next few years regarding the specific variable of interest (i.e., profitability). Inherent in this approach is the notion that we would randomly select the firms to place into the treatment and control groups (i.e., the MNC and DC groupings) in our sample. In general, the iconic test procedure involves developing a random experiment, regardless of whether the unit of analysis centres on individuals, firms, industries or countries.

Unfortunately, in international business research, we are seldom afforded the luxury of a randomized controlled experiment. In addition, in many business situations the treatment may not be a simple binary choice – become an MNC or a DC – but instead may have a continuous element to it that corresponds to firms receiving various doses of internationalization (differing treatment amounts). In the absence of randomized trials with placebos and variable doses, we focus on observational data and use cross-sectional regressions to make inferences about the treatment effect (Angrist and Krueger 2001). Continuing with our MNC vs DC example, a common approach is to estimate the relation between an observed firm characteristic (e.g., profits) and a measure of firm-level internationalization (either as a binary or continuous variable) across a broad sample of firms. Although this approach seems intuitively appealing, it creates an interpretation problem because it is difficult from this test to make causal inferences about the question of interest.

Using Observational Data: The Non-Random Sample

The challenge in using observational data and cross-sectional tests is that the individuals or firms in our treatment and control groups are not randomly selected. More specifically, in the cross-section of firms that we actually observe, firms emerge in distinct organizational and industry patterns. The variable of interest may even influence how firms emerge as multinational or domestic companies (the particular case of reverse causality). For instance, in comparing MNCs and DCs it seems plausible that more profitable firms can afford to develop international operations or that firm internationalization arises due to differences in managerial

experience that also affect firm profitability. This creates a non-random treatment problem, and it is not one that simply inflates the "*t*-statistics." Instead, we obtain inconsistent estimates of the impact of firm internationalization on firm profitability in our regressions, potentially leading to the rejection of true hypotheses or failure to reject false hypotheses (Woolridge 2010). Thus, our empirical tests are distorted, and we may draw the wrong policy implications.

An illustration at the individual level often serves as the best example to highlight this non-random treatment problem. Consider an international business researcher who is interested in testing a program to help facilitate cross-cultural teamwork. For convenience, the researcher provides the training to a group of professors at the university where s/he is employed. One year later s/he observes faculty effectiveness in cross-cultural teams and compares this to cross-cultural team effectiveness in the general population. Specifically, s/he regresses the cross-cultural teamwork effectiveness on faculty appointment and discovers that, consistent with a positive treatment, the university professors in the sample have greater cross-cultural team effectiveness than does the general population. The researcher then reports an effective cross-cultural teamwork effect with the treatment group and concludes that firms should consider approving the training program for workers in their companies.

Clearly, the above test procedure may give the wrong answer to our true question of interest because the university professors may have greater cultural awareness relative to the general population without receiving the treatment. The assignment to the treatment group was not random. Unfortunately, our typical regressions in international business are often even more problematic than this particular example. In this example, we have an idea about the direction of the bias because we have an educated guess about the nature of cultural sensitivity among university professors and the general population. Yet in most international business issues of interest, the direction of the bias is unknown. Moreover, in IB we are rarely able to give the treatments to the subjects even if they are randomly assigned; rather, the individuals or firms themselves often select to take the treatment (or not). In IB studies that do not take into account the non-random assignment problem, we routinely observe that *JIBS* reviewers recommend rejection.

The Prevalence of Non-Random Treatment Problems

Even a cursory glance at real world data indicates that firms do not emerge randomly or uniformly around the world. Similarly, individuals are not randomly assigned postings nor do they uniformly develop managerial expertise. As such, it is difficult to interpret the cross-sectional tests that we commonly employ in IB because our analysis violates the necessary conditions to make them a valid test (Roberts and Whited 2011). Of course we are all familiar with this potential issue, which is often known under the broad title of "endogeneity." A common misconception in the papers we review is that ruling out reverse causality solves the endogeneity problem. Unfortunately, the problem is more pervasive, and reverse causality is only one distinct case of the non-random treatment effect. Statisticians and econometricians have been discussing the issue for decades, and over the past several years their remedies have become quite common in empirical business research. At the *Journal of International Business Studies (JIBS)* we find that the most successful studies in IB use the intuition and insights behind these methods in developing their research design to facilitate causal inference from their observational data.

In IB research the objective usually centres on providing evidence about the causal effects of some particular IB phenomenon. Because this research usually involves observational data, rather than random trials, the relevant goal in IB research design centres on the development of a test that best approximates a controlled experiment (Angrist and Krueger 2001). As a result, studies that explicitly identify the source of variability in the dependent variable can develop appropriate tests that improve the researcher's ability to make causal inferences (Angrist and Pischke 2008). This issue applies to a variety of approaches in IB research, not just the examples used for illustration in this essay. Research on the determinants of multinationality or studies that use data items such as individual patents also face this same endogeneity problem. Brenner (2011) exemplifies this approach to careful research design at *JIBS* in his study determining if resource advantages cause managers in firms with illegal international activity to cooperate with government prosecutions.

Approximating the Randomized-Controlled Experiment

This section provides a brief (and hopefully intuitive) explanation of some of the statistical remedies that IB scholars use in their cross-sectional tests. These short descriptions of several common methods for dealing with the non-random treatment problem are not the main focus of this essay (however, Roberts and Whited 2011 provide a thorough analysis). Rather our emphasis is on the importance of careful research design that incorporates field research or institutional knowledge to develop tests with observational data to facilitate causal inference.

We primarily focus on non-random treatment (endogenous binary variable) rather than the continuous case endogeneity because it provides an intuitive framework for discussing strategies to exploit variation in the main independent variable to develop testable hypotheses. In this context, we believe that discussing some of the potential remedies may help researchers who submit articles to *JIBS* identify the manner in which their observational data can be used to approximate a randomized controlled experiment. In sum, we seek to highlight the notion that IB research that recognizes the variability in the casual relationship and clearly identifies the strategy being used to approximate a controlled experiment has the best chance of success in the *JIBS* review process.

Control Variables and Fixed Effects

Theoretical predictions in international business research are often direct and straightforward, suggesting that internationalization causes some activity to occur. The simplest test in this circumstance is to focus on univariate statistical differences between the groups of interest (i.e., MNCs vs DCs). Yet we all appreciate that we must control for other individual or firm attributes to properly gauge the relation of interest. At the most basic level this occurs because we do not have randomized controlled experiments. In essence, the inclusion of control variables in a multivariate regression is an attempt to deal with the nonrandom nature of the treatment effect in our analysis. Unfortunately, in many

circumstances, this control variable approach is insufficient to deal with the non-random treatment effect problems that we encounter. Potential sources of problems include the omission of some important variables, reverse causality, and measurement error in the variables of interest (Roberts and Whited 2011). Thus, it appears to reviewers that this empirical approach is chosen because of its ease of use rather than because it emerged as a well-designed strategy to make the tests more like an experiment.

As the non-random treatment problem has been recognized for decades, several statistical techniques have been developed and included in standard statistical software packages to address these concerns (e.g., in STATA, SPSS). Perhaps one of the earliest empirical approaches to dealing with endogeneity is mechanical in nature, namely including unit-level fixed effects in the regression (Woolridge 2010). Unit-level fixed effects, such as firm fixed effects, are well suited for circumstances with panel data and are strongly endorsed in many econometric textbooks (e.g., Greene 2008). This approach essentially includes a dummy variable for each individual or firm and relies on changes of the causal variable within a given individual or firm.

Although fixed effects are easy to implement, their ability to effectively curb the non-random treatment problem depends on the nature of the endogeneity problem. Business researchers often find that in a dynamic setting where firm characteristics slowly change over time, the use of fixed-effects removes the theoretical, cross-sectional variation of interest (Zhou 2001).[1] The implication is that it can be difficult to find a meaningful relationship between the causal variable and the outcome variable with fixed effects, even if one truly exists. Evidence of causal relation with unit-level fixed effects can be quite compelling even though it may be difficult to interpret a lack of evidence.

Matching and Propensity Score Models

The matched sample approach essentially attempts to address the non-random treatment effect by creating a pseudo random sample. In many international business situations the most obvious approach to matching

centres on firm size or industry in order to develop a sample where the treated and untreated firms overlap. Cassiman and Golovko (2011), for instance, use a matching model framework to address endogeneity in a *JIBS* study on innovation and exports. At the individual level, matching on education and experience represent common approaches. Intuitively, matching is a method to add control variables and allow the treatment effect to differ across firm type. Matching achieves this goal by eliminating firms from either the treatment or control group that do not have comparable firms in the other group and therefore minimize extrapolation (Angrist and Pischke 2008). The cost of this improved estimation in terms of robustness is that such analysis is less generalizable to the broader universe.

In recent years, an approach labelled as "propensity score matching" has gained popularity because it allows a refined matching process along multiple individual or firm characteristics (Dehejia and Wahba 1999). In the effort to create a matched sample in a study on MNCs for instance, researchers may attempt to effectively randomize the data by matching MNCs to DCs along several different dimensions such as total assets, industry, ownership structure, analyst following, and so forth. This particular approach of matching often uses a logit or probit model with the variable of interest (i.e., propensity to become an MNC) as the dependent variable. The researcher then matches MNCs to DCs based on their predicted propensity to become MNCs. Often these propensity score models use one-to-one firm matching and attempt to match firms on their predicted values (Caliendo and Kopeinig 2008). Although one-to-one matching exemplifies the spirit behind matching, alternative propensity score approaches such as one-to-many, kernel matching and reseeding may also be relevant.

Implementing a propensity score model with a binary treatment is straightforward. The first step is to predict the variable of interest for each individual or firm (i.e., predict their likelihood of becoming an MNC) based on multiple individual or firm characteristics. Second, using the predicted value for the variable of interest (i.e., chance of becoming an MNC) match individuals or firms with high and low values of the variable of interest (i.e., MNCs to DCs). Third, test the original equation of interest (i.e., profits in MNCs and DCs) using only the individuals or

firms in the matched sample. In essence, this approach attempts to correct for the non-random treatment effect by matching a treated firm (or person) to an untreated firm which has similar characteristics.

Using our cross-cultural training example from earlier, each of our treated faculty members would be matched to someone in the general population, with similar age, gender, education, activity levels, marital status and so forth. Although this may not solve the non-treatment problem, it can potentially mitigate some of the associated problems. A limitation of the matching approach is that for a given propensity score, one might have a lot of the treated firms (e.g., the MNCs) but only a very few of the counter-factual firms to be matched (e.g., the DCs), making one-to-one matching difficult. A major strength of the matching approach is that it obliges us to explicitly identity the non-random component of the treatment effect and to determine the appropriate counter-factual firms (Heinrich et al. 2010 provide a primer on using propensity score matching).

Instrumental Variable Approach

Another popular approach to dealing with endogeneity is to seek an exogenous proxy for the treatment or independent variable of interest (Larcker and Rusticus 2010). This classic approach centres on finding a variable, called an instrument, which influences the independent variable (the right-hand-side variable) but appears unlikely to affect the dependent variable (the left-hand-side variable) except through its effect on the independent variable (Wintoki et al. 2012). Cull et al. (2011) provide an example in *JIBS* of using an instrumental variable approach in their analysis on related lending and the development of banking systems.

Focusing again on our cross-cultural example, an instrument would need to be something that is significantly related to the likelihood of being in the treatment group (i.e., related to being a university professor) but unlikely to be related to cultural sensitivity. For instance, a sudden and unexpected increase in the job market opportunities in the year a person received their graduate degree might be related to the decision to become a university professor but unrelated to the cultural sensitivity

which often starts earlier in life. We then use this "instrument" to predict the treatment effect and use this predicted variable in the test of interest.

Ideally, an instrument should affect the main dependent variable through a single channel and in a single direction (Angrist and Krueger 2001). Unfortunately, exogenous instruments are rare and difficult to find. However, as the instrumental variable is part of the standard toolkit of many business scholars, we often see attempts to use some other firm choice variable as an "instrument." In a high percentage of the empirical papers, it appears that the chosen instrument(s) often turn out to be some other endogenous variable(s). It is common, for instance, to see leverage or firm size used as instruments, when these are obviously related to the dependent variable. This approach is usually justified by pointing to some other articles that also choose to use this particular endogenous instrument. Murray (2006) provides a detailed discussion of the problem with invalid instruments. Ultimately, the instrumental variable approach depends on the quality of the instrument being used.[2] Larcker and Rusticus (2010) provide a step-by-step guide to using instrumental variables.

Natural Experiments

Another approach to dealing with endogeneity centres on evaluating the variable of interest after some shock, such as the death of a CEO, a natural disaster, or a regulatory change. Using a specific intervention, such as a change in regulation, can be thought of as natural or quasi experiment (Bertrand et al. 2004). The natural experiment approach uses the regulatory change as the treatment effect and allows the same firm or individual to be analyzed before and after the shock. To implement this approach one computes the difference between the variable of interest before the shock and after the shock in each firm affected by the shock or regulatory change. Of course other things may be changing as well, so ideally we would like another set of firms or individuals that did not receive a shock to use as a control group. We then can compare the difference in the shock group to the difference in the non-shock group over the same time period. This difference-in-difference test provides a robust environment

for evaluating cause and effect. The effectiveness of this approach depends on the exogeneity of the shock. For example, if a group of firms lobby to induce a regulatory change, then this regulatory change cannot really be considered an exogenous shock for these firms. In contrast, unexpected events like financial or political crisis can provide ideal test environments, especially when the shock and non-shock groups are similar along other firm or individual characteristics.

As an illustration, consider a researcher concerned about the impact of taxes on the investment strategies of multinational firms. A country changes its tax code in such way that taxes are increased for repatriated income, which primarily affects multinational firms. We then compare the investments by each MNC before and after the tax change. This difference provides an estimate of the effect of taxes on investments. Of course other issues in the economy may influence investments so we can compute this same difference in investments for domestic firms (who were unaffected by the change in the law). Computing the difference in these differences then provides a strong test of the effect of taxes on the investment decisions of multinational firms.

Regression Discontinuity Design

Another emerging technique for dealing with the non-random treatment effect centres on an approach labeled as regression discontinuity design (Lee and Lemieux 2010). This method attempts to alleviate concerns about the non-random treatment effect by exploiting how people or firms become part of the treatment group. This approach focuses on identifying an observable characteristic that defines how someone or some firm becomes part of the treatment group and seeks to exploit the cutoff point. Essentially, the regression discontinuity method seeks to utilize the similarities of those individuals/firms just above and just below the cutoff point (Almond and Doyle 2011). Thus, in a regression discontinuity design we would seek to compare firms (or persons) who were just above the cut-off point (became part of the treatment effect) to those who were just below the cut-off point.

As an example, assume we wish to compare the value of an expatriate posting, relative to a similar posting in a domestic subsidiary, on managerial career advancement in a sample of Finnish managers. Comparing the

post placement salaries of the managers in both subsidiaries will provide a biased (upward) estimate of the foreign subsidiary posting because they are likely to be assigned to better managers. Even in the absence of the posting in the foreign subsidiary, the expat manager would, on average, likely earn a high wage in the future. To illustrate the process, assume that managers are assigned to subsidiary postings based on their IQ and that those with an IQ above 160 receive the expatriate posting and those between 140 and 160 receive assignments in the domestic subsidiary (with mean IQ of 100 and standard deviation of 15). Then, even though the managers are not randomly assigned to subsidiary postings, we may be able to extract the expatriate treatment effect because IQ data is available for male Finnish citizens as part of their compulsory military service.[3] Presumably the ability of those with an IQ of 159 does not differ that much from those with an IQ of 161.

To evaluate the value of an expat posting we might regress post assignment pay on an indicator variable for expat posting by using the subset of the managers with IQs between 158 and 163. The counter-factual or control group is comprised of the managers with 158–160 IQs, while the treatment group is comprised of the expat managers with IQs between 161 and 163. The difference in pay between these two similar groups will be captured by the coefficient estimate on expatriate posting. In a sense, this approach endeavours to randomise the treatment group in a similar spirit to the propensity score model by suggesting the appropriate control and treatment groups are those on either side of the cut-off point. As such, this approach also represents a subset analysis.

A Guidepost to Research Design: The Role of Theory

The empirical approaches that seek to analyze data using standard regressions with matched samples, instrumental variables, natural experiments or regression discontinuity designs are valid and relevant. Unfortunately, we frequently use these statistical techniques as crutches or substitutes for critically thinking about the problem of interest, resulting in dubious analyses (Thomas et al. 2011). In our haste to discover the truth, we often

seek to let the data speak by running regressions and then fashioning a story around the results. Yet this approach intensifies and exacerbates the non-random treatment problem, leading to deceptive results and improper policy implications. The systematic manner in which the underlying data emerge needs to be considered before the first test is performed (Heckman and Urzua 2009). Thus, the first step in developing our hypotheses is to identify how firms/individuals are assigned to the treatment group and why this assignment occurs (Roberts and Whited 2011).

While this first step sounds simple, we often find this critical step is skipped in the papers submitted to *JIBS*. Take an example of a study of the relationship between a country's legal system and the behavior of firms in a country (e.g., propensity of foreign direct investment or FDI) based on the panel data of multiple countries and firms. A country's legal system is determined by its resources, history, culture, industry structure and so on, and it does not change quickly. Therefore the legal system affects firm behavior, not the other way around, and hypotheses should be developed in that direction. In the long run, however, the firms in a country (or even foreign firms) can affect the country's legal system. For example, more profitable firms might demand stricter intellectual property rights protection and their taxes may help fund the legal system, so one might be interested in investigating this kind of phenomenon. If one wants to study how firm behavior affects legal systems, then the potential for non-random legal institution assignment should be fully examined and incorporated in hypothesis development. In this case, one cannot delegate the examination of potential reverse causality (i.e., legal system affects firm behavior) to statistical tests.

A Systematic Approach

One formal approach to dealing with the nonrandom treatment effect centres on developing a structural model. Structural models provide rigid and explicit equations of individual or firm behavior that rely on idealistic assumptions.[4] Although structural models are often couched in technical jargon, the intuition behind using them suggests a simple framework

for developing the theoretical underpinnings of the eventual empirical specification. Fundamentally, one should think about how the observed variations in the right-hand-side variable of interest may have emerged. The scholar's institutional knowledge and ideas about how the treatment decision emerges are critical components to sound empirical research design (Angrist and Pischke 2008). As we are unable to randomly assign firms into the treatment/control groups, understanding how the firms were initially assigned to the treatment or control group is essential to developing testable hypotheses.

Consider an example regarding the determinants of FDI. An IB scholar might be interested in evaluating the idea that firm-level FDI is driven by firms seeking to find low cost employees. Accounts in the business press routinely describe investment in China and job migration of the US in this fashion. One approach to test this idea would be to compare FDI within a country, across different states/provinces, based on the average wage rate. Alternatively, one could make the same sort of comparison across multiple countries or geographic regions. A typical premise to test this maintained hypothesis might be: FDI is negatively related to wages. Basile (2004) provides such a test in the context of FDI across Italy using foreign acquisitions. Specifically, evidence is found to suggest foreign direct investment is positively related to wage rates. One might be tempted to conclude that our theoretical prediction was incorrect; instead we found that firms were targeting FDI in provinces with high wages.

Yet a finding of a positive relation between wages and FDI may stem from the non-random sample that we used. Recall that our tests are based on the premise that we randomly assigned high wages to some countries/provinces and low wages to others. Ideally, the wage rate is supposed to be randomly assigned across countries to generate a reliable test. Of course, wage rates are not randomly distributed across countries/provinces but instead may arise due to differences in human and physical capital. Thus, our hypothesis and research design needs to incorporate the notion that wage rates are not exogenous. For instance, it may be that wage rates are a function of education and experience, suggesting that low wage environments may have limited human capital. In terms of identifying the relevant control group this non-random component needs to be incorporated into the hypothesis development.

Focusing on the theoretical development in this FDI story, our hypotheses need to explicitly acknowledge that FDI should be negatively related to wages for a given level of human capital. This type of hypothesis might naturally lead to the construction of a propensity score matched sample of workers from high and low wage provinces with similar levels of human capital in order to identify the wage rate differential. Ultimately, none of the procedures developed by econometricians are magic pills (Roberts and Whited 2011). Instead, they all highlight the need for careful theoretical development that leads to the proper identification of the relevant control group as the best alternative to randomized controlled experiments.

Research Implications

The success of empirical business research over the past three decades is based on simple, straightforward theories that provide qualitative predictions and conform to observed real world phenomena. Yet we still need to incorporate the notion that the independent variable is unlikely to be randomly distributed across firms. Generally speaking, the theoretical predictions that we develop should incorporate, by design, the nonrandom component of our right-hand-side variable. Reinterpreting Marschak (1953), it is not necessary or desirable to fully specify a structural model of the dependent variable, but one does need to consider the fundamental economic issues that lead to nonrandom assignments of the treatment effect among the firms in the sample. In the absence of a randomized controlled experiment, we need to incorporate the non-random assignment into the treatment group in our research design to improve causal inference.

More fundamentally, the strength of IB research depends on the ability to identify the main theoretical mechanisms by which the dependent variable arises. These mechanisms can be identified using both qualitative and quantitative methods. One approach that seems to be gaining ground in more recent years centres on combining field research (anecdotal or systematic qualitative) and quantitative analysis. Because this approach relies on insights from the insiders – managers and employees of the firm – to inform econometric analysis, it is sometimes known as "insider

research" (Ichniowski and Shaw 2009). In insider research, the rich micro data collected through field interviews help identify the behavioral mechanisms that explain how the treatment (for instance, a certain type of management work practice) affects firm performance including productivity and profitability (Siegel and Larson 2009). Similarly, speaking with professionals about the nature of causality for researchers using secondary data can also be useful.

By helping the researcher to model the adoption of treatments more accurately, insider research helps identify any selection bias in the estimation of the treatment effect. The key issues though centre on developing a strong theoretical argument for how a phenomenon causes a particular effect and how that phenomenon emerged among the observations in the sample.

Across multiple international business subfields, we find that researchers who carefully consider how a phenomenon arose in the cross section are often the most successful at *JIBS*. Following these examples we encourage researchers to speak to managers, market participants, bankers or consultants in the area to obtain the institutional details that are crucial to understanding the nature of causality in a particular phenomenon. The acid test is whether the research design in an empirical study with observational data is the one that best approximates a randomized-controlled experiment for the hypothesis of interest.

Notes

1. Specifically, the fixed effect absorbs the time invariant characteristics of the firm, which mitigates endogeneity but also reduces our ability to study the effects of some time-invariant variables of interest.
2. Using our cross-cultural training example again, a poor choice for predicting treatment might be religious tolerance. Although religious tolerance might be related to university employment, it may also be related to cross-cultural sensitivity.
3. The subsidiary assignment could be done with some other set of observable criteria such as employee rankings or assessments. IQ, however, provides a useful illustration due to its fine gradation and familiarity to

academics. Hoekstra (2009) provides an example of using regression discontinuity design to evaluate flagship university attendance on salaries.

4. Consequently, most business questions do not lend themselves to using formal quantitative structural models that focus on complex predictions (see Welch 2010).

References

Almond, D., and J.J. Doyle. 2011. After midnight: A regression discontinuity design in length of postpartum hospital stays. *American Economic Journal: Economic Policy* 3 (3): 1–34.

Angrist, J., and A. Krueger. 2001. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives* 15 (4): 69–85.

Angrist, J., and J. Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

Basile, R. 2004. Acquisitions versus greenfield investment: The location of foreign manufacturers in Italy. *Regional Science and Urban Economics* 34 (1): 3–25.

Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119 (1): 249–275.

Brenner, S. 2011. Self-disclosure at international cartels. *Journal of International Business Studies* 42 (2): 221–234.

Caliendo, M., and S. Kopeinig. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys* 22 (1): 31–72.

Cassiman, B., and E. Golovko. 2011. Innovation and internationalization through exports. *Journal of International Business Studies* 42 (1): 56–75.

Cull, R., S. Haber, and M. Imai. 2011. Related lending and banking development. *Journal of International Business Studies* 42 (3): 406–426.

Dehejia, R., and S. Wahba. 1999. Causal effects in nonexperimental studies: Re-evaluating the evaluation of training programs. *Journal of the American Statistical Association* 94 (448): 1053–1062.

Greene, W. 2008. *Econometric analysis*. 6th ed. Upper Saddle River: Prentice Hall.

Heckman, J., and S. Urzua. 2009. *Comparing IV and structural models: What simple IV can and cannot identify*. Working Paper, National Bureau of Economic Research.

Heinrich, C., A. Maffioli, and G. Vazquez. 2010. *A primer for applying propensity-score matching*, Technical Note IDB-TN-161. Washington, DC: Inter-American Development Bank.

Hoekstra, M. 2009. The effect of attending the flagship state university on earnings: A discontinuity based approach. *Review of Economics and Statistics* 91 (4): 717–724.

Ichniowski, C., and K. Shaw. 2009. *Insider econometrics: Empirical studies of how management matters*. Working paper 15618. National Bureau of Economic Research.

Larcker, D., and T. Rusticus. 2010. On the use of instrumental variables in accounting research. *Journal of Accounting and Economics* 49 (3): 186–205.

Lee, D., and T. Lemieux. 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48 (2): 281–355.

Marschak, J. 1953. Economic measurements for policy and prediction. In *Studies in econometric method*, ed. W. Hood and T. Koopmans, 1–26. New York: Wiley.

Murray, M. 2006. Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives* 20 (4): 111–132.

Roberts, M., and T. Whited. 2011. *Endogeneity in corporate finance.* Working Paper, Wharton.

Siegel, J., and B.Z. Larson. 2009. Labor market institutions and global strategic adaptation: Evidence from Lincoln Electric. *Management Science* 55 (9): 1527–1546.

Thomas, D.C., A. Cuervo-Cazurra, and M.Y. Brannen. 2011. From the editors: Explaining theoretical relationships in international business research: Focusing on the arrows, NOT the boxes. *Journal of International Business Studies* 42 (9): 1073–1078.

Welch, I. 2010. *A critique of quantitative structural models in corporate finance.* Working paper, Brown University.

Wintoki, M., J. Linck, and J. Netter. 2012. Endogeneity and the dynamics of internal corporate governance. *Journal of Financial Economics* 105 (3): 581–606.

Woolridge, J. 2010. *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, MA: MIT Press.

Zhou, X. 2001. Understanding the determinants of managerial ownership and the link between ownership and performance: Comment. *Journal of Financial Economics* 62 (3): 559–571.

19

Endogeneity in International Business Research: A Commentary

J. Myles Shaver

Introduction

As Reeb, Sakakibara and Mahmood (2012)—hereafter RSM—highlight in their essay, explicitly considering endogeneity is important for empirical studies in international business (IB) because of the types of questions that we address and the nature of the data that we can access to answer these questions. To demonstrate the importance of this point, RSM provide a number of conceptual examples. More compelling than these examples, however, are studies showing that carefully accounting for endogeneity changes the conclusions of previous IB research (e.g., Shaver 1998; Berry and Kaul 2016).

In this commentary, I hope to provide three insights to complement the material that RSM present. To do this, it is important to focus on the

Reeb, D., M. Sakakibara, and I.P. Mahmood. 2012. Endogeneity in international business research. *Journal of International Business Studies* 43 (3): 211–218.

J. M. Shaver (✉)
University of Minnesota, Minneapolis, MN, USA
e-mail: mshaver@umn.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_19

main issue that underlies RSM's focus on endogeneity—the desire to establish causality in IB research. Let me start by being clear what it means—in a practical sense—to try and establish causality in applied empirical research (i.e., establishing casual identification).[1]

The Practical Process of Establishing Causal Identification

Establishing causal identification means trying to isolate the underlying theoretical mechanism that causes the relationship between a treatment and outcome. Practically, this means trying to establish that the observed relationship between variables is not spurious and not the result of a competing causal theoretical mechanism (e.g., Cuervo-Cazurra et al. 2016).

In order to effectively do this, it is important that theory guides our research—as RSM note. However, being precise about how theory leads to a hypothesized cause and effect is only a starting point. What is also important is that we expend effort to conceptually understand what might be among the host of other explanations or alternative theories that would lead to the same relationship between the variables we study. Let me demonstrate why.

As RSM note, the use of panel data and fixed-effects estimators is a viable strategy to help establish causal identification. The benefit of this approach, over a cross-sectional regression, is that the interpretation of coefficient estimates from a fixed-effect estimator is the partial correlation of changes in the X variable and changes in the Y variable. This compares to the interpretation of coefficient estimates in a cross-sectional regression, which is the partial correlation between the level of X and the level of Y. The fixed-effect estimator, therefore, will rule out many alternative explanations that (i) would lead to an association between the level of X and the level of Y and (ii) are not related to a cause and effect mechanism of X on Y. The reason is that if X causes Y, then changing X must be associated with a change in Y.

Nevertheless, ruling out such alternative explanations does not preclude that an alternative causal mechanism could lead to the partial

correlation of changes in X and changes in Y. For example, an unmeasured variable might have a causal relationship with Y. If this variable and X happen to change at the same time, then we would observe a partial correlation between changes in X and changes in Y that is spurious.

Therefore, having a well-articulated theory of why X affects Y *is only a starting point*. At some point, focusing on causal identification means that we must not only generate precise theories, but that we must also critically think of alternative theories that would lead to the same empirical relationship of our preferred theory. In other words, just because we have not thought hard or precisely about alterative causal mechanisms to a finding that we have in the literature does not preclude that these unknown effects underlie the relationships we observe. This also means that the process in establishing causal identification requires that we do not support every theoretical mechanism that we advance.

Research Design Strategies and Causal Identification

Another admirable aspect of RSM's discussion is that they highlight several research design strategies to account for endogeneity that can aid in establishing causal identification.[2] In addition to accessible descriptions of each approach, the authors also present its constraints or limitations. This, however, leads to the following question. What does it mean when there are multiple research design approaches to establish causal identification when none of them offers a perfect solution?

Steps towards accounting for endogeneity are exactly that—steps. Each of the research design strategies that RSM present are important ways to advance our understanding of causal identification because they aid ruling out plausible alternative explanations with specific characteristics. At the same time, none is exhaustive, and all possess a set of underlying assumptions. It is important for scholars to understand and acknowledge the limitations of the approaches that they take.

As I make the case in detail elsewhere (Shaver forthcoming), when there exist multiple approaches to establishing causal identification and

when all have their limitations, it is important that we establish causal identification through a cumulative body of research. What I mean by this is that we should not expect any one study to provide unequivocal conclusions with respect to causal identification. Rather, to reach unequivocal conclusions will often require multiple studies, focusing on the same question and theory, and using a variety of research design strategies.

This does not call into question any of the recommendations that RSM provide. Instead, it suggests that tackling the same question with different approaches will be important to advance understanding in our field. It also means that we must welcome empirical contributions that advance identification as a meaningful contribution to the literature. We should view causal identification as a research process.

Advances in Data Availability and Causal Identification

The points that RSM make in their article are equally valid as we see advances in data availability. The reason is that the problem of causal identification is not a problem of lack of data.[3] It is a problem of the nature of the data. 'Big data' that are not collected with an eye to research design and causal identification will suffer the same concerns that RSM highlight.

Conclusion

To conclude, I wish to highlight three points to complement RSM's essay:

- Alternative approaches to establish causal identification have their limitations. Therefore, the way in which we need to establish causal identification is through a cumulative body of research—which will require a plurality of approaches.

- Although establishing causal identification requires that we develop well-crafted theories, it also requires that we expend effort to advance well-crafted alternative theoretical mechanisms.
- The difficulty of establishing causal identification tends not be a problem of too little data; it is a problem of the nature of the data that we collect.

Notes

1. Across social science disciplines, many authors comment on the importance and research strategies to establish causal identification—including economics (e.g., Angrist and Pischke, 2010), political science (e.g., Samii 2016), sociology (e.g., Gangl 2010), psychology (e.g., Rohrer, 2018), and management and organizations (e.g., Shaver forthcoming).
2. Shaver (forthcoming) notes that research design choices as described in RSM are only a subset of tools we have to help establish causal identification. Other tools include measurement and interpretation.
3. Endogeneity biases are not small-sample biases—they hold asymptotically.

References

Angrist, J.D., and J.S. Pischke. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives* 24 (2): 3–30.

Berry, H., and A. Kaul. 2016. Replicating the multinationality-performance relationship: Is there an S-curve? *Strategic Management Journal* 37: 2275–2290.

Cuervo-Cazurra, A., U. Andersson, M.Y. Brannen, B.B. Nielsen, and A.R. Reuber. 2016. Can I trust your findings? Ruling out alternative explanations in international business research. *Journal of International Business Studies* 47 (8): 881–897.

Gangl, M. 2010. Casual inference in sociological research. *Annual Review of Sociology* 36: 21–47.

Reeb, D., M. Sakakibara, and I.P. Mahmood. 2012. Endogeneity in international business research. *Journal of International Business Studies* 43 (3): 211–218.

Rohrer, J. M. 2018. Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science* 1 (1): 27–42.

Samii, C. 2016. Causal empiricism in quantitative research. *The Journal of Politics* 78 (3): 941–955.

Shaver, J.M. 1998. Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science* 44 (4): 571–585.

———. forthcoming. Causal identification through a cumulative body of research in the study of strategy and organizations. *Journal of Management*. https://doi.org/10.1177/0149206319846272.

Part X

Common Method Variance

20

Common Method Variance in International Business Research

Sea-Jin Chang, Arjen van Witteloostuijn, and Lorraine Eden

What Is Common Method Variance?

A large number of papers submitted to *JIBS* use data collected from a survey instrument. When self-report questionnaires are used to collect data at the same time from the same participants, common method variance (CMV) may be a concern. This concern is strongest when both the dependent and focal explanatory variables are perceptual measures derived from the same respondent (Podsakoff and Organ 1986).

S.-J. Chang
National University of Singapore, Singapore, Singapore
e-mail: schang@nus.edu.sg

A. van Witteloostuijn
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
e-mail: a.van.witteloostuijn@vu.nl

L. Eden (✉)
Department of Management, Texas A&M University, College Station, TX, USA
e-mail: leden@tamu.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_20

CMV is "variance that is attributable to the measurement method rather than to the constructs the measures represent" (Podsakoff et al. 2003: 879). CMV creates a false internal consistency, that is, an apparent correlation among variables generated by their common source. For example, this could occur if a researcher asks respondents to evaluate an MNE's organizational capabilities and the firm's international performance in the same survey. In such cases, self-report data can create false correlations if the respondents have a propensity to provide consistent answers to survey questions that are otherwise not related. Thus, common methods can cause systematic measurement errors that either inflate or deflate the observed relationships between constructs, generating both Type I and Type II errors.

Scholarly views of CMV differ. Campbell (1982: 692), a former editor of the *Journal of Applied Psychology*, provides a strongly negative assessment (italics added): "If there is no evident construct validity for the questionnaire measure or *no variables that are measured independently of the questionnaire*, I am biased against the study and believe that it contributes very little". On the other hand, some scholars argue that the CMV problem may be overstated (Crampton and Wagner 1994; Lindell and Whitney 2001; Spector 1987), and even an "urban legend" (Spector 2006). A recent exhaustive review of research on CMV in behavioral research reaches a more balanced conclusion: "common method variance is often a problem and researchers need to do whatever they can to control for it" (Podsakoff et al. 2003: 900).

The specific details of the research methodology are clearly relevant in determining the likelihood and degree of common method bias. Podsakoff et al. (2003) explore four general sources of CMV: the use of a common rater, the manner in which items are presented to respondents, the context in which items on a questionnaire are placed, and the contextual influences (time, location and media) used to measure the constructs.[1] Some CMV sources may be more problematic than others, for example, perceptual data from single raters may be more worrisome than the manner in which items are presented in the survey instrument. A manuscript that suffers from more potential sources of CMV should, in general, be more problematic than one with fewer sources. The most worrisome example of CMV, according to the authors (2003: 885), occurs when "the data for both the predictor and criterion variable are obtained from

the same person in the same measurement context using the same item context and similar item characteristics".

The *JIBS* editors see CMV as a potentially serious concern for researchers using survey-based data,[2] especially where the dependent and independent variables are perceptual and from the same source.[3] *JIBS* editors and reviewers must therefore assess not only whether CMV exists, but also how likely it is to generate Type I and Type II errors. Our position is to inform potential contributors that they should, prior to journal submission, deal with potential CMV biases in their research both seriously and explicitly.

How Can CMV Be Addressed by Researchers?

In general, four approaches have been recommended in the literature as methods that researchers should use to avoid or correct CMV (see, for example, Podsakoff et al. 2003):

1. The obvious strategy is, of course, to avoid any potential CMV in the research design stage by using other sources of information for some of the key measures. In particular, if possible, the dependent variable should be constructed using information from different sources than the independent variables.
2. A number of procedural remedies in designing and administering the questionnaire, from mixing the order of the questions to using different scale types, can reduce the likelihood of CMV.
3. Complicated specifications of regression models reduce the likelihood of CMV. Specifically, respondents are unlikely to be guided by a cognitive map that includes difficult-to-visualize interaction and nonlinear effects. This is less likely the more complicated the model.
4. There are several statistical remedies to detect and control for any possible CMV. A post hoc Harman one-factor analysis is often used to check whether variance in the data can be largely attributed to a single factor. Additionally, other statistical procedures can be applied to partial out common factors or to control for them.

Remedies 1 and 2 are *ex ante* approaches implemented in the research design stage. Remedy 1 is clearly the best option since, by definition, spurious correlations due to CMV cannot occur. Remedy 2 can also *ex ante* reduce the likelihood of the consistency motive and theory-in-use biases in the informant responses. Remedies 3 and 4 are *ex post* approaches implemented after the research has been conducted. Remedy 3 implies that the likelihood of CMV can be reduced by specifying complex relationships that are unlikely to be part of the respondents' cognitive maps, while remedy 4 corrects for CMV through a variety of statistical procedures.

We first discuss these four basic remedies below and then offer our advice to international business scholars on how to handle the possibility of CMV in their research.

Remedies 1 and 2: Avoid CMV in the *ex ante* Research Design Stage

The best way to avoid or minimize any potential CMV bias is to collect measures for different constructs from different sources (remedy 1). Ideally, the dependent variable(s) are collected from a different source than the independent variables are collected from.[4] For example, subsidiary-level autonomy or decision-making power of a subsidiary-level manager could be measured with a local survey, whereas a subsidiary performance measure could be constructed from information provided by the multinational headquarters or from a local archival source. If it is not possible to obtain data from different sources, another possibility is to collect data at different points in time. Preferably, all this is part of the *ex ante* research design. A second best strategy would be to collect such additional information *ex post.*

Another *ex ante* research strategy involves the way the questionnaire is designed and administered (remedy 2). Respondents should be assured of the anonymity and confidentiality of the study, that there are no right or wrong answers, and that they should answer as honestly as possible.[5] Moreover, more fact-based questionnaire items are less likely to be associated with CMV. Podsakoff et al. (2003, 888) state that "these procedures should reduce people's evaluation apprehension and

make them less likely to edit their responses to be more socially desirable, lenient, acquiescent and consistent with how the researcher wants them to respond". Additionally, great care must be taken to systematically examine the construction of items so as to ensure that ambiguous, vague and unfamiliar terms are not included, and that the questionnaire as a whole and the individual items are formulated as concisely as possible (see Harrison et al. 1996; Lindell and Whitney 2001; Podsakoff et al. 2003). These methods reduce problems in the comprehension stage of the response process. Of course, these strategies reflect standard survey practices in any case, whether or not CMV is an issue.

Specifically related to CMV, Podsakoff et al. (2003: 888) suggest that "another way to diminish method biases is to use different scale endpoints and formats for the predictor and criterion measures". This should reduce method biases caused by commonalities in scale endpoints and anchor effects. In a recently published *JIBS* article on the international location decision, for instance, Galan et al. (2007) measured the dependent variable in two different ways so as to avoid potential CMV with independent variables that were operationalized as perception-based measures.

Moreover, counterbalancing the order of questions relating to different scales and constructs makes CMV less likely, as the respondent cannot then easily combine related items to cognitively "create" the correlation needed to produce a CMV-biased pattern of responses (Murray et al. 2005). One option is to randomize the order of the questions using survey software. For instance, in a *JIBS* article examining the influence of parent control on conflict in international joint ventures, Barden et al. (2005) used different response anchors across measured constructs, separated measurements in time, and manipulated the order of questionnaire items in such a way that CMV across dependent, independent and control variables became very unlikely.

Remedies 3 and 4: Deal with CMV in the *ex post* Statistical Analyses

CMV is more likely to emerge in models that are overly simple. A third approach is therefore to specify relationships among the dependent and independent variables that are not so simple that these relationships are

likely to be part of the individual raters' cognitive maps. In this context, Harrison et al. (1996: 248) refer to the cognitive miser principle. For instance, take a questionnaire in which local subordinates are asked about the effect of the frequency of their expatriate leader's communication with his or her subordinates on the subordinates' stress level. Suppose that the theory predicts that this effect is non-linearly moderated by the subordinates' locus of control trait. Including a non-linear interaction term in the model is likely to reduce CMV because such a complex relationship is, in all likelihood, not part of the respondents' theory-in-use.

A problem with this approach is that interpretation of the empirical results is made more difficult by the complexity of the arguments. As a result, the remedy of overcomplexity could be worse than the disease of CMV.[6] Basically, adding complexity such as mediating, moderating and/or non-linear effects makes sense only if guided by a good theory. In the end, sound theory that directs design and method is, of course, the bottom line that characterizes all good research, be it survey-based or not.

The fourth remedy is to apply *ex post* statistical approaches. Indeed, there are quite a few of them; here, we only briefly refer to some of the more popular ones since there are several other papers with more details (please refer to the references attached to this Letter). Perhaps the most common but ineffective response by authors to address CMV (other than ignoring it) is to rely on Harman's single-factor test to assert that their research is not pervasively affected by CMV. This method loads all items from each of the constructs into an exploratory factor analysis to see whether one single factor does emerge or whether one general factor does account for a majority of the covariance between the measures; if not, the claim is that CMV is not a pervasive issue. However, Podsakoff et al. (2003) explain that this claim is likely to be incomplete because Harman's test is insensitive. It is unlikely that a single-factor model will fit the data, and there is no useful guideline as to what would be the acceptable percentage of explained variance of a single-factor model. The *JIBS* team therefore believes that simply reporting seemingly reassuring outcomes from Harman's single-factor test is insufficient to prove that CMV is not a pervasive issue.

Lindell and Whitney (2001), Podsakoff et al. (2003) and Malhotra et al. (2006) review several statistical methods that are more sophisticated

than Harman's test, which can be used to test and possibly control for CMV. Different statistical remedies are available for different types of research settings and different sources of CMV. Promising statistical remedies include a partial correlation procedure and a direct measure of a latent common method factor. The former method partials out the first unrotated factor from the exploratory factor analysis, and then continues to determine whether the theoretical relationships among the variables of interest do still hold. The latter method allows questionnaire items to load on their theoretical constructs, as well as on a latent CMV factor, and examines the significance of theoretical constructs with or without the common factor method. Both methods have their own limitations, however, one of which is the assumption that the sources of CMV can be well identified and validly measured.[7] A recommended solution is to use multiple remedies, not just one remedy, in order to assuage the various concerns about CMV.

Could CMV Be a Problem in *JIBS*?

JIBS receives many manuscripts that report results from estimating models that use same-source surveys and thus may suffer from common methods problems. However, while psychology journals have worried about CMV since the early 1980s, and have implemented techniques to handle this problem since at least the mid-1990s, the common methods issue has been slow to be recognized, and is much less addressed, by international business scholars. Best practices in the "macro" disciplines appear to have lagged behind those in the "micro" disciplines, probably because the micro areas (e.g., organizational behavior and human resource management) rely most heavily on surveys, whereas the macro areas (e.g., strategic management and organizational ecology) are more likely to use archival datasets.

In March 2009, we reviewed all the articles published in *JIBS* between 2000 and the present for evidence of potential sources of CMV. Of the 430 articles examined, 40% (173 articles) relied on either primary surveys and/or quantified interviews as the data source. The 173 articles were then characterized by potential sources of common methods bias.

Almost all the articles (167 articles) contained one or more sources of CMV, and most appeared to have multiple sources. Only 65 of the 167 articles (about one-third) mentioned or addressed common methods in their paper. Of this group, half the articles (32) used Harman's single-factor test or something similar to test for CMV. Fifteen articles used another approach. Only nine articles used both Harman's test and at least one other correction method to control for CMV.

These statistics, of course, tell us only how frequently common methods appear in recently published *JIBS* articles, not the magnitude of the potential bias from CMV in these articles. Previous research estimating the magnitude of the effects did not include *JIBS* articles; see for example, Doty and Glick (1998) and Cote and Buckley (1987). So, the most we can say is there may be a problem based on frequency of usage of common methods, but at present we have no estimates of the magnitude of the problem.

Based on this short survey, it appears that common method bias has not been recognized nor addressed by most IB scholars, even in *JIBS*, the top journal in the field of international business. We recognize, of course, that standards for rigor in empirical work are continually rising. What were acceptable methodological practices even five years ago can easily and rapidly become unacceptable as social science scholars better understand the limitations of their empirical techniques and develop more rigorous methods for identifying and correcting for potential biases in their work. The purpose of our Letter from the Editors is therefore not to criticize earlier research, but rather to encourage IB scholars to implement current best practices in research methods. We argue that the hurdle barrier must now be set higher in *JIBS vis á vis* CMV. It is time for IB scholars to address, and reduce or offset where feasible, the use of common methods in their empirical work.

Recommendations

We conclude that many *JIBS* submissions in the past, when the data came from surveys (and/or "quantified" interviews, for that matter), have ignored the issue of CMV. Where the authors have addressed CMV, the

typical response has been to report seemingly reassuring results from Harman's single-factor test.

The *JIBS* editors believe this approach has been insufficient. For the current editorial team, it is now standard practice to return a manuscript to the author when it appears to suffer from common method bias and the issue has been ignored in the manuscript. The desk rejection letter asks the author to perform validity checks and resolve any CMV issues before resubmitting the manuscript.

Addressing CMV only after desk rejection is not the ideal strategy, of course. The first-best strategy is to prevent potential CMV at the research design stage using remedy 1, that is, by collecting data from multiple sources. *Ex ante*, before running any analyses, the collection of key information from other sources should be planned, using where possible archival data and multiple respondents. Alternatively, additional information can be collected afterwards. An example of remedy 1 is Carraher et al. (2008), who surveyed expatriate employees but obtained their expat performance measure from company records filed by the individual's supervisor rather than from the expat respondents themselves.

In addition, we recommend that the survey questionnaire be carefully designed, applying all or a large subset of the procedural remedies listed above (remedy 2). Depending upon the nature of the questionnaire, tailor-made CMV measures can be included, a well-known example being social desirability scales. *Ex post*, in the empirical stage, options are to run Harman's single-factor test, specify a complex model (including interaction and non-linear terms), and explicitly control for or partial out CMV statistically (remedies 3 and 4). The options are summarized in Fig. 20.1.

More often than not, a perfect solution is out of reach. If *ex ante* methods are not doable, the *JIBS* editors recommend that IB scholars use multiple *ex post* procedural remedies including possibly a more complex model specification, and partialing out or controlling for CMV (remedies 2, 3 and 4). What we ask for is that CMV-related methodological issues should be discussed carefully and explicitly in any manuscript submitted to *JIBS* that uses single-respondent data. This will often imply the need to apply a number of the remedies referred to above. While the problems with CMV were not well understood by IB scholars in the past

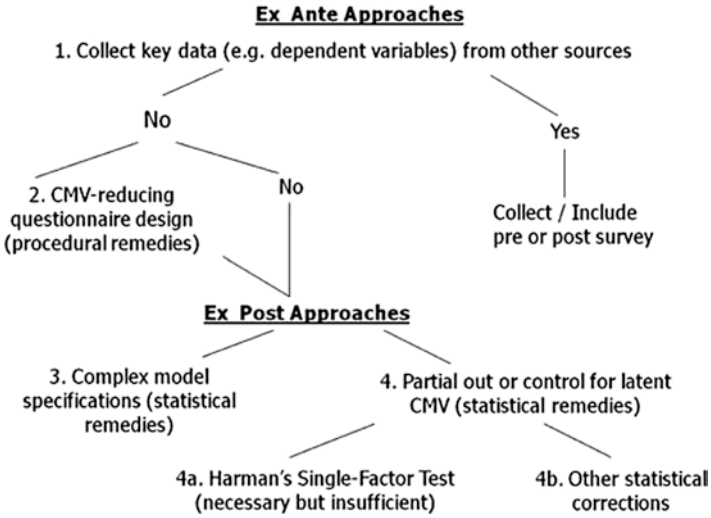

Fig. 20.1 Approaches for handling common method variance (CMV)

and, as a result, there were many *JIBS* articles published that might suffer from multiple sources of common methods bias, the standards have changed and IB scholars must adopt current best practices.

Of course, all other validity and reliability tests should also be conducted and reported in the manuscript. In singling out CMV in this Letter from the Editors we do not want to give *JIBS* authors the impression that they should ignore other validity and reliability tests of their quantitative and qualitative research methods – they should not. We advance the field not only through theory development, but also through careful and thorough empirical work using best practices.

As international business researchers, we also recognize that sometimes common methods cannot be totally avoided, for example, if the research probes into difficult waters where data of any kind are scarce such as in severely understudied parts of the world (Africa, the Middle East), or where the research undertaken is so novel or insightful that this may be considered over standard methodological considerations. Editors and reviewers should not reject innovative manuscripts that push the boundaries of our knowledge of international business solely on the grounds of common methods. In such situations, the gains in creativity and impact might outweigh the loss of methodological purity.

Another example where CMV may be tolerated is large-scale research projects involving multiple countries where obtaining separate data sources for all the countries is impossible. Such large-scale projects typically deal with potential national differences in response bias by using within-subject standardization, mean centering the scores of individuals, or using regression approaches to control for national differences in response biases. These studies also typically link some sort of aggregate nation scores to a separate source archival predictor or criterion. Cross-country comparative studies where the intent of the study is to treat cognitive structures and processes of a population as a dependent variable might well want to use same-source correlations as a good indicator of shared aspects of cognition, with a national indicator as a separate-source predictor of these same-source correlations.

Lastly, we want to make it clear to the *JIBS* community that it is not our intent in this Letter from the Editors to privilege IB researchers who use large, readily available datasets such as Compustat, Orbis or SDC Platinum. Rather, we want to encourage primary and qualitative research in international business – including surveys – but at the same time increase awareness among IB researchers of potential CMV biases so that they can be avoided in the design stage. Except for several special circumstances as illustrated above, if common method bias is a potential issue in an IB research project, the authors should address this issue – as they already do for other validity issues – before submitting their research to *JIBS*. We are looking for more careful examination from all IB researchers in proving the robustness of their results, and hope this Letter from the Editors will help raise the quality standards for empirical research by future IB scholars.

Acknowledgements This letter reflects the opinion of the *JIBS* editors. We thank Wendy Boswell and Duane Ireland for helpful early discussions about CMV. We particularly appreciate the empirical work by Li Dai (PhD student at the Department of Management at Texas A&M), who browsed through many issues of JIBS in search of CMV best practices, and Les Graham (Durham Business School, UK), who provided an excellent overview of the literature. We gratefully acknowledge the excellent suggestions made by other *JIBS* editors and members of the *JIBS* Consulting Editors and Editorial Review Board. This letter went through two rounds of single blind review. In total, we received about 30

reports from editorial board reviewers, many of them providing multiple pages of detailed comments. We tried to accommodate as many comments as possible, within the constraints of the Letter format. In particular, we thank Laurence Booth, Roger Calantone, Dave Gilliland, Klaus Meyer, Mark Peterson, and Arndt Sorge for extensive comments; any remaining errors or omissions remain the responsibility of the authors. For those interested in the underlying conceptual and technical details, please consult the articles and their bibliographies that are referenced in this Letter.

Notes

1. For example, a researcher may be interested in a hypothesized relationship between constructs A and B (for example, A can be an organizational characteristic of an MNE and B the firm's international performance). If the measures of A and B are derived from information provided by a single rater within each MNE, aggregated across multiple MNEs, there may be spurious systematic correlations between these two measures. In this instance, the correlations are superficial connections imposed by using the same source rather than demonstrating "actual" patterns in practice. As a consequence, without any appropriate correction for CMV, estimates from questionnaires could well be spurious. The estimates may not reveal any real underlying theoretical relationship, but rather the artificial cognitive maps of reality that are hidden in the respondents' minds. Consequently, the reported analyses are likely to suffer from Type I and Type II errors.
2. CMV is by no means a problem only for primary researchers. Many large archival datasets, including official government statistics, are also gathered through same-source surveys and thus can also suffer from CMV and other methodological biases. Thus, researchers using archival data sources should also be aware of potential biases and address them in a proper manner.
3. Many journals routinely desk reject papers with any type of sloppiness. For good advice on this, albeit from a different discipline, see http://www.math.ucla.edu/~tao/submissions.html.
4. Method 1, however, faces a potential ethical barrier if the survey respondents were anonymous since survey data (e.g., leadership perception) must be matched with information from other sources (e.g., employee

appraisal) that could compromise the respondents' anonymity. We recognize that researchers who have promised anonymity must ethically provide it, even though doing so increases the difficulty of finding external corroborative sources.

5. Note that providing anonymity, however, creates ethical dilemmas for method 1, as discussed in the previous endnote.
6. See, for instance, McClelland and Judd (1993), who point out difficulties in detecting and interpreting interactions and moderator effects.
7. Note that a complete identification of all sources of CMV is not necessary. For instance, partial correlation adjustment works if a variable theoretically unrelated with others, preferably the dependent variable, can be identified and used in the adjustment (Lindell and Whitney 2001; Malhotra et al. 2006).

References

Barden, J.Q., H.K. Steensma, and M.A. Lyles. 2005. The influence of parent control structure on parent conflict in Vietnamese international joint ventures: An organizational justice-based contingency approach. *Journal of International Business Studies* 36 (2): 156–174.

Campbell, J. 1982. Editorial: Some remarks from the outgoing editor. *Journal of Applied Psychology* 67 (6): 691–700.

Carraher, S.M., S.E. Sullivan, and M.M. Crocitto. 2008. Mentoring across global boundaries: An empirical examination of home- and host-country mentors on expatriate career outcomes. *Journal of International Business Studies* 39 (8): 1310–1326.

Cote, J.A., and M.R. Buckley. 1987. Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research* 24 (3): 315–318.

Crampton, S., and J. Wagner. 1994. Percept-percept inflation in microorganizational research: An investigation of prevalence and effect. *Journal of Applied Psychology* 79 (1): 67–76.

Doty, D.H., and W.H. Glick. 1998. Common methods bias: Does common methods variance really bias results? *Organizational Research Methods* 1 (4): 374–406.

Galan, J.I., J. González-Benito, and J.A. Zuñiga-Vincente. 2007. Factors determining the location decisions of Spanish MNEs: An analysis based on the

investment development path. *Journal of International Business Studies* 38 (6): 975–997.

Harrison, D.A., M.E. McLaughlin, and T.M. Coalter. 1996. Context, cognition, and common method variance: Psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision Processes* 68 (3): 246–261.

Lindell, M.K., and D.J. Whitney. 2001. Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology* 86 (1): 114–121.

Malhotra, N.K., S.S. Kim, and A. Patil. 2006. Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science* 52 (12): 1865–1883.

McClelland, G.H., and C.M. Judd. 1993. Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin* 114 (2): 376–390.

Murray, J.Y., M. Kotabe, and J.N. Zhou. 2005. Strategic alliance-based sourcing and market performance: Evidence from foreign firms operating in China. *Journal of International Business Studies* 36 (2): 187–208.

Podsakoff, P.M., and D.W. Organ. 1986. Self-reports in organizational research: Problems and prospects. *Journal of Management* 12 (4): 531–544.

Podsakoff, P.M., S.B. MacKenzie, J.Y. Lee, and N.P. Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88 (5): 879–903.

Spector, P.E. 1987. Method variance as an artifact in selfreported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology* 72 (3): 438–443.

———. 2006. Method variance in organizational research: Truth or urban legend? *Organizational Research Methods* 9 (2): 221–232.

21

Common Method Variance in International Business Research: A Commentary

D. Harold Doty and Marina Astakhova

Introduction

Common method variance (CMV) is a topic that has displayed remarkable staying power in the social sciences. The topic has its roots in the early work on construct validity such as Cronbach and Meehl's 1955 article. The dominant analytical framework is Campbell and Fisk's (1959) articulation of the multi-trait multi-method matrix. For perspective, these early references indicate that social scientists have been dealing with the problem of CMV since before most of our current social scientists were born!

In the late 1980s interest in the topic was reinvigorated when Spector analysed ten published studies, reported little if any evidence of significant

Chang, S.-J., A. van Wittleloostuijn, and L. Eden. 2010. From the editors: Common method variance in international business research. *Journal of International Business Studies* 41 (2): 178–184.

D. H. Doty • M. Astakhova (✉)
The University of Texas at Tyler, Tyler, TX, USA
e-mail: hdoty@uttyler.edu; mastakhova@uttyler.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_21

methods variance, and concluded "that the problem may in fact be mythical" (1987: 442). Needless to say, this unexpected conclusion triggered a plethora of competing studies (Cote and Buckley 1987; Doty and Glick 1988; Williams et al. 1989). Summarizing across these studies it is fair to conclude that while Spector (1987) was perhaps optimistic, his conclusion was not without merit. In the intervening years the awareness of CMV has diffused broadly across the organizational literatures (e.g., Podsakoff et al. 2003; Sharma et al. 2009).

Recently Chang, van Witteloostuijn, and Eden (2010) published an excellent editorial letter in *JIBS* emphasizing the importance of CMV to international researchers. They provide a short and concise overview of the CMV problem, summarize a set of ex ante design remedies to avoid CMV, and detail ex post strategies that incorporate both conceptual and statistical approaches for assessing CMV. Their cautionary discussion of the Harmon one-factor test reminds us to think more deeply and critically about common practices for assessing the threat of CMV and to implement more recent and sophisticated techniques for addressing it.

Considering the maturity of the field and the efficiency of Chang and associates' work, contributing to this literature in a short commentary is a daunting task. We organize our effort to address this task around four questions triggered by the *JIBS* editorial letter. The first three questions focus on measurement and data concerns. The fourth question asks how reviewers might use information already available in the study's results to evaluate the threat of CMV.

Does Common Method Variance Matter?

In the current context this question may seem absurd (Davis 1971). However, there is a growing body of literature that validates our question. For example, although Doty and Glick (1998) report that CMV results in an average 26% inflation of observed correlations they argue that this level of bias may not be sufficient to invalidate research findings. Spector (2006: 222) concludes that CMV has attained the status of an urban legend in that it "is based on truth, but has been distorted and exaggerated […] over time." Fuller and colleagues report that "at the typical level

of reported scale reliabilities, correlations do not exhibit inflating CMB until a high level of CMV exists – at least 60% common variance depending on the measurement scale" (2016: 3195). Lance and associates state that the inflationary effect of CMV "is almost completely offset by the attenuating effect of measurement error" (2010: 435).

These conclusions seem more reasonable when the problem of CMV is expressed in classical test theory (Nunnally 1967). In the context of CMV, the foundational equation of classical test theory, Eq. 21.1, is revised to include systematic method error as presented in Eq. 21.2:

$$\text{Test score} = \text{true score} + \text{random error}. \tag{21.1}$$

$$\text{Test score} = \text{true score} + \text{random error} + \text{systematic error}. \tag{21.2}$$

The systematic error term in Eq. 21.2 is assumed to be methods error, which is systematic variation in the measure of a construct that is attributable to the measurement technique. CMV is "systematic error variance shared among variables measured with and introduced as a function of the same method and/or source" (Richardson et al. 2009: 763). Most of the conversation about CMV focuses only on how the methods error component inflates the observed relationship between constructs causing researchers to risk falsely rejecting the null hypothesis (Type I error).

A more complete incorporation of classical test theory, however, requires that we also acknowledge the random error components of measures that attenuate observed relationships. This poses an interesting paradox – the lower the reliability of self-report survey measures the lower the potential for CMV to bias results (Fuller et al. 2016; Lance et al. 2010). In fact, limited amounts of CMV may offset the attenuation caused by random error, increase the accuracy of the observed correlations between constructs, and reduce the risk of Type II error.

While we would never argue that CMV is good, we do suggest that low levels of CMV do not jeopardize the validity of a study. To be clear – scholars should not embrace measures with low reliability to avoid CMV. However, the review process should incorporate a balanced perspective about the countervailing effects of random error and CMV. Scholars should also continue extending the work of Fuller et al.

(2016) and Lance et al. (2010) by investigating the offsetting effects of random error and methods error.

How Important Is Uncommon Method Variance?

A second critical question focuses on a recent advancement, uncommon method variance (UMV), which is defined as extraneous unshared influences on measures that attenuate observed correlations (Williams and Brown 1994; Spector et al. 2019). The arguments presenting UMV deny the loosely held assumption embedded in Eq. 21.2 that the systematic error term is monolithic.

Spector and associates (2019) argue that there may be many different sources of systematic error that could contaminate individual measure-construct pairings in unique ways. For example, some constructs are more subject to social desirability than others. If social desirability introduces systematic error into the measure of one construct but not into the measure of a second construct, then the observed correlation between the constructs will be attenuated. Similarly, when two constructs are measured with different methods and each method introduces unique systematic error variance into each measure-construct pair, the observed correlation will be attenuated. Thus, much like random error, UMV deflates the magnitude of observed relationships and increases the risk of Type II error. Unlike random error, however, UMV is systematic error that cannot be detected as unreliability using traditional methods such as Cronbach's alpha.

At the current time, the magnitude of the UMV problem is simply unknown. What is clear, however, is that multi-method data are not a panacea to achieve data that are free of systematic error. Each data source, and even each method-construct paring, may be as likely to suffer from CMV as from UMV. How these different sources of systematic error variance combine with random error variance to bias observed relationships requires extensive investigation.

Potential Systematic Error in Emerging Data Sources

Outsourcing data through Amazon's Mechanical Turk, Qualtrics panels, StudyResponse, and other open recruitment methods is on the rise in IB research (e.g., Grinstein and Riefler 2015; Shaffer et al. 2016). Such emerging methods allow relatively quick and inexpensive data collection that can be tailored to the research question and study design. Participants are typically unsupervised and anonymous, have completed surveys in unknown locations, and are highly motivated by financial incentives. "Outsmarting" such participants through *ex ante* techniques to reduce CMV is challenging because participants are quick learners and are well aware of social desirability and attention checks in surveys (Hauser and Schwarz 2016).

Another emerging trend is the use of Big Data to create and capture value for individuals, businesses, communities, and governments (George et al. 2014). We are aware of no systematic effort or investigations that evaluate Big Data sources for either random or systemic error. Given the newness and undefined nature of Big Data, we can, at best, speculate about some potential relationships with CMV and UMV. For example, a considerable number of data points are typically "objective measures" (e.g., number of sales, mobile transactions, yearly profit) that are more concrete, and thus should be less subject to either CMV or UMV. Big Data are generated from an increasing plurality of sources including internet clicks, user-generated content, and social media. These different methods render a single-source monomethod concern moot but increase the threat from UMV. The high volume and large sample size associated with Big Data may demand a switch from *p*-values that will almost always be significant to basic Bayesian statistics or even more complex analyses, such as genetic algorithms, natural language processing, or neural networks. These complex models should reduce the threat of CMV without increasing the threat of UMV.

What's a Reviewer to Do?

Our final question shifts the focus of our comments from authors to reviewers. We agree wholeheartedly with Chang and associates (2010:182) that "[e]ditors and reviewers should not reject innovative manuscripts that push the bounds of our knowledge of international business solely on the grounds of common methods."

Reviewers and editors tend to express scepticism regarding research that makes use of same-source, self-reported data because they believe that such data are unable to provide accurate parameter estimates of inter-construct relationships. Such scepticism is justified. However, rather than automatically dismissing such manuscripts, we suggest reviewers use other information available in the manuscript to evaluate the extent to which CMV is a concern. We offer four points as initial considerations for reviewers. Our intent is not to suggest a methodology for reaching definitive conclusions about a study's validity, but rather to expand the CMV-related evidence reviewers might consider.

Is a Single Source or Self-Report Design Mandated by the Constructs or Theory?

Before dismissing studies employing self-report measures reviewers should consider the extent to which the theoretical constructs require such measures. When research questions involve internal states, managerial cognitions, or participants' perceptions, self-report measures may be the most theoretically appropriate measurement approach. For example, Su and associates (2009) appropriately employed self-report measures in their study of channel member's (i.e., supplier and retailer) perceived dependence.

Some Study Constructs Are More Subject to Bias Than Others

In general, abstract constructs that are difficult or ambiguous to answer are likely to be more susceptible to CMV than more concrete constructs (Cote and Buckley 1987; Doty and Glick 1998; Malhotra et al. 2006). When answering questions measuring abstract constructs respondents are more likely to introduce bias by subjectively interpreting questions or using implicit theories to guide their responses. In contrast, measures of more concrete and observable constructs are subject to less bias. Thus, reviewers can use the content of the constructs to help judge the potential for biased results.

Scale Reliabilities are Informative As previously discussed, scale reliabilities influence the degree to which CMV biases. "Low reliability" data (reliabilities ranging between 70 and 80) may potentially mitigate CMV concerns (Fuller et al. 2016). Thus, unless the reliabilities reported in the study are unusually high, reviewers might conclude that the likelihood of biased observed correlations is not too great.

Inter-Construct Correlations Are Informative

The zero-order correlation matrix is also informative. For example, a moderate number of small positive and negative correlations suggests that the study is not contaminated by sufficient levels of CMV to cause unanticipated strong correlations across the nomological network. Reviewers should further consider whether the larger nomological network seems to make sense. Are measures that should be unrelated unrelated? Are measures that should be moderately related moderately related?

Conclusion

In closing, we acknowledge that we have at times taken a somewhat contrarian perspective – a tactic we employed to stimulate continuing interest in a topic that was already a concern before either of us was born. To aspiring authors, we caution that such a perspective might not be beneficial in either their manuscripts or in their responses to reviewers. We suggest following the advice and strategies to avoid CMV outlined by authors such as Chang, van Witteloostuijn and Eden (2010), Podsakoff and associates (2003; 2012), and many others. For reviewers we have attempted to balance the perspective offered by Campbell (1982: 692) who stated that "[i]f there is no evident construct validity for the questionnaire measure or no variables that are measured independently of the questionnaire, I am biased against the study and believe that it contributes very little." As a field, we should begin the difficult task of developing decision guidelines that help reviewers determine when the threat of CMV is too great. In the role of reviewer it continues to be fairly easy to think up all the many ways that CMV might creep into data. It is much more difficult to assess whether the threat of CMV is sufficiently severe that the conclusions of the manuscript are no longer valid and, as a consequence, the manuscript should be rejected.

References

Campbell, J. 1982. Editorial: Some remarks from the outgoing editor. *Journal of Applied Psychology* 67 (6): 691–700.

Campbell, D.T., and D.W. Fisk. 1959. Convergent and discriminant validation by the multi-trait – Multi-method matrix. *Psychological Bulletin* 56 (2): 81–105.

Chang, S.-J., A. van Wittleloostuijn, and L. Eden. 2010. From the editors: Common method variance in international business research. *Journal of International Business Studies* 41 (2): 178–184.

Cote, J.A., and R. Buckley. 1987. Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research* 24 (3): 315–318.

Cronbach, L.J., and P.C. Meehl. 1955. Construct validity in psychological tests. *Psychological Bulletin* 52 (4): 281–302.

Davis, M.S. 1971. That's interesting: Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences* 1 (2): 390–344.

Doty, D.H., and W.H. Glick. 1988. *Method variance in I/O research major effect or mythical beast?* Dallas: Society for Industrial and Organizational Psychology Conference.

———. 1998. Common methods bias: Does common methods variance really bias results? *Organizational Research Methods* 1 (4): 374–406.

Fuller, C.M., M.J. Simmering, G. Atinc, Y. Atinc, and B.J. Babin. 2016. Common methods variance detection in business research. *Journal of Business Research* 69 (8): 3192–3198.

George, G., M.R. Haas, and A. Pentland. 2014. Big data and management. *Academy of Management Journal* 57 (2): 321–326.

Grinstein, A., and P. Riefler. 2015. Citizens of the (green) world? Cosmopolitan orientation and sustainability. *Journal of International Business Studies* 46 (6): 694–714.

Hauser, D.J., and N. Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods* 48 (1): 400–407.

Lance, C.E., B. Dawson, D. Birkelbach, and B.J. Hoffman. 2010. Method effects, measurement error, and substantive conclusions. *Organizational Research Methods* 13 (3): 435–455.

Malhotra, N.K., S.S. Kim, and A. Patil. 2006. Common method variance in IS research: A comparison of alternative approaches and a reanalysis of past research. *Management Science* 52 (12): 1865–1883.

Nunnally, J. 1967. *Psychometric theory*. New York: McGraw-Hill.

Podsakoff, P.M., S.B. MacKenzie, J.Y. Lee, and N.P. Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88 (5): 879–903.

Podsakoff, P.M., S.B. MacKenzie, and N.P. Podsakoff. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology* 63: 539–569.

Richardson, H., M. Simmering, and M. Sturman. 2009. A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods* 12 (4): 762–800.

Shaffer, M.A., B.S. Reiche, M. Dimitrova, M. Lazarova, S. Chen, M. Westman, and O. Wurtz. 2016. Work- and family-role adjustment of different types of global professionals: Scale development and validation. *Journal of International Business Studies* 47 (2): 113–139.

Sharma, R., P. Yetton, and J. Crawford. 2009. Estimating the effect of common method variance: The method–method pair technique with an illustration from TAM research. *MIS Quarterly* 33 (3): 473–490.

Spector, P. 1987. Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology* 72 (3): 438–443.

Spector, P.E. 2006. Method variance in organizational research: Truth or urban legend? *Organizational Research Methods* 9 (2): 221–232.

Spector, P.E., C.C. Rosen, H.A. Richardson, L.J. Williams, and R.E. Johnson. 2019. A new perspective on method variance: A measure-centric approach. *Journal of Management* 45 (3): 855–880.

Su, C., Z. Yang, G. Zhuang, N. Zhou, and W. Dou. 2009. Interpersonal influence as an alternative channel communication behavior in emerging markets: The case of China. *Journal of International Business Studies* 40 (4): 668–689.

Williams, L.J., and B.K. Brown. 1994. Method variance in organizational behavior and human resources research: Effects on correlations, path coefficients, and hypothesis testing. *Organizational Behavior and Human Decision Processes* 57 (2): 185–209.

Williams, L.J., J.A. Cote, and M.R. Buckley. 1989. Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology* 74 (3): 462–468.

22

Common Method Variance in International Business Research: Further Reflections

Arjen van Witteloostuijn, Lorraine Eden, and Sea-Jin Chang

We want to thank Harold Doty and Marina Astakhova (Chap. 21, this volume) for their excellent commentary on our *JIBS* (2010) editorial on common method variance (CMV), which is reprinted as Chap. 20 (this volume). We would like to support in particular their suggestions for reviewers. After reflecting on this insightful commentary, we like to take the opportunity to add the following elements to the debate.

We thank Bo Bernhard Nielsen for helpful comments on an earlier draft.

A. van Witteloostuijn
Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
e-mail: a.van.witteloostuijn@vu.nl

L. Eden
Department of Management, Texas A&M University, College Station, TX, USA
e-mail: leden@tamu.edu

S.-J. Chang (✉)
National University of Singapore, Singapore, Singapore
e-mail: schang@nus.edu.sg

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_22

In the nine years since our *JIBS* piece appeared, CMV has moved from a problem facing psychology researchers to a challenge recognized by all social scientists. Scholars have extended the CMV literature to disciplines such as entrepreneurship (Tehseen et al. 2017), international marketing (Jean et al. 2016) and public administration (George and Pandey 2017).

As Doty and Astakhova (Chap. 21, this volume) observe, CMV has become a controversial issue. While the consensus is that CMV exists (Schwartz et al. 2017), scholars differ in their views as to whether CMV matters a little or a lot (George and Pandey 2017) – or even is simply an "urban legend" (see Spector 2006; for a contrary view, see Schwartz et al. 2017). Some authors also argue that both uncommon method variance (UMV) and CMV should be considered, and at the level of the individual variable, rather than the method (Spector et al. 2019).

The "politicization" of the dialogue on CMV was something that we did not expect when we wrote our editorial. We had attempted to provide a balanced assessment, noting that CMV might need to be tolerated in IB research, especially when investigating empirical phenomena where data are scarce (as is often the case in IB research in general and developing countries in particular). Moreover, we noted that large readily available datasets such as ORBIS and Compustat had their own problems. As *JIBS* editors, we wanted to encourage primary-data sourced, quantitative IB studies to take CMV seriously. We recommended that authors carefully consider whether CMV is potentially a problem, and contemplate whether and how to remediate the issue, but we did not intend to discourage them with the "big stick" of CMV supposedly invalidating empirical results "by definition".

We continue to believe, and cannot emphasize enough, that many CMV issues can be avoided or minimized through making appropriate *ex ante* research design decisions. Systematically considering alternative designs, such as experimental and longitudinal ones, in combination with focused econometric identification strategies is key. Of course, this does not imply that the single-respondent–one-shot survey design should be ruled out altogether. On the contrary, for specific research questions, this design is still useful. As an illustration, let us use the example of studying self-identified victims of perceived opportunism in international business transactions (dependent variable), where we assess the

antecedents or drivers of opportunism. The suggestion by reviewers that researchers should also survey the perpetrators who engaged in opportunism makes little sense. A credible study of victims does not require validating their views by surveying those who made them victims. At the same time, often other research designs would have been more powerful, including multi-person and/or multi-period survey ones. For example, in Urbig, Terjesen, Procher, Muehlfeld and van Witteloostuijn (2016), the dependent variable is based on a lab experiment, with information on the independent variables collected through a survey, whereby both measurements are separated by a time lag.

Before measuring anything (which in our field usually involves collecting secondary data), a researcher should analyse *why* s/he wants to know *what*, and what the answers to why and what imply for measures. As a field of study, we probably do not devote sufficient attention to this issue. Let us address the question of measures by making an analogy with the uses of a river. When we ask the question as to whether we can wade from one bank of the river to the other, we can answer this question by measuring the maximum depth of the river in the season of maximum flow, not its average depth at a random point in time. However, if we want to build a reliable bridge over the river, we should again consider the season of maximum flow, but this time measure the distance between the two banks. If we seek to estimate the risk of overflowing, we need to measure the maximum flow itself, rather than focusing on the depth of the river or the distance between the banks. Finally, if we want to find out what area can be irrigated on a continuous basis thanks to the river, we should measure the average flow, rather than the maximum flow. And so on. This issue of why and what may seem trivial, but it is not. The principle that a researcher should start by carefully considering why s/he is collecting what data before starting to do so is often violated, we fear. But by doing this carefully, issues of CMV can often be avoided or minimized as well.

In our editorial, we recommended that both *ex ante* and *ex post* approaches be deployed to handle CMV; that *ex ante* approaches are preferable; and that *ex post* solutions should involve multiple methods and tests. Schwartz et al. (2017) recently reached similar conclusions to ours in their empirical tests of CMV, whereby they found that the Harmon single-factor test did not detect CMV and that the best approach

was an *ex ante* design approach coupled with using instrumental variables in two-stage least squares (2SLS) to obtain consistent estimates. Antonakis, Bendahan, Jacquart and Lalive (2010) similarly recommended using 2SLS to handle CMV. It is also possible that the addition of non-linear, mediation and moderation relationships to the estimated model will increase complexity beyond what any respondent would reasonably be aware of at the time of the survey (Podsakoff et al. 2003; Siemsen et al. 2010; and Podsakoff et al. 2012).

Thinking ahead, we know that CMV is often associated with single-respondent–one-shot survey designs. But to what extent are other designs (potentially) affected by CMV? We find this question particularly interesting in the context of the data science revolution, which is likely to have an impact on IB research as well. For instance, to what extent is – or can – CMV be associated with measures based on text analysis of scraped data? One of the authors recently did his first "data science" paper (van Witteloostuijn and Kolkman 2018) and was intrigued by what we can now do with data science techniques, particularly machine learning algorithms; however, we are still in the middle of the process of finding out all the pros and cons of such techniques, with CMV being a possible issue.

We think that the CMV challenge provides an additional argument as to why the IB field needs more replication of different types (cf. Walker et al. 2019), whereby we should to the extent possible study a research question through different designs, methods and measures – apart from using different samples.

References

Antonakis, J., S. Bendahan, P. Jacquart, and R. Lalive. 2010. On making causal claims: A review and recommendations. *The Leadership Quarterly* 21: 1086–1120.

George, Bert, and Sanjay K. Pandey. 2017. We know the Yin – But where is the Yang? Toward a balanced approach on common source bias in public administration scholarship. *Review of Public Personnel Administration* 37 (2): 245–279.

Jean, Rueey-Jer Bryan, Ziliang Deng, Daekwan Kim, and Xiaohui Yuan. 2016. Assessing endogeneity issues in international marketing research. *International Marketing Review* 33 (3): 483–512.

Podsakoff, P.M., S.B. MacKenzie, and J. Lee. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88: 879–903.

Podsakoff, P.M., S.B. MacKenzie, and N.P. Podsakoff. 2012. Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology* 63: 539–569.

Schwartz, Andrew, Tracey Rizzuto, Jose Luis Roldan, and Ramon Barrera-Barrera. 2017. Examining the impact and detection of the "Urban Legend" of common method bias. *The DATABASE for Advances in Information Systems* 48 (1): 93–118.

Siemsen, E., A. Roth, and P. Oliveira. 2010. Common method bias in regression models with linear, quadratic, and interaction effects. *Organizational Research Methods* 13 (3): 456–476.

Spector, Philip E. 2006. Method variance in organizational research: Truth or urban legend? *Organizational Research Methods* 9 (2): 221–232.

Spector, P.E., C.C. Rosen, H.A. Richardson, L.J. Williams, and R.E. Johnson. 2019. A new perspective on method variance: A measure-centric approach. *Journal of Management* 45 (3): 855–880.

Tehseen, Shehnaz, T. Ramayah, and Sulaiman Sajilan. 2017. Testing and controlling for common method variance: A review of the available methods. *Journal of Management Studies* 4 (2): 146–175.

Urbig, D., S. Terjesen, V. Procher, K.S. Muehlfeld, and A. van Witteloostuijn. 2016. Come on and take a free ride: Contributing to public goods in native and foreign language settings. *Academy of Management Learning & Education* 15: 268–286.

van Witteloostuijn, A., and D. Kolkman. 2018. Is firm growth random? A machine learning perspective. *Journal of Business Venturing Insights* 11: e00107.

Walker, R.M., G.A. Brewer, M.J. Lee, N. Petrovsky, and A. van Witteloostuijn. 2019. Best practice recommendations for replicating experiments in public administration. *Journal of Public Administration Research and Theory*. https://academic.oup.com/jpart/advance-article/doi/10.1093/jopart/muy047/5074357.

Part XI

Multilevel Models

23

Multilevel Models in International Business Research

Mark F. Peterson, Jean-Luc Arregle, and Xavier Martin

Introduction

Research published in the *Journal of International Business Studies* often links institutional or cultural characteristics of nations to features of businesses and business people that are nested within nations. Other research analyses subsidiaries nested within multinational enterprises and economic regions. Previously, such work has matched samples on selected characteristics such as organizational (Hofstede 1980/2001) or industry context (Fischer et al. 2010; House, Hanges, Javidan, Dorfman, Gupta, and GLOBE Associates 2004) in a way that reduced internal validity

M. F. Peterson
Aarhus University, Aarhus, Denmark

J.-L. Arregle (✉)
EM Lyon Business School, Lyon, France
e-mail: arregle@em-lyon.com

X. Martin
Tilburg University, Tilburg, The Netherlands
e-mail: X.Martin@uvt.nl

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_23

problems, but which required external validation for application in other contexts.

One option for dealing with that problem was to aggregate variables at the lower level (Level 1) to a higher level (Level 2). Alternatively, Level 2 data were often disaggregated to Level 1. Both approaches have limitations. Aggregation removes Level 1 variance and eliminates the opportunity to control for Level 1 confounding variables. Hence it creates the risk of ecological fallacy, that is, "the assumption that relationships between variables at the aggregate level imply the same relationships at the individual level" (Jargowsky 2005: 715). Disaggregation produces biased statistics by treating Level 2 values that are assigned to Level 1 observations as independent (Arregle et al. 2006).

By using multi-level modeling (MLM), however, researchers studying higher-level constructs can effectively control for theoretically extraneous individual-level variables (Cullen et al. 2004), and researchers studying lower-level constructs can treat Level 2 contingencies as continuous rather than categorical variables (Smith, Peterson, Thomason, and the Event Meaning Management Research Group 2011). MLM thus allows IB scholars to accurately model context *and* lower-level effects.

This editorial is not intended to resolve all MLM controversies in IB by editorial fiat, but explains some critical considerations and literatures that are often overlooked in MLM research. We first introduce statistical tools, and discuss issues in using them. We then consider directions for international comparative research and MNE research.

Multilevel Analytical Tools and Issues

The basics of MLM methods for analyzing nested data are well known (e.g., Hofmann 1997), but criteria for deciding when to use them remain controversial. Recent advances in analysis methods create unrealized potential for IB research.

MLM Basics

Although MLM models can be used with more than two levels of nesting, two-level models are the most common. They take the basic form.

Level 1:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij} \tag{23.1}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j} \tag{23.2}$$

and

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j} \tag{23.3}$$

where i is a Level 1 observation, j is a Level 2 observation, X is a Level 1 predictor, and W is a Level 2 predictor.

In "intercepts and slopes as outcomes" models, β_{0j} (an intercept) and β_{1j} (a slope) are explained by variables measured at the higher level (Level 2). The intercepts part of the model predicts differences in the average level of a dependent variable at Level 1 from some Level 2 independent variable. The slopes part of the model predicts the relationship between two Level 1 variables from some Level 2 variable. In simple nested models, each Level 1 observation belongs only to one Level 2 group. By contrast, in cross-nested (or cross-classified) MLM a Level 1 observation can be nested in two Level 2 groups. Longitudinal studies can nest separate occasions of events (as Level 1 observations) within characteristics of situations that are modeled as stable (Level 2) characteristics (Martin et al. 2007).

Criteria for Using Mixed Linear Models

The criteria for deciding whether to use MLM are generally well known, but the specific cut-off points for an appropriate application are more controversial. Consider three cases.

First, when a research problem suggests that the means of variables measured at Level 1 can be predicted by a Level 2 variable, intraclass correlation coefficients (ICCs) or similar methods are used to determine that the Level 1 observations differ significantly between Level 2 groups (Bliese and Hanges 2004; Hanges and Dickson 2004). As a rule of thumb, Hox (2010) considers ICCs of 0.05, 0.10, and 0.15 as small, medium, and large, respectively, for organizational research.

Second, other research problems simply require that Level 2 effects be controlled. For example, a study might test for a consistent relationship between strategy and organizational effectiveness across a small number of nations, multiple-nation regions, or time periods. When the purpose is to control for nesting rather than to test Level 2 hypotheses, and such small numbers of Level 2 groups limit the power to identify significant Level 2 effects, the criterion that one should first demonstrate significant Level 2 effects is relaxed.

Third, when the researcher is interested in differences in relationships between Level 1 variables depending on a Level 2 variable, then the ICC test to show mean differences in Level 1 variables among Level 2 groups is not relevant. Instead, researchers should test whether Level 1 slopes differ between Level 2 units before attempting to explain differences in slopes (Raudenbush et al. 2004). IB researchers should heed the difference in the utility of ICCs and similar coefficients for evaluating the appropriateness of MLM to handle these three very different research problems.

Level 1 and Level 2 Sample Sizes

Sample-size problems arise in multilevel international research because of the limited number of nations in the world, data quality and accessibility problems for many nations, interest in studying multiple-nation groupings, availability of data from few time periods, or the small number of members in teams or subsidiaries in MNEs. Sample-size criteria are based on numbers of predictors at each level, and on whether fixed effects or random effects models are being studied.

In MLM, fixed effects describe group-specific features assumed to affect the dependent variable, whereas random effects assume that the group is drawn randomly from a larger population. Fixed-effects models, which can thus be useful for researchers interested only in interpreting the specific Level 2 groups being studied, require fewer Level 2 groups than do random-effects models that seek generalization to populations represented by the Level 2 groups.

Kreft (1996) suggested 30 Level 2 groups and 30 Level 1 observations per group. Simulation studies indicate that the balance swings toward more than 30 groups and fewer than 30 observations per group for hypotheses about the effects of Level 2 variables (e.g., Maas and Hox 2005; Snijders and Bosker 1993). For a two-level model, Hox (2010) recommends at least 20 observations for 50 groups to test cross-level interactions, and at least 10 observations for 100 groups to test random effects. He also presents detailed power analyses based on both sample sizes and anticipated effect sizes. Spybrook et al. (2011) provide software for power analysis for specific MLM situations.[1]

MLM can also use unbalanced data with different numbers of observations per group. Bell et al. (2008) even apply it when a small percentage of groups have only one or two observations. For research with few Level 2 groups, moderated regression or multi-group structural equation modeling that treats the groups as categories only may be more appropriate than MLM (Meuleman and Billiet 2009).

IB researchers should stay abreast of research about MLM sample sizes, since this literature is still evolving. They should also note that most simulation studies are based on simple models with only one Level 1 and one

Level 2 predictor, so their results may yet understate the sample size demands for the more complex multivariate models of interest to IB scholars.

Measure Equivalence at Multiple Levels

MLM research relating higher level (e.g., Level 2) predictors to group means of Level 1 criteria should consider whether or not the variables represented by group means are comparable to their individual-level measurement structures. Factor structures can be compared at multiple levels of analysis (Hanges and Dickson 2004; Muthen 1994), but scholars continue to debate whether factor equivalence across levels is always desirable (Fischer et al. 2010). For example, using measures of values at both national and individual levels requires that a researcher consider how parties at a lower level respond to their shared experiences with the values that characterize their societies (Peterson and Wood 2008). Obviously, not all groups or individuals wholly accept their society's values. The frequent finding of more differentiated value-related factors at individual than at aggregate levels can either be an artifact of smaller samples and variance in aggregated data or have a substantive basis (Ostroff 1993; Peterson 2009). Scholars who argue that either Level 1 and Level 2 data structures should be similar, or that they should differ, need to provide both statistical and substantive explanations for their position.

Recent Advances in Analysis Methods

New methods and statistical software for MLM research regularly appear. MLM has become available for analyzing dichotomous, nominal, count, ordinal, and truncated dependent variables (Cuypers and Martin 2010). Methods for testing mediating effects are also available, although they are complex and sensitive (Zhang et al. 2009).

In the next two sections, we consider specific directions for the fruitful use of MLM in IB: international comparative research and MNE research, respectively.

International Comparative Research

Multilevel datasets and analysis methods draw attention to controversies about concepts such as nations, multiple-nation groupings, and within-nation regions (Hitt et al. 2007; Peterson and Soendergaard 2011).

Nations

The concept of nation is surprisingly controversial (Smith 2004). At one extreme, the nation is either a taken-for-granted construct, or is something that distinguishes what is legitimately IB research from what is not (Tung and van Witteloostuijn 2008). At the other extreme, nations are seen as being so ephemeral that they are inconsequential (Tsui et al. 2007).

Addressing these extremes has encouraged scholars to think about nations in increasingly nuanced ways (Peterson and Soendergaard 2011; Tung 2008). Studies testing Level 1 hypotheses that control for nesting of individuals or organizations within nations should use MLM when ICCs show significant nation differences in Level 1 variables. Studies using simple designs that have no interest in generalizing beyond the nations studied (e.g., nations of Latin America) can use fixed-effects statistics to test nation-level hypotheses that require relatively few Level 2 observations. Studies with multiple nation-level predictors, or which use random-effects statistics to make inferences to nations in general, need to pay special attention to power limitations (Hox 2010).

Although MLM cannot solve the theoretical problem of the significance of nation-states, it is and should continue to be used by IB researchers to probe the utility of studying nation-states as compared with multiple-nation clusters and within-nation regions.

Multiple-Nation Clusters

IB scholars sometimes combine sets of nations. Cultural rationales for grouping nations can be based on functional similarities, ancient history, historical institutional spread, traditional occupations, modern institutional arrangements, economics, or economic change (Gupta and Hanges 2004; Inglehart and Baker 2000; Schwartz 2004). Economic rationales for clustering have been proposed when nations are physically contiguous or economically integrated (Arregle et al. 2009; Rugman and Verbeke 2004). The cultural basis and the economic basis for clustering nations can overlap when economic ties have cultural implications (or the opposite). Proximity can promote cultural convergence through cross-border interaction and the development of international integrative mechanisms that promote cross-border business relationships.

Since applications of research about multiple-nation clusters are typically to the specific clusters studied rather than to a larger population of clusters, fixed-effects models are often appropriate. The number of multiple-nation clusters is typically too limited to test hypotheses about them as the higher level in MLM (i.e., Level 2 in two-level models or Level 3 in three-level models). Nevertheless, theoretical considerations often suggest that nation clusters should be treated as controls or studied using other methods. Nation clusters could also be used at the intermediate level (e.g., Level 2 in a three-level model) in MLM with more than two levels, as long as the sample sizes at the different levels are sufficient.

Within-Nation Regions

The increasingly documented variability in cultural, institutional, and economic characteristics within nations suggests that IB researchers should consider studying within-nation regions (e.g., Chan et al. 2010; Crone 2005; Lenartowicz and Roth 2001). When something that is typically conceived as a nation characteristic also shows meaningful within-nation differences, considering regions as a level of analysis has the potential to overcome the Level 2 sample size problem that plagues MLM research in IB.

Large-Scale Datasets About Societies

IB scholars should take advantage of the large international datasets about social, cultural, and economic characteristics that are now publicly available. For example, data at the individual, nation and sometimes the within-nation region levels are available for versions of the Schwartz Value Survey (Fischer et al. 2010), World Value Survey (Inglehart et al. 1998), European Social Survey (2011), and Global Entrepreneurship Monitor (2010), and from Eurostat. While promising, these databases have common limitations that are only partially compensated for by their large sample sizes. One is that items often use simple words and dichotomous response alternatives to accommodate respondents with limited education. Another is the frequent use of single-item scales, making translation equivalence and response bias difficult to evaluate (Hult et al. 2008).

Research About MNEs

Our discussion to this point has emphasized comparative projects where nation is one level of nesting. We also see several research directions for multilevel research about MNEs.

From Globalization to Semi-Globalization

A "global" perspective on MNEs that connects global headquarters with country-level subsidiaries dominates IB (Rugman and Verbeke 2004). However, a recent semi-globalization approach holds that incomplete cross-border integration generates neither global integration nor national market fragmentation (Ghemawat 2003). Consequently, multi-nation regions affect MNEs' international strategy. Firms investing in particular regions recognize that some countries share economic and political interdependence, common cultural or ethnic heritages, or historically developed relations that can support firm-specific advantages (Rugman and Verbeke 2004). Such regions can be more culturally or institutionally

meaningful than nations. When region-level sample sizes allow, IB researchers should use MLM to consider whether well-known relationships at the country or foreign direct investment (FDI) level also appear at the region level. Scholars should also consider such unique effects of a region as, for instance, whether region-level experience complements or substitutes for country-level experience in determining entry mode.

Interorganizational Effects and Clusters Within Countries

Recognizing that FDI decisions are affected by the actions of other firms, IB scholars have studied the relative effects of buyers and rivals (Martin et al. 1998), timing as well as count effects (Martin et al. 1998), and institutional forms such as business groups (Belderbos and Sleuwaegen 1996) – all potentially fruitful areas for MLM (Martin et al. 2007). Recognizing the pros and cons for each firm in agglomerating with rival or complementary firms (Shaver and Flyer 2001), researchers should continue to examine clusters in foreign entry (Martin et al. 2010). MLM should also be used to connect such cluster effects with differences in economic and cultural attractiveness of regions within nations.

Teams and Subsidiaries Within the MNE

Much research on intercultural teams in MNEs is conceptual, experimental, or ethnographic (Leung and Peterson 2010). Large-scale surveys can contribute as well. Thus Leung and Peterson (2010) propose team characteristics (e.g., diversity of knowledge and perspectives) and outcomes at Level 1, vs communication technology (e.g., amount of physical presence) at Level 2. Intercultural negotiation research (Brett and Crotty 2008) compares the influence of the personal values of team members on various negotiation processes and outcomes (potentially Level 1) vs national culture characteristics of the negotiating teams (potentially Level 2). Research on MNEs has also focused on the effects of MNE-subsidiary relationships, subsidiary environments, and networks across subsidiaries

on various organizational outcomes. MLM not only perfectly fits the multilevel structure of these research questions; it also allows a better and more elaborated modeling, opening up new theoretical perspectives for quantitative studies on this topic. Accordingly, MLM should be used to study how higher-level nation or MNE variables, and lower-level subsidiary or team variables, and their interactions, explain lower-level decisions and outcomes.

Conclusion

Studies using MLM methods occasionally appear in *JIBS,* but MLM continues to have unrealized potential for addressing a broad range of international comparative and MNE research topics. Methodologically, we have pointed out that researchers using MLM must first thoroughly understand the nature of the multilevel nesting and relationships in their data and hypotheses, and the corresponding diagnosis and sample size requirements. In our collective editing experience, we frequently find submissions in which the level of analysis of the theory, measures, or analysis techniques do not match. We also find MLM submissions based on datasets that lack the sample sizes for the sorts of hypotheses being tested, and we find regression-based submissions using solely either disaggregated or aggregated analyses that would be better conducted using MLM. Equally importantly, MLM is attuned to the increasing conceptual sophistication of IB research. The combination of MLM with advances in theory about the significance of nations, as well as about levels such as clusters of nations, within-nation regions, and MNE groups and subunits, suggests that MLM should and will have an increasing role in IB research well into the future, whenever conditions for its use are met.

Acknowledgements The authors thank Shige Makino and Anand Swaminathan for contributions to an earlier version of this editorial; and Paul Bliese, Fons van de Vijver, the *JIBS* Editors and Area Editors, and three anonymous reviewers for their useful comments.

Note

1. Beside this Optimal Design software, other packages to determine the appropriate sample size based on power estimates in specific cases of MLM include PinT (http://www.stats.ox.ac.uk/~snijders/multilevel.htm) and RMASS (http://www.rmass.org/)

References

Arregle, J.L., L. Hebert, and P.W. Beamish. 2006. Mode of international entry: The advantages of multilevel methods. *Management International Review* 46 (5): 557–618.

Arregle, J.L., P.W. Beamish, and L. Hebert. 2009. The regional dimension of MNEs' foreign subsidiary localization. *Journal of International Business Studies* 40 (1): 86–107.

Belderbos, R., and L. Sleuwaegen. 1996. Japanese firms and the decision to invest abroad: Business groups and regional core networks. *Review of Economics and Statistics* 78 (2): 214–220.

Bell, B.A., J.M. Ferron, and J.D. Kromrey. 2008. Cluster size in multilevel models: The impact of sparse data structures on point and interval estimates in two-level models. In *Proceedings of the joint statistical meetings, survey research methods section*, 1122–1129. Alexandria: American Statistical Association.

Bliese, P.D., and P.J. Hanges. 2004. Being both too liberal and too conservative: The perils of treating grouped data as though they were independent. *Organizational Research Methods* 7 (4): 400–417.

Brett, J.M., and S. Crotty. 2008. Culture and negotiation. In *Handbook of cross cultural management research*, ed. P.B. Smith, M.F. Peterson, and D.C. Thomas, 269–284. Beverly Hills: Sage Press.

Chan, C.M., S. Makino, and T. Isobe. 2010. Does subnational region matter? Foreign affiliate performance in the United States and China. *Strategic Management Journal* 31 (11): 1226–1243.

Crone, T.M. 2005. An alternative definition of economic regions in the United States based on similarities in state business cycles. *Review of Economics and Statistics* 87 (4): 617–626.

Cullen, J.B., K.P. Parboteeah, and M. Hoegl. 2004. Cross-national differences in managers' willingness to justify ethically suspect behaviors: A test of institutional anomie theory. *Academy of Management Journal* 47 (3): 411–421.

Cuypers, I.R.P., and X. Martin. 2010. What makes and what does *not* make a real option? A study of equity shares in international joint ventures. *Journal of International Business Studies* 41 (1): 47–69.

European Social Survey. 2011. www.europeansocialsurvey.org

Fischer, R., M. Vauclair, J.R.J. Fontaine, and S.H. Schwartz. 2010. Are individual-level and country-level value structures different? Testing Hofstede's legacy with the Schwartz Value Survey. *Journal of Cross-Cultural Psychology* 41 (2): 135–151.

Ghemawat, P. 2003. Semiglobalization and international business strategy. *Journal of International Business Studies* 34 (2): 138–152.

Global Entrepreneurship Monitor. 2010. www.gemconsortium.org

Gupta, V., and P.J. Hanges. 2004. Regional and climate clustering of societal cultures. In *Culture, leadership, and organizations: The GLOBE study of 62 societies*, ed. R.J. House, P.J. Hanges, M. Javidan, P.W. Dorfman, and V. Gupta, 178–218. Thousand Oaks: Sage.

Hanges, P.J., and M.W. Dickson. 2004. The development and validation of the GLOBE culture and leadership scales. In *Culture, leadership, and organizations: The GLOBE study of 62 societies*, ed. R.J. House, P.J. Hanges, M. Javidan, P.W. Dorfman, and V. Gupta, 122–151. Thousand Oaks: Sage.

Hitt, M.A., P.W. Beamish, S.E. Jackson, and J.E. Mathieu. 2007. Building theoretical and empirical bridges across levels: Multilevel research in management. *Academy of Management Journal* 50 (6): 1385–1399.

Hofmann, D.A. 1997. An overview of the logic and rationale of hierarchical linear models. *Journal of Management* 23 (6): 723–744.

Hofstede, G. 1980/2001. *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks: Sage Publications.

House, R.J., P.J. Hanges, M. Javidan, P.W. Dorfman, V. Gupta, and GLOBE Associates. 2004. *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks: Sage.

Hox, J.J. 2010. *Multilevel analysis: Techniques and applications*. 2nd ed. New York: Routledge.

Hult, G.T.M., D.J. Ketchen, D.A. Griffith, C.A. Finnegan, T. Gonzalez-Padron, N. Hamancioglu, Y. Huang, M.B. Talay, and S.T. Cavusgil. 2008. Data equivalence in cross-cultural international business research: Assessment and guidelines. *Journal of International Business Studies* 39 (6): 1027–1044.

Inglehart, R., and W.E. Baker. 2000. Modernization, cultural change, and the persistence of traditional values. *American Sociological Review* 65 (1): 19–51.

Inglehart, R., M. Basañez, and A. Moreno. 1998. *Human values and beliefs: A cross-cultural sourcebook*. Ann Arbor: University of Michigan Press.

Jargowsky, P.A. 2005. Ecological fallacy. In *Encyclopedia of social measurement*, ed. K. Kempf-Leonard, 715–722. San Diego: Elsevier.

Kreft, I.G.G. 1996. *Are multilevel techniques necessary? An overview, including simulation studies*. Los Angeles: California State University at Los Angeles.

Lenartowicz, T., and K. Roth. 2001. Does subculture within a country matter? A cross-cultural study of motivational domains and business performance in Brazil. *Journal of International Business Studies* 32 (2): 305–325.

Leung, K., and M.F. Peterson. 2010. Globally distributed workforce: Social and international issues. In *Handbook of industrial-organizational psychology*, ed. S. Zedeck, H. Aguinis, W.F. Casio, M.J. Gelfand, K. Leung, S.K. Parker, and J. Zhou, 771–805. Washington, DC: APA.

Maas, C.J.M., and J.J. Hox. 2005. Sufficient sample sizes for multilevel modeling. *Methodology* 1 (3): 86–92.

Martin, X., A. Swaminathan, and W. Mitchell. 1998. Organizational evolution in the interorganizational environment: Incentives and constraints on international expansion strategy. *Administrative Science Quarterly* 43 (3): 566–601.

Martin, X., A. Swaminathan, and L. Tihanyi. 2007. Modeling international expansion. In *Research methodology in strategy and management*, ed. D.J. Ketchen and D.D. Bergh, vol. 4, 101–117. Amsterdam: Elsevier.

Martin, X., R.M. Salomon, and Z. Wu. 2010. The institutional determinants of agglomeration: A study in the global semiconductor industry. *Industrial and Corporate Change* 19 (6): 1769–1800.

Meuleman, B., and J. Billiet. 2009. A Monte Carlo sample size study: How many countries are needed for accurate multilevel SEM? *Survey Research Methods* 3 (1): 45–58.

Muthen, B. 1994. Multilevel covariance structure analysis. *Sociological Methods and Research* 22 (30): 376–398.

Ostroff, C. 1993. Comparing correlations based on individual-level and aggregated data. *Journal of Applied Psychology* 78 (4): 569–582.

Peterson, M.F. 2009. Cross-cultural comparative studies and problems of coordinating international research teams. In *Handbook of organizational research methods*, ed. D.A. Buchanan and A. Bryman, 328–345. London: Sage.

Peterson, M.F., and M. Soendergaard. 2011. Traditions and transitions in quantitative societal culture research in organization studies. *Organization Studies* 32 (11): 1539–1558.

Peterson, M.F., and R.E. Wood. 2008. Cognitive structures and processes in cross cultural management. In *Handbook of cross-cultural management research*, ed. P.B. Smith, M.F. Peterson, and D.C. Thomas, 15–33. Thousand Oaks: Sage.

Raudenbush, S.W., A.S. Bryk, Y.F. Cheong, R. Congdon, and M. du Toit. 2004. *HLM 6: Linear and nonlinear modeling*. Lincolnwood: Scientific Software International.

Rugman, A.M., and A. Verbeke. 2004. A perspective on regional and global strategies of multinational enterprises. *Journal of International Business Studies* 35 (1): 3–18.

Schwartz, S.H. 2004. Mapping and interpreting cultural differences around the world. In *Comparing cultures: Dimensions of culture in a comparative perspective*, ed. H. Vinken, J. Soeters, and P. Ester, 43–73. Leiden: Brill.

Shaver, J.M., and F. Flyer. 2001. Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strategic Management Journal* 21 (12): 1175–1193.

Smith, P.B. 2004. Nations, cultures, and individuals: New perspectives and old dilemmas. *Journal of Cross-Cultural Psychology* 35 (1): 6–12.

Smith, P.B., M.F. Peterson, S.P. Thomason, and the Event Meaning Management Research Group. 2011. National culture as a moderator of the relationship between leaders' use of guidance sources and how well work events are handled. *Journal of Cross Cultural Psychology* 42 (6): 1103–1123.

Snijders, T.A.B., and R.J. Bosker. 1993. Standard errors and sample sizes for two-level research. *Journal of Educational and Behavioral Statistics* 18 (3): 237–259.

Spybrook, J., S.W. Raudenbush, R. Congdon, and A. Martínez. 2011. Optimal design for longitudinal and multilevel research: Documentation for the "optimal design" software. Available from http://www.wtgrantfoundation.org. or from http://sitemaker.umich.edu/group-based

Tsui, A.S., S.S. Nifadkar, and A.Y. Ou. 2007. Cross-national, cross-cultural organizational behavior research: Advances, gaps, and recommendations. *Journal of Management* 33 (3): 426–478.

Tung, R.L. 2008. The cross-cultural research imperative: The need to balance cross-national and intra-national diversity. *Journal of International Business Studies* 39 (1): 41–46.

Tung, R.L., and A. van Witteloostuijn. 2008. From the editors: What makes a study sufficiently international? *Journal of International Business Studies* 39 (1): 180–183.

Zhang, Z., M.J. Zyphur, and K.J. Preacher. 2009. Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods* 12 (4): 695–719.

24

Multilevel Models in International Business Research: A Commentary

Robert J. Vandenberg

In general, the Peterson, Arregle and Martin (2012) article is a well-written summary of multilevel modelling (MLM). The article covers the basics including MLM's use as a test for substantive multilevel questions as well as a control for the lack of independence among the units of observations. Further, the authors address intraclass correlation as a test as to whether or not aggregation is warranted when addressing substantive multilevel questions. The article also addresses the very important topic of sample size, which needs to be addressed at all levels in the MLM.

These benefits will be obvious to most international business (IB) and management scholars who are familiar with the MLM technique. However, I would like to use this commentary on the Peterson et al. (2012) article to address things that are not so obvious to the reader,

Peterson, M.F., J.L. Arregle, and X. Martin. 2012. Multilevel models in international business research. *Journal of International Business Studies* 43 (5): 451–457.

R. J. Vandenberg (✉)
University of Georgia, Athens, GA, USA
e-mail: rvandenb@uga.edu

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_24

specifically: (a) uncovering the "hidden" jewels within the manuscript, and (b) introducing the reader to advances since the article was published.

With respect to "hidden" jewels or nuggets that may not be obvious to the reader, there are four in the Peterson et al. (2012) article. The first one is the authors not mentioning tests of cross-level direct effects. This is not a criticism. Rather, it is a positive feature of the manuscript. A cross-level direct effect is a hypothesis that a level-2 variable directly impacts a level-1 variable; for example, the financial performance of a firm directly impacts individual employee productivity. However, as noted by LoPilato and Vandenberg (2015), it is statistically impossible to test cross-level direct effects because there is no shared covariance between levels. You either have to bring the level 2 variable down to level 1 and assign the level 2 value to each person in the unit, or you have to aggregate the level 1 variable upward and make it a level 2 variable. The point is that in order for the variables to covary they have to reside at one level or the other. Consequently, there is no "cross-level anything" about testing the association. Peterson et al. (2012) are to be commended for not even mentioning this as a possibility since it could consequently encourage the reader to engage in a faulty practice.

A second hidden nugget is Peterson et al.'s (2012) avoidance of the term "cross-level" when discussing how a level 2 variable may be predicted to impact the relationship (slope) between two level 1 variables within each level 2 unit. This is frequently and inappropriately referred to as a cross-level interaction in the research literature. As was the case above, there is nothing cross-level about testing such hypotheses. Let's assume for illustration purposes that we have 100 groups (level 2 units) with 10 people in each group, each measured on some independent variable, x, and a dependent variable, y. Let's further assume that we have strong substantive reasons to predict that some variable, z, measured only at level 2 will impact the slopes representing the regression of y on x. As was the case above, there is no shared covariance across levels. The common practice, therefore, is to create a random vector of slopes by regressing y onto x within each group. The random vector of slopes in the current illustration would consist of 100 slopes, but most importantly, it is now a level 2 variable, which can now be regressed onto the level 2 variable z. Consequently, the question really being addressed is whether the strength

of the slopes across the 100 groups varies as a function of the level 2 variable. Peterson et al.'s (2012) method of describing such cases is appropriate and accurate in that they made no reference to these cases as being cross-level tests.

The third hidden nugget is Peterson et al.'s (2012) brief mention of cross-classified cases (see the top of the second column on page 452). I understand that they were limited to a relatively small number of pages to introduce MLM and could not, therefore, delve into all of its complexity. However, I would encourage the reader not to take cross-classification lightly, but rather to obtain a deeper understanding of it and of the implications cross-classification may have for MLM substantive questions. Cross-classification in its simplest form means that some unit of observation may be nested into more than one level 2 unit, or a level 2 unit may be nested within more than one level 3 unit. Since a large aspect of undertaking MLM analyses is partitioning the variance of the variables (i.e., what proportion of its variance is due to level 1 vs. level 2), a researcher may incorrectly estimate the amount of variance at a level because of not accurately accounting for cross-classification. In turn, this may result in making a false conceptual inference concerning that variable's relationship with other variables in the MLM. Leckie (2013) provides an excellent explanation and starting place for understanding this topic.

The fourth hidden nugget from Peterson et al. (2012) is the topic of centering. Actually, centering was not addressed at all in the article and therefore, it's only a nugget because it should be mentioned. At least within the realm of the organizational sciences, Hofmann and Gavin (1998) were among the first researchers to address the importance of centering when considering multilevel models. An excellent follow-up on that article was provided by Enders and Tofighi (2007).

When discussed in the context of testing interaction hypotheses, centering is used to reduce the non-essential ill-conditioning relationship of the proposed moderator variable and the proposed independent variable when the interaction is represented through the product of those two variables. That is not the reason for discussing centering in the context of the vast majority of MLM tests. Rather, centering can in most circumstances improve the interpretability of the MLM results. Assume, for

example, that we aggregated some level 1 variable to level 2, and then regressed that variable onto some level 2 independent variable such as firm performance. Without centering, we would interpret the intercept as being that value of the aggregated dependent variable when the independent variable is truly zero. However, a zero for firm performance is not a realistic value. If firm performance is grand mean centred first, though, then the intercept is that value of the aggregated dependent variable when firm performance is at its mean. We can then discuss how deviations above and below mean firm performance impacts the aggregated dependent variable. There are also times when group-mean centering improves the interpretability of MLM results, particularly if the level 2 independent variable is aggregated from level 1.

In terms of advancements in MLM since Peterson et al. (2012) was published, there have been great strides in recent years in multilevel structural equation modelling (MLSEM). Vandenberg and Richardson (2019) provide a summary of these advancements. MLSEM's appropriateness for IB research will depend upon the types of measures used to operationalize the underlying constructs. Specifically, MLESM is most useful for models in which some of the variables are measured at the individual level using multi-item scales that are known to validly represent the underlying construct also known as the latent variable. If that is the case, then MLSEM permits the inclusion of measurement error resulting in more accurate outcomes when examining the relationships of the latent variables with other variables in the model. As illustrated by Vandenberg and Richardson (2019), a researcher can use MLSEM to test relatively simple multilevel confirmatory factor analysis models. MLSEM may also be used to test very complex path models consisting of a mix of latent variables with observed variables (e.g., measures from archival sources such as proxy statements, etc.).

Another advancement for IB researchers has been the ease with which a researcher may now test models with more than 2 levels. All of the considerations raised by Peterson et al. (2012) are still valid such as the need to justify aggregation and to also ensure that sample sizes are adequate. Further, centering needs to be carefully considered across the levels. A good primer on MLM with sample syntax from both commercial packages as well as R is provided by Hox, Moerbeek, and Schoot (2017).

In closing, the Peterson et al. (2012) article provides an excellent introduction to those considering the use of MLM for their international business research needs, as well as for researchers in other disciplines (e.g., management). However, MLM even in its simplest form is a complex technique. As such, the uninitiated should not expect to master it without some training and in depth reading on the issues. There are two books in particular that I would highly recommend. Each book is an edited volume with chapters contributed by the leading experts in multilevel modelling. The first volume is Klein and Kozlowski (2000); the second is Humphrey and LeBreton (2019). The second volume is replete with examples, and perhaps most useful is its companion website containing the syntax for many of the examples provided by the contributors.

References

Enders, C.K., and D. Tofighi. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods* 12 (2): 121–138.

Hofmann, D.A., and M.B. Gavin. 1998. Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management* 24 (5): 623–641.

Hox, J.P., M. Moerbeek, and R. van de Schoot. 2017. *Multilevel analysis: Techniques and applications*. 3rd ed. New York: Routledge.

Humphrey, S.E., and J.M. LeBreton. 2019. *The handbook of multilevel theory, measurement, and analysis*. Washington, DC: American Psychological Association.

Klein, K.J., and S.W.J. Kozlowski. 2000. *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions*. San Francisco: Jossey-Bass.

Leckie, G. 2013. Cross-classified multilevel models – Concepts. *LEMMA VLE Module* 12: 1–60. http://www.bristol.ac.uk/cmm/learning/course.html

LoPilato, A.C., and R.J. Vandenberg. 2015. The not so direct cross-level direct effect. In *More statistical and methodological myths and urban legends*, ed. C.E. Lance and R.J. Vandenberg, 292–310. New York: Routledge.

Peterson, M.F., J.L. Arregle, and X. Martin. 2012. Multilevel models in international business research. *Journal of International Business Studies* 43 (5): 451–457.

Vandenberg, R.J., and H.A. Richardson. 2019. A primer on multilevel structural equation modeling: User-friendly guidelines. In *The handbook of multilevel theory, measurement, and analysis*, ed. S.E. Humphrey and J.M. LeBreton, 449–472. Washington, DC: American Psychological Association.

25

Multilevel Models in International Business Research: Broadening the Scope of Application, and Further Reflections

Xavier Martin

Vandenberg (2019) provides a perspective and update of the use of multilevel models (MLM) as explicated by Peterson et al. (2012) for international business (IB) research. In this note, I touch on the implications of the fact that many IB topics involve non-continuous criterion (dependent) variables and thus non-linear models. The illustration pertains to topics in international strategy, but similar issues can be expected with micro-IB research, with the addition of measurement equivalence issues in cross-cultural research (Peterson et al. 2012).

As a starting point, consider a typical MLM specification with subsidiaries (level 1) nested in a parent MNE (level 2). The level 1 may involve repeated measures, which are a common form of multilevel data (Raudenbush and Bryk 2002) and can make a difference to sample

I am grateful for feedback from Jean-Luc Arrègle, Lorraine Eden, Bo Bernhard Nielsen, and Mark Peterson on earlier drafts. All errors remain mine.

X. Martin (✉)
Tilburg University, Tilburg, The Netherlands
e-mail: X.Martin@uvt.nl

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_25

requirement as discussed below. In our illustration, the corresponding level 1 would then be a subsidiary-year observation.

Peterson et al.'s guidelines for "mixed linear" models (ICC, sample sizes) apply as such for a continuous outcome like subsidiary sales or profitability (ratio). However, in practice, many international strategy studies predict versions of discrete variables (Martin and Li 2015; Martin et al. 2007). For example, binary variables commonly describe dichotomous FDI outcomes such as choice between entry modes, or subsidiary survival (Arrègle et al. 2006; Shaver 1998). Ordinal variables can be used to generalize such outcomes, such as the choice among three or more entry modes with increasing levels of commitment, or survey-based outcomes pertaining to below-expected, roughly-as-expected, and above-expected performance (Nielsen 2007). Categorical (nominal) outcomes commonly describe location choices (Wu et al. 2017). Count variables may also describe outcomes such as international experience; their use with MLM in IB is scarce although related examples do exist (Ghosh et al. 2014).

Such outcomes mean that the level-1 random effects cannot be normally distributed, as assumed in standard multilevel models, and the variance of these effects depends on the predicted value (Raudenbush and Bryk 2002). Solutions exist to estimate multilevel models with each of these criterion distributions, which are based on a hierarchical specification of a generalized linear model that generates close approximation to restricted or—better yet—maximum likelihood estimates (Raudenbush et al. 2011; Rodríguez 2008). Each specification involves a link function to transform the level-1 predicted value so it fits into the requisite range 3-level available (Raudenbush et al. 2011; Rodríguez 2008).

However, researchers should be aware of several complications relative to the recommendations of Peterson et al. (2012) and Vandenberg (2019), when modelling non-continuous outcomes. First, if the question pertains to the substantiveness of level-2 (mean) effects, different measures of intraclass correlation are suitable for each distribution (Snijders and Bosker 2012). Second, (quasi) maximum likelihood estimation is sensitive to empty cells or rare cases, that is, to instances where few observations pertain to a given outcome. This is a common issue with IB samples, especially with highly "skewed" distributions of FDI outcomes such

as due to rarer entry modes or countries with few investments. Researchers thus face a dilemma of losing otherwise informative but rare groups (classes) and ascertaining whether there is sufficient power to interpret estimates and obtain stable results (including avoiding false positives).

Simulation analyses to determine suitable sample sizes for non-linear MLM have focused exclusively on logistic regression, implying that the available recommendations are directly relevant to binary outcomes. Upon reviewing these, Hox (2010: 237) concludes "that multilevel logistic models require larger sample sizes than models for normal data, and that the sample size requirements increase when the modeled proportions are close to zero or one." Thus, by contrast with Kreft's (1996) rule of thumb of 30 level-2 groups and 30 level-1 observations for each group for an MLM analysis with a continuous criterion, Moineddin, Matheson, and Glazier (2007) recommend minima of 50 groups and 50 observations per group—nearly a three-fold increase in required overall sample size. Furthermore, they recommend that the group size (i.e., the number of observations per group) should be adjusted for low-prevalence outcomes, such that (at a minimum) both outcomes should be expected to occur at least once in each group. Unfortunately, I am not aware of similar comparisons for other rules of thumb mentioned by Peterson et al. (2012), that is, level-2 predictions, cross-level interactions, and random-effects models.

The results for logistic regression can be taken as partly informative for other dependent variables insofar as the link function for multinomial and ordinal models is the logistic function—although the link function for the ordinal model is a cumulative logit (Raudenbush et al. 2011). (The basic link function for the count model involves a logarithmic transformation, not to be confused with a logit, and also has different generalizations such as for rare events.)

However, even for the multinomial and ordinal models, further challenges arise relative to the binary model. The ordinal model involves extra parameters (to create the cumulative logit) and an assumption of proportional odds. The multinomial model involves comparisons of each category with a reference category, and thus requires multiple level-1 equations and multiple level-2 equations as a function of the total number of categories—typically N-1 of each, though MLM specifications

with extra contrasts exist (Rodríguez 2008). Each of these factors should increase the target numbers of groups and of observations per group, although—subject to further study—the increase should be less than proportional to the extra estimates. Furthermore, with the increase in the number of outcomes, it is more likely that at least one group will have a (very) high proportion of zeros. Authors should be mindful of such imbalances and aim to add observations in groups with low prevalence of non-zero outcomes.

A further complication, which is not unique to non-linear MLM models, is that most power simulation studies use a single predictor (at level 2). Researchers should be mindful that adding hypothesized predictors increases requisite sample size since both point estimates (for fixed effects) and standard errors can be affected by insufficient sample sizes. On the other hand, since most sample size issues pertain to standard errors (Hox 2010; Raudenbush et al. 2011), the use of MLM to control for level-2 variables (such as parent MNE or country characteristics) is relatively benign and researchers should not be reluctant to take advantage of MLM.[1]

Two factors can help researchers overcome the issues resulting from small sample size. First, the sample requirements can be alleviated when measures are repeated (Wang et al. 2011). Specifically, using repeated measures at level 1 reduces the number of unique subjects required; for example, having more subsidiary-year observations makes it possible to run MLM with fewer unique subsidiaries. The use of panel data with a longitudinal cross-section structure, even if the panels are unbalanced, is increasingly common in international strategy research (Martin et al. 2007; Martin and van den Oever 2013). Insofar as some covariates are not time-varying (e.g., certain parent company traits, culture measures, etc.) whereas level-1 variables are time-varying, the use of MLM is warranted. At the same time, repeated measures can enable more appropriate use of MLM by alleviating the sample size issue. A caveat is that adding repeated observations that are "empty," in that they exhibit no or minimal variation in the dependent variable, may worsen the issue of low-prevalence outcomes even as it increases the sample size. Even then, this trade-off is virtuous when it comes to MLM applications to discrete-time event history analysis (Allison 1982; Hedeker 2008; Rodríguez 2008). In

IB research, entry timing and survival are prime examples of outcomes to which this applies (Martin 2013a; Martin et al. 2007).

Second, a set of solutions in the presence of small sample sizes and low prevalence of certain categories is to use bootstrapping and a Bayesian estimator leveraging Gibbs sampling (Hox 2010; Rodríguez 2008). Although the use of Bayesian estimation for MLM in IB is scarce, an example in support of hypothesis testing is Cuypers and Martin (2006).

Unfortunately, of the power analysis software applications reviewed by Peterson et al. (2012), only one specifically accounts for a non-continuous outcome. This "Optimal Design" application also has the advantage of providing sample calculations for up to three levels (Spybrook et al. 2011). However, the feature for non-continuous models only pertains to a binary criterion. Furthermore, it does not accommodate repeated observations. The same is true of the "beta version" of the MLPowSim add-on to MLwiN (Browne et al. 2009), which Peterson et al. (2012) did not mention. Browne et al. (2009) also describe how to use MLPowSim to compute appropriate sample size for a count model, albeit their method is limited to a random effects MLM Poisson specification.

With this caveat, a brief overview of MLM software packages and related user guides may be helpful to researchers who seek to implement MLM with non-continuous criterion variables.

First, all general statistical packages now include multilevel functions, although the depth of application and guide detail for non-continuous MLM ranges from cursory (SPSS: Heck et al. 2014), to partially featured (SAS: Wang et al. 2011), to extensive and dedicated (Stata: Rabe-Hesketh and Skrondal 2012). An excellent guide is Hox, Moerbeek, and van de Schoot (2017), which covers such general packages as well as R in an accessible manner.

Second, three specialized packages stand out for the relative completeness of their features for binary, ordinal, multinomial, and count analyses. HLM benefits from an excellent user guide, which links statistical theory with application (Raudenbush et al. 2011). User resources for MLwiN are more scattered, but the package is notable for its features that allow the user to constrain variances and conduct parametric and non-parametric bootstrap analyses (Hox 2010). Finally, Mplus contains a suite of MLM modules (see also Finch and Bolin 2016), including the

type of structural equation models (MLSEM) path models mentioned by Vandenberg (2019). All three packages allow the construction of models with three levels (and more), although this remains an area where guidelines for sample size are incompletely settled.[2]

The takeaway is that MLM solutions exist to fit the variegated distributions of outcomes typical of IB research. However, scholars should be especially careful about specifying the relationships between levels. This starts with careful theorizing about the nature of these relationships (Martin 2013b, 2014). Furthermore, researchers should sample conservatively (use more cases) when examining complex decisions involving some of the more interesting but scarce outcomes that contribute to the richness of IB research.

Notes

1. I thank Mark Peterson for raising the issues discussed in this paragraph.
2. In particular, guidance for the number of level-2 groups in a three-level model is lacking. This is often critical in designing IB studies, where level 2 may refer to the parent company (which is the typical sample starting point) while level 1 refers to subsidiary or firm-year and level 3 refers to country. I thank Mark Peterson for pointing this out.

References

Allison, P.D. 1982. Discrete-time methods for the analysis of event histories. *Sociological Methodology* 13: 61–98.

Arrègle, J.-L., L. Hébert, and P. Beamish. 2006. Mode of international entry: The advantages of multilevel methods. *Management International Review* 46 (5): 597–618.

Browne, W.J., M.G. Lahi, and R.M.A. Parker. 2009. *A guide to sample size calculations for random effect models via simulation and the MLPowSim software package*. Bristol: University of Bristol.

Cuypers, I.R.P., and X. Martin. 2006. What makes and what does not make a real option? A study of international joint ventures. In *Academy of Management Best Paper Proceedings, Business Policy and Strategy Division*. New York: Academy of Management.

Finch, W.H., and J.E. Bolin. 2016. *Multilevel modeling using Mplus*. Boca Raton: Chapman and Hall/CRC.

Ghosh, A., X. Martin, J.M. Pennings, and F.C. Wezel. 2014. Ambition is nothing without focus: Compensating for negative transfer of experience in R&D. *Organization Science* 25 (2): 572–590.

Heck, R.H., S.L. Thomas, and L.N. Tabata. 2014. *Multilevel and longitudinal modeling with IBM SPSS*. 2nd ed. New York: Routledge.

Hedeker, D. 2008. Multilevel models for ordinal and nominal variables. In *Handbook of multilevel analysis*, ed. J. de Leeuw and E. Meijer. New York: Springer.

Hox, J.J. 2010. *Multilevel analysis: Techniques and applications*. 2nd ed. New York: Routledge.

Hox, J.J., M. Moerbeek, and R. van de Schoot. 2017. *Multilevel analysis: Techniques and applications*. 3rd ed. New York: Routledge.

Kreft, I.G.G. 1996. *Are multilevel techniques necessary? An overview, including simulation studies*. Los Angeles: California State University at Los Angeles.

Martin, X. 2013a. Pre-emption and entry timing. In *Palgrave encyclopedia of strategic management*, ed. D.J. Teece and M. Augier. New York: Palgrave.

———. 2013b. Solving theoretical and empirical conundrums in international strategy research: Linking foreign entry mode choices and performance. *Journal of International Business Studies* 44 (1): 28–41.

———. 2014. Institutional advantage. *Global Strategy Journal* 4 (1): 55–69.

Martin, X., and C. Li. 2015. What do we know about state-owned emerging-economy firms, and how? Evaluating literature about inward and outward multinational activities. In *Advances in International Management*, ed. L. Tihanyi, E.R. Banalieva, T.M. Devinney, and T. Pedersen. Bingley: Emerald.

Martin, X., and K. van den Oever. 2013. Progress, maturity or exhaustion? Sources and modes of theorizing on the international strategy – Performance relationship (1990–2011). In *Advances in International Management*, ed. T.M. Devinney, T. Pedersen, and L. Tihanyi. Bingley: Emerald.

Martin, X., A. Swaminathan, and L. Tihanyi. 2007. Modeling international expansion. In *Research methodology in strategy and management*, ed. D.J. Ketchen and D.D. Bergh. Amsterdam: Elsevier.

Moineddin, R., F.I. Matheson, and R.H. Glazier. 2007. A simulation study of sample size for multilevel logistic regression models. *BMC Medical Research Methodology* 7 (1): 34.

Nielsen, B.B. 2007. Determining international strategic alliance performance: A multidimensional approach. *International Business Review* 16 (3): 337–361.

Peterson, M.F., J.-L. Arrègle, and X. Martin. 2012. Multilevel models in international business research. *Journal of International Business Studies* 43 (5): 451–457.

Rabe-Hesketh, S., and A. Skrondal. 2012. *Multilevel and longitudinal modeling using Stata, (3rd ed), volume II: Categorical responses, counts, and survival*. College Station: Stata Press.

Raudenbush, S.W., and A.S. Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*. 2nd ed. Thousand Oaks: Sage.

Raudenbush, S.W., A.S. Bryk, Y.F. Cheong, and R.T. Congdon. 2011. *HLM 7: Hierarchical linear and nonlinear modeling*. Lincolnwood: Scientific Software International.

Rodríguez, G. 2008. Multilevel generalized linear models. In *Handbook of multilevel analysis*, ed. J. de Leeuw and E. Meijer. New York: Springer.

Shaver, J.M. 1998. Accounting for endogeneity when assessing strategy performance: Does entry mode affect FDI survival? *Management Science* 44 (4): 571–585.

Snijders, T.A.B., and R. Bosker. 2012. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. 2nd ed. Thousand Oaks: Sage.

Spybrook, J., H. Bloom, R. Congdon, C. Hill, A. Martinez, and S.W. Raudenbush. 2011. *Optimal design plus empirical evidence: Documentation for the "Optimal Design" software*. New York: William T. Grant Foundation.

Vandenberg, R.J. 2019. Multilevel models in international business research: A commentary. In *Research methods in international business*, ed. L. Eden, B.B. Nielsen, and A. Verbeke. New York: Springer.

Wang, J., H. Xie, and J.H. Fisher. 2011. *Multilevel models: Applications using SAS*. Boston: De Gruyter.

Wu, Z., R. Salomon, and X. Martin. 2017. Complex strategic choices: A new approach with an application to adaptive agglomeration strategies among foreign entrants. *Global Strategy Journal* 7 (3): 286–311.

Part XII

Distance

26

Conceptualizing and Measuring Distance in International Business Research: Recurring Questions and Best Practice Guidelines

Sjoerd Beugelsdijk, Björn Ambos, and Phillip C. Nell

Introduction

Distance may well have become an international business research workhorse (Salomon 2016; Verbeke et al. in press), but the distance construct as well as its operationalization are continuously being debated in practice (Kirkman et al. 2006; Shenkar 2001; Zaheer et al. 2012). That debate can be quite impassioned. Some find the use of a composite cultural distance index appropriate (e.g., Cuypers et al. 2018), while others reject this outright as a vestige of the "dark middle ages" of cross-cultural

S. Beugelsdijk (✉)
University of Groningen, Groningen, The Netherlands
e-mail: s.beugelsdijk@rug.nl

B. Ambos
University of St. Gallen, St. Gallen, Switzerland
e-mail: bjoern.ambos@unisg.ch

P. C. Nell
Wirtschaftsuniversität Wien, Vienna, Austria
e-mail: phillip.nell@wu.ac.at

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_26

research (Tung and Verbeke 2010: 1270). Similar debates exist around other aspects of distance research. Some of these debates seem unresolved and complex, and it is our experience that authors, reviewers, and editors respond to these issues differently.

In this editorial, we address recurring disagreement on theory, methods, data, and the relationship between different distance dimensions, complementing and updating existing editorials (Beugelsdijk and Mudambi 2013; Zaheer et al. 2012), commentaries (Brouthers et al. 2016; Van Hoorn and Maseland 2016), debates (Cuypers et al. 2018; Maseland et al. 2018), and surveys of distance research (Berry et al. 2010; Hutzschenreuter et al. 2014; Shenkar 2001; Tung and Verbeke 2010). We then make a series of recommendations which we believe will help achieve convergence in research practice.

Our recommendations center on: (1) how to theorize on distance, and (2) what method and (3) what data to use to calculate a distance index. Where relevant and possible, we support our argument by leveraging all available country–pair data on the most used distance dimensions. While we discuss distance in a broad sense, we illustrate our points predominantly by drawing on cultural and institutional distance. Nonetheless, we think that our reasoning and recommendations are relevant for other types of distance. Our goal is to set out a disciplined approach to crafting and reviewing distance studies in a positive and constructive way. We summarize and elaborate on current practices, explain the nature of the debate regarding distance, and, where we can, provide best practice guidelines. The data that we use are available on the website of the *Journal of International Business Studies* for replication and extension purposes.

Distance Research in International Business

In principle, distance can be measured between any two entities, be it individuals, teams, organizations, nations, ethnic groups, language groups, even organizational fields. In most international business and management research, the distance measured is between countries. Zaheer et al. (2012: 19) define international management as "management of distance". The attractiveness of distance is rooted in its literal

meaning related to geographic or physical distance, and its metaphorical one (Shenkar 2012) referring to "the collective differences between countries" (Zaheer et al. 2012: 20). The importance of country as a unit of analysis also applies to psychic distance, which can be defined as the perceived distance that individuals or groups hold regarding a particular country (Dow and Karunaratna 2006; Håkanson et al. 2016).

The origins of the inter-country distance concept can be traced back to early work in international economics. Geographic distance plays an important role in the gravity models commonly used in classic and modern trade theory to explain trade flows between countries (Anderson and van Wincoop 2003; Bergstrand 1985; Feenstra et al. 2001). Beckerman (1956) suggested that psychic distance can partly explain intra-European trade, thereby extending the meaning of distance beyond its geographical dimension. Interestingly, the empirical evidence for Beckerman's 1956 claim that distance perceptions explain trade flows is relatively recent (Dow and Karunaratna 2006; Guiso et al. 2009; Håkanson 2014; Yu et al. 2015).

Some two decades would pass before Beckerman's concept of psychic distance would be used in international business research. A series of studies published by the Uppsala school (e.g., Johanson and Vahlne 1977; Johanson and Wiedersheim-Paul 1975) argued that psychic distance significantly influences the location choice and internationalization paths of firms. In one of those studies, Johanson & Wiedersheim-Paul (1975: 308) define psychic distance as the "sum of factors preventing or disturbing the flow of information between firm and market" (see also Vahlne and Wiedersheim-Paul 1973),[1] a definition that has become a classic in distance research (Dow and Karunaratna 2006), and is echoed by Zaheer et al.'s (2012: 20) definition of psychic distance as the "collective differences between countries".

To operationalize their construct, Vahlne and Wiedersheim-Paul (1973) created a compound measure including characteristics of the target market such as GDP per capita, educational level, differences between the home country (in their case, Sweden) and the host country, including language and culture, as well as trade relations measured by the relative level of imports as a further proxy for information flows.[2] Johanson, Vahlne, and Wiedersheim-Paul's contributions, although

seminal, introduced ambiguity in transferring what was originally a perceptual measure that complemented the cost of geographical distance into a measure of objective differences between trading partners (Håkanson and Ambos 2010). Although the psychic distance construct as such is generally accepted in international business research and practice, Håkanson and Kappen (2017) assert that the theoretical predictions of the associated Uppsala school of internationalization lack robust empirical support.

A decade later, Kogut and Singh (1988) introduced a composite cultural distance index based on the country scores for the four national cultural dimensions developed by Hofstede (1980) and Cuypers et al., (2018). They considered their cultural distance index to be in many ways similar to the psychic distance index of the Uppsala school (Kogut and Singh 1988: footnote 10). In their original article, the authors were also very explicit about the internal validity of their measure, acknowledging that:

> The indices of Hofstede can be criticized for a number of reasons, especially regarding the internal validity of the dimensions and the method of constructing the scales. Whereas the criticism has a sound basis, Hofstede's study has some appealing attributes, namely, the size of the sample, the codification of cultural traits along a numerical index, and its emphasis on attitudes in the workplace. (Kogut and Singh 1988: 422)

In the years that followed, the Kogut and Singh (1988) index was widely adopted, in part because of the ease of calculating it and in part because of increasing use of secondary datasets in international business research (Cuypers et al. 2018). The cultural distance index, which originated as a psychological complement to geographical distance, has become the de facto standard instrument to measure distance in international business studies. Over time, the index has turned into a quasi-objectified, single measure of differences between internationally distant actors (Ambos and Håkanson 2014).

The Kogut and Singh (1988) index has been the subject of serious conceptual and methodological criticism (Kirkman et al. 2006, 2017; Shenkar 2001, 2012; Maseland et al. 2018). Shenkar (2001) identified a

set of weaknesses plaguing it, yet a decade later he would note that a large majority of studies simply cited his criticism of the index in order to "acknowledge" the problem, then went on to use it without any further discussion (Shenkar 2012).

There have nevertheless been efforts to address problems with the index; for instance, attempts to introduce additional dimensions. Kostova (1996) developed institutional country profiles to ground the concept of institutional distance (e.g., Eden and Miller 2004; Xie and Li 2017; Xu and Shenkar 2002). Ghemawat (2001) introduced the CAGE framework, referring not only to cultural but also to economic, geographic, and administrative distance (Nell and Ambos 2013; Mingo et al. 2018). Others have proposed new metrics to calculate the index (Berry et al. 2010), created new databases with additional dimensions (Dow and Karunaratna 2006), and developed perception-based psychic distance measures (Håkanson and Ambos 2010). Collectively, these advances have been valuable, but there are still many unanswered questions raised by authors, reviewers and editors.

We have organized our discussion of distance research around three questions, for which we provide theoretically-grounded practical recommendations: (1) How should a theoretical framework on distance (e.g., distance in general vs. distance on a specific dimension) be constructed? (2) What method should be used to calculate a distance index, specifically, does (co-) variance correction (e.g., using Mahalanobis distance) affect the results, and if so, how? And (3) what data should be used to construct a distance index, and specifically, does it matter if one chooses a particular dataset of cultural (Hofstede, Schwartz or Globe) and institutional dimensions (e.g., Quality of Governance, Economic Freedom Index, or International Country Risk Guide)?

How to Think Theoretically About Distance

There is no such thing as a general distance theory in the sense of a single, internally consistent set of assumptions, mechanisms, and boundary conditions, but the lack of a single distance theory is no reason to stop exploring the meaning of distance in international business. Similarly, there is

no grand theory of national culture, but this has not stopped scholars from investigating the relevance of national culture for international business.[3] Distance is a construct and as such is meaningful only in the context of a specific theory, be it, for example, agency theory, transaction costs theory, or learning theory. This may appear obvious, but it is important to make it explicit as it has several important implications for theory development.

Distinguishing Between Geographical and Contextual Distance

As alluded to earlier, we see the concept of distance as the joining of two essential elements of doing business across borders. The first one is the geographic distance between two or more locations. Narrowly defined, geographic distance is the distance between two points on the surface of the earth, as given by latitudinal and longitudinal coordinates. In keeping with this definition, geographic distance has three properties: it is (1) symmetrical (i.e. the geographic distance between countries A and B is the same as that between countries B and A), (2) continuous, and (3) stable over time.

Second, distance refers to the change in context that occurs when firms cross national borders. In this case, distance serves as a metaphor for the kinds and varying degrees of differences in context (Shenkar 2012). In contrast to geographic distance, contextual distance can be (1) asymmetric, (2) non-continuous, and (3) it can change over time. Shenkar (2001) makes the point that the distance from one country to another may be asymmetric (e.g., between a country with a low level of economic development and one with a high level) and that this has implications for internationalizing firms. Similarly, psychic distance research has shown that the perceived distance between country A and country B may be different from that between B and A (Shenkar 2001; Håkanson and Ambos 2010; Håkanson et al. 2016). Contextual distance may also be non-continuous, as it is subject to border effects (Beugelsdijk and Mudambi 2013). National borders are powerful discrete delineators of context (Peterson et al. 2018).[4] Finally, contextual distance can change over time

as countries go through economic, institutional, and cultural change. We will return to this point later, but note here that many distance dimensions have been relatively stable over time.

Spelling Out the Mechanisms

In addition to distinguishing between geographic and contextual distance, the relationship between theoretical argument and the distance construct must be made explicit (see also Zaheer et al. 2012; Maseland et al. 2018). Distance may have a different meaning in learning theories (Stahl and Tung 2015) as compared to agency and transaction cost theories (Kostova et al. 2017; Nell et al. 2017; Shenkar et al. 2008). While learning theories would emphasize that doing business in a different context can stimulate creativity, agency theory and transaction costs theory would highlight the increased uncertainty, and the potential for misunderstandings. Similarly, a large economic distance may give rise to additional costs, as products and business models must be adapted, but it may also generate arbitrage opportunities (Ghemawat 2001). Finally, if the argument is that managers prefer to enter countries that are relatively similar to the home country, then perhaps psychic distance (that captures overall perceived dissimilarity) might be the more appropriate construct. Hence, the functional role of distance depends on the type of distance (the specific distance construct) and the theoretical context in which it is used.

Unfortunately, explicit theorizing on the channels through which geographic and/or contextual distance affects outcomes is often underspecified, even missing altogether. For instance, while the standard assumption – whether implicit or explicit – in many studies is that distance results in costs, why that may be true is seldom discussed (Beugelsdijk et al. 2018; Maseland et al. 2018; Zaheer et al. 2012).

Distance as an Aggregate Construct

Given the multidimensional nature of contextual differences, the question is whether distance should be measured on a specific dimension or as a composite construct. This discussion applies specifically to cultural

distance as an index based on the distance between the home and host country on multiple cultural dimensions. In the original Kogut and Singh (1988) index, four of Hofstede's cultural dimensions were included (individualism–collectivism, power distance, uncertainty avoidance, and masculinity–femininity).

Some have argued against the use of composite distance indices, such as the Kogut and Singh index (Dow and Karunaratna 2006; Kirkman et al. 2006), as the dimensions included are conceptually different, and aggregating them makes the composite index imprecise and noisy. Using instead the home–host distance on individual dimensions (for example, the distance on Hofstede's individualism dimension) allows for more precise theorizing. Others counter that the use of composite indices such as the cultural distance index is valuable, composite indices being more tractable and connecting well to prior research (Cuypers et al. 2018). It may be a moot point as the Kogut and Singh cultural distance index has become the de facto research standard, and, as observed by Shenkar (2012), many distance studies simply continue to use it because it has been used before.

In our view, whether one should use a composite index or one based on an individual dimension, hinges on the nature of the theoretical argument. Such an approach is unfortunately often lacking in current research practice (for an exception on aspects of cultural distance, see Dikova et al. 2010; for an exception on aspects of institutional distance, see Abdi and Aulakh 2012). Frequently, a model using individual dimensions of distance is used as a robustness test for a model with composite distance indices (or vice versa) without any discussion of the theoretical implications. We contend that a composite index is required when the nature of the theoretical argument has to do with distance in general. When the theoretical argument involves a specific dimension, for example a difference in degree of individualism, then the authors should address why it matters. For instance, it does not make sense to replace a composite index of distance by one based on individual cultural dimensions if one aims to study the effect of overall cultural distance on the frequency of knowledge exchange among MNE subsidiaries.

An illustration of the difference between theorizing on distance in general or on a specific dimension can be found in Kogut and Singh (1988).

In their study on culture and entry mode choice, they examined the effect on entry mode choice of both overall cultural distance and of uncertainty avoidance (one of the cultural dimensions included in their distance index).[5] Their first hypothesis builds on the logic that increased cultural distance between two country pairs will foster uncertainties and thus affect entry mode choice. This is a classic argument about overall distance. In their second hypothesis, they suggest that firms from countries that score high on uncertainty avoidance will choose a different entry mode than firms based in countries that score low on that dimension. This example underscores that: (1) arguments on the impact of distance in general will differ from those on a specific dimension of it, and (2) theorizing on individual dimensions will typically require specifying the direction of the effect, in this case from high to low uncertainty avoidance (see Hennart and Larimo 1988, for an example of how power distance in the home country affects entry modes in the host country).

Consistency of Distance Effect Assumptions

More precise theorizing on the mechanisms through which distance affects outcomes is required, as many of the outcomes studied in distance research are the result of multi-stage decision-making processes. In many distance–performance studies, for example, it is argued that distance leads to lower levels of MNE subsidiary or MNE parent performance because it results in a liability of foreignness, and hence in higher costs of doing business abroad (Beugelsdijk, Nell, and Ambos 2017). However, should we not expect distance to have influenced location and entry mode decisions in the first place (Brouthers 2002)? It is inconsistent from a conceptual perspective to argue that distance affects performance but has no impact on location decisions.

Another example of multi-stage decision-making applies to studies of the impact of distance on the choice of foreign market entry modes. Following transaction costs logic, it is often argued that contextual distance between home and host country will discourage entry with high commitment modes (such as a wholly owned subsidiary) because distance makes access to information and its interpretation more dif-

ficult (Morschett et al. 2010). However, multinational firms often develop a portfolio of activities in different countries, and the distance between the home country (where the headquarters is located) and the host country of a new foreign entry may not be the most relevant distance with which the multinational firm has to cope (Hendriks et al. 2017). In this case, what matters may be the "added distance", i.e., the distance between the country of the new foreign entry and the closest country in which the firm is already active (Hutzschenreuter et al. 2011).

These distance–performance and distance–entry mode examples highlight the need to develop a theoretical logic that explains how distance-induced costs and benefits affect the different stages of the firm's internationalization process.

Recommendation

If we want to make progress, recognizing the need for an explicit discussion of the mechanisms through which a particular type of distance (be it geographic or contextual) affects the phenomenon of interest would appear paramount. This requires taking into account the multi-stage nature of the decision-making process associated with firm internationalization, as well as a careful evaluation of whether the distance we want to study is symmetric (in the case of geographic distance) or asymmetric (in the case of contextual distance). We urge authors to make sure their chosen distance construct is aligned with their theorizing, and not just added to a regression model for convenience purposes. Without such an argumentation, adding a distance variable becomes a pointless exercise. Hence, "the use of the aggregate index must be theoretically justified and where appropriate, substituted by cultural distance measures calculated separately for one or more of the five dimensions *as necessitated by theoretical and domain considerations*" (Shenkar 2001: 529). The italics are ours, as we want to underscore that one should not use individual dimensions of distance being conceptual equivalents of a composite measure.

How to Construct a Distance Index

Kogut and Singh's (1988) cultural distance index (KSIndex) has become a 'must have' variable in international business and management research (Shenkar et al. 2008: 908). It calculates national cultural differences by the composite difference on a set of cultural dimensions (see Kogut and Singh 1988: 422):

$$\text{KSIndex}_j = \sum_{i=1}^{n} \left\{ \left(I_{ij} - I_{iHOME} \right)^2 / V_i \right\} / N \tag{26.1}$$

where I_{ij} refers to the host country j's mean score on Hofstede's ith dimension, Ii_{HOME} to the home country's mean score on this same dimension, V_i to the variance of the ith dimension and N to the number of dimensions. The distance index can be calculated for any multidimensional construct. As we discuss in detail below, the cultural dimensions need not be taken from Hofstede but can also be derived from the Schwartz (1994, 1999, 2006) or Globe (House et al. 2004) cultural frameworks. The Kogut and Singh formula has been used to operationalize other types of distance besides cultural distance (e.g., regulatory distance in Wu and Salomon 2016; institutional distance in Campbell et al. 2012).

The Kogut and Singh index belongs to the family of Euclidean distance metrics. Kogut and Singh applied the Euclidean distance metric to measure *national cultural* differences, but it can be applied to other units of analysis (teams, firms, or subunits). The Kogut and Singh index represents an adaptation of the standard Euclidean method of calculating a composite distance index on a set of individual dimensions. The Euclidean distance between a home country and country j on an i-dimensional construct I is calculated as follows[6]:

$$\text{EuclidenDist}_j = \sqrt{\sum_{i=1}^{n} \left(I_{ij} - I_{iHOME} \right)^2} \tag{26.2}$$

Taking Care of Variance Differences

The key difference between Eqs. 26.1 and 26.2 is the correction for differing variances across the dimensions, because one "problem with Euclidean distance is that it does not take into account the variance of the [individual] variables" (Berry et al. 2010: 1469). That is, the Kogut and Singh index is a Euclidean distance with variance correction. In addition, Kogut and Singh divided the overall distance by the number of dimensions, while the Euclidean distance formula takes the square root of the overall difference.

The Kogut and Singh index and the Euclidean distance index are often presented as alternatives, and therefore used in robustness tests (e.g., Barkema and Vermeulen 1997; Drogendijk and Slangen 2006).[7] Below, we illustrate the relationship between these alternative indices for two of the most used distance constructs: (1) cultural distance, and (2) institutional distance.

There are three cross-cultural frameworks used in the management literature. A summary of the key characteristics of each, including their dimensions, can be found in "Appendix A". The first is Hofstede's (1980, 2001) well-known cultural framework. The first version consisted of four cultural dimensions, to which an additional two were later added (Hofstede et al. 2010). Whereas the first four dimensions were derived from surveys of IBM employees carried out between 1968 and 1972, the latter two are based on a set of six questions from the World Values Survey–European Value Studies (WVS-EVS).[8] Berry et al. (2010) and Beugelsdijk et al. (2015) have used the WVS-EVS data to develop Hofstede-inspired dimensions.[9] Schwartz (1994, 1999, 2006) developed the Schwartz Value Survey, which consists of seven national cultural value orientations. House et al. (2004) developed nine national cultural dimensions for values and for practices, a framework commonly referred to as Globe.

All these culture frameworks can be used to compute cross-country cultural distances. We discuss below whether using the Kogut–Singh formula or the Euclidean distance formula yield radically different results. We calculate the two indices for all country pairs for which data are available. Table 26.1 shows that the correlations between the Kogut and

Singh index (Eq. 26.1) and the Euclidean distance index (Eq. 26.2) are very high, ranging from 0.89 (Globe) to 0.97 (Hofstede's six dimensions).[10]

As mentioned earlier, additional distance constructs have been developed to complement cultural distance, often measured by applying the Kogut and Singh approach. One of these is institutional distance (Eden and Miller 2004; Malhotra and Gaur 2014; Xie and Li 2017; Xu and Shenkar 2002), which has been measured using a variety of databases, including the Quality of Governance database (QoG; also referred to as World Governance Indicators) developed by the World Bank (e.g., Kaufmann et al. 2008; Abdi and Aulakh 2012; Ang et al. 2015; Campbell et al. 2012; Hutzschenreuter et al. 2014; Li et al. 2014; Salomon and Wu 2012), the Economic Freedom Index (EFI) provided by the Heritage Foundation (e.g., Demirbag et al. 2011; Gubbi et al. 2010; He et al. 2013), and the International Country Risk Guide (ICRG) developed by the Political Risk Services group (e.g., Makino and Tsang 2011; Valentino et al. 2018).

Table 26.1 Pairwise correlation between Kogut–Singh index (*KSI*) and Euclidean distance (*ED*)

Distance construct	Database	Correlation between KSI and ED	Number of country pairs	Number of countries
Cultural distance	Hofstede 4 dimensions	0.96	4830	70
	Hofstede 6 dimensions	0.97	3782	62
	Schwartz	0.94	4970	71
	Globe-values	0.89	3306	58
Institutional distance	Quality of governance (QoG)	1.00	38,612	197
	Economic freedom index (EFI)	0.95	28,390	169
	International country risk guide (ICRG)	0.96	19,182	139

Pairwise correlations between the Kogut and Singh index (Eq. 26.1) and the Euclidean distance (Eq. 26.2) for different distance constructs. For QoG, EFI, and ICRG, we used the 2013 scores. Correlations are based on unique country pairs (home–home combinations are excluded)

The QoG data consist of six dimensions: rule of law, control of corruption, government effectiveness, voice and accountability, political stability, and regulatory quality. The World Bank calculates standardized country scores for these six dimensions; hence re-scaling the dimensions by correcting for variance differences is not required; however, re-scaling may matter for the EFI-based distance index. The EFI consists of ten indicators of property rights protection, corruption levels, fiscal freedom, government spending, and a set of six indicators measuring freedom of doing business, trade, finance, and investment. EFI scores are not standardized. Nevertheless, the correlation between the Kogut–Singh and the Euclidean versions of this index is 0.95 (see Table 26.1). The ICRG consists of 12 dimensions related to government and political stability, levels of socio-economic development, conflict and corruption, and religious and ethnic tensions. As shown in Table 26.1, the Kogut and Singh index using ICRG dimensions correlates 0.96 with the Euclidean version.

The need to re-scale and correct for the variance differences between the dimensions included in a distance index depends on the data used. Both for cultural and institutional distance, re-scaling matters little. For the QoG-based institutional distance index, it does not matter at all. The high correlations between distance indices applying variance correction or not (Kogut–Singh vs. Euclidean) have implications for the interpretation of the results of distance studies. Given these high correlations, we would not expect results to differ substantially between studies using these two methods (all else being equal).

However, one important clarification needs to be made. It is not always clear whether researchers use the variance of a dimension that is available for all country data, or the variance within a dataset that consists of only a subsample of countries (e.g., only European countries or only dyads between Germany and any other country). For Table 26.1, we have used the variance as based on all available countries. Obviously, the choice of what variance to use can have a significant influence on the final distance index. From a theoretical perspective, we think it would be best to correct for the variance that is considered relevant for the firms under consideration. In most cases, however, we do not know the actual country exposure of a firm (either because we do not know the portfolio of countries in which a firm has invested, or because we do not know the

countries the firm has possibly considered for a location choice decision), and it may thus be most practical to use the variance of all available data. We would urge authors to be transparent in this regard.

Recommendation

For the cultural and institutional distances that we computed, rescaling only has a small impact on the resulting index. For these indices, it does not matter much whether we use the Kogut and Singh index or the Euclidean distance index. Yet, this could be different for other types of distance. We generally recommend that researchers re-scale individual dimensions of distance – especially when there are substantial differences in variance across dimensions – and that they are transparent about which variance is used when doing so.

Taking Care of Co-Variance

In addition to the need to correct for variance differences across dimensions, a second concern with Euclidean approaches is that they disregard potential correlations between the individual distance dimensions. Shenkar (2001) pointed out that correlated dimensions may exert an undue influence on the final index. The most frequently used methods to correct for co-variance across distance dimensions is the Mahalanobis index (Mahalanobis 1937). The popularity of this method has grown since Berry et al. (2010) introduced it to the international business field.

The Mahalanobis approach takes the full variance–co-variance matrix into account when computing distance between country pairs. As Berry et al. (2010) note, Mahalanobis' technique is especially interesting when the dimensions included in the distance index are measured on a different scale (e.g., GDP per capita and inflation rates). This argument is less relevant to cultural and institutional distance because both are commonly measured using similarly scaled dimensions (e.g., the Hofstede dimensions and the EFI dimensions are measured on a 0–100 scale, and the QoG dimensions are standardized).

The Mahalanobis distance is frequently misunderstood, perhaps because the technique itself is relatively complex. Often, Mahalanobis distance is perceived to be the most advanced or the best technique to create a composite index (Flury and Riedwyl 1986), but this is not necessarily true (Brereton and Lloyd 2016). When the individual distance dimensions included in the index are totally uncorrelated, the resulting Mahalanobis index is perfectly correlated with a variance-corrected Euclidean index (De Maesschalck et al. 2000). In this case, applying a Mahalanobis technique and correcting for the covariance does not add value.

The Mahalanobis index also has no added value when all the dimensions are very highly correlated with each other (Brereton and Lloyd 2016). For example, the correlations between the six QoG indicators range between 0.62 and 0.94.[11] A principal component factor analysis on these six indicators shows that they reflect one single construct explaining 86% of the variation across the six indicators. Given these very high correlations, it makes sense to use the factor score and to measure institutional quality as one single reflective construct (e.g., Lavie and Miller 2008; Zaheer and Hernandez 2011; Klopf and Nell 2018).

Thus, Mahalanobis' technique becomes relevant when there is a mix of high and low correlations between the indicators included. Under these circumstances, it may – albeit not necessarily – yield quite different results as compared to Euclidean approaches.

The six Hofstede dimensions, as well as the Schwartz and Globe dimensions, show such a mix of correlations (see "Appendix B"). While the QoG indicators are highly correlated, this does not hold for the 10 EFI dimensions (range between 0.01 and 0.92) and the 12 ICRG dimensions (range between 0.02 and 0.80) (see "Appendix C"). Table 26.2 compares the (variance-corrected) Euclidean distance index with the Mahalanobis distance index for cultural and institutional distance using alternative databases.

The Euclidean (four dimensional) Hofstede-based cultural distance correlates 0.88 with the Mahalanobis Hofstede-based cultural distance. For Hofstede's six-dimensional model, this correlation is 0.84. Using alternative culture frameworks, we find that the correlation between the Euclidean distance and the Mahalanobis distance drops to 0.58 (Schwartz)

and 0.72 (Globe). For the EFI-based institutional distance index, we find a correlation of 0.62, and for ICRG this correlation is 0.58.

Table 26.3 shows the correlations between QoGbased institutional distance constructs using Euclidean, Mahalanobis, and factor score techniques. The correlation between the Euclidean and Mahalanobis construct is only 0.40. The correlation between the Euclidean and the factor score using the first principal component of all six QoG indicators is 0.97.

Table 26.2 Pairwise correlation between Euclidean distance (*ED*) and Mahalanobis distance (*MD*)

Distance construct	Database	Correlation between ED and MD	Number of country pairs	Number of countries
Cultural distance	Hofstede 4 dimensions	0.88	4830	70
	Hofstede 6 dimensions	0.84	3782	62
	Schwartz	0.61	4970	71
	Globe-values	0.73	3306	58
Institutional distance	Economic freedom index (EFI)	0.62	28,390	169
	International country risk guide (ICRG)	0.58	19,182	139

Pairwise correlations between Euclidean distance and Mahalanobis distance for different constructs. The Euclidean distance is variance corrected. For EFI and ICRG, we use the 2013 scores. Correlations are based on unique country pairs (home–home combinations are excluded)

Table 26.3 Pairwise correlation between Euclidean distance (*ED*), Mahalanobis Distance (*MD*), and distance on first principal component of QoG indicators

Distance construct	Database	Correlation between ED and MD	Correlation between ED and distance on first principal component
Institutional distance	Quality of governance (QoG)	0.40	0.97

Euclidean distance is variance-corrected. We use the 2013 scores for QoG. Correlations are based on unique country pairs (home–home combinations are excluded). Number of countries is 197.

The discussion on co-variance correction relates to the literature on index construction methods, and the distinction between formative and reflective constructs (Bollen and Diamantopoulos 2017; Coltman et al. 2008; Diamantopoulos et al. 2008; Diamantopoulos and Winklhofer 2001). Starting with the Kogut and Singh index (1988), cultural distance has been treated as a formative construct based on the four individual dimensions. The Mahalanobis approach continues this tradition as it essentially represents a formative approach to index construction. QoG, however, has been interpreted more as a reflective construct, whereby the latent institutional distance variable is reflected by all the individual dimensions (Lavie and Miller 2008; Zaheer and Hernandez 2011; Klopf and Nell 2018; Slangen and Beugelsdijk 2010).

We do not argue here that correcting for covariance by using Mahalanobis' approach is wrong. However, researchers should be aware that it represents a formative approach based on a given number of dimensions. It is debatable whether cultural distance and other distance constructs are theoretically of a formative nature or a reflective nature, or whether there is a more complex factor structure where both formative and reflective aspects are present. We think that highly aggregated constructs, such as distance constructs, often possess characteristics of reflective as well as formative constructs, a common phenomenon in the field of index construction (Bollen and Diamantopoulos 2017), but which has been neglected in most distance research. In fact, the six cultural dimensions developed by Hofstede are already based on a factor analytic procedure, using the original survey questions based on a reflective logic. The Hofstede-based Mahalanobis distance index thus already represents a complex factor structure with formative and reflective elements.

The relatively high correlations between the Hofstede indices, whether or not applying covariance correction, have implications for the interpretation of results of distance studies. Given these high correlations, we would not expect results to differ substantially between studies using these two methods (all else equal). In fact, meta-analysis of cultural distance and its relation to firm performance shows that there is no significant difference between the results obtained with the Hofstedebased Kogut and Singh index or with its Mahalanobis equivalent (Beugelsdijk et al. 2018). Yet, we do not know whether this result can be generalized

to other cultural or institutional distance indices. The correlations shown in Tables 26.2 and 26.3 give reason for concern.

Recommendation

We think that Mahalanobis' approach is valuable for correcting potential co-variance between the dimensions. Our analysis of the three most used cultural and institutional distance indices shows that co-variance correction matters, but need not yield radically different distance indices. In the case of Hofstede-based cultural distance, using Mahalanobis' approach does not fundamentally alter the index as compared to a Euclidean approach and can therefore safely be ignored. For the other distance measures used here, co-variance correction matters more and should thus be carefully examined. We recommend that scholars be transparent about their approach to co-variance correction. We also think that more research is needed on leveraging different, more complex index construction methods using structural equation modeling techniques, and that researchers should explain more clearly whether they want to treat distance as a formative or a reflective construct.

Data Selection Challenges for Building a Distance Index

We now turn to two data-related questions. First, how does one add a dimension to an already existing multidimensional framework? For instance, Hofstede et al. (2010) added two dimensions to the original four of the framework. Second, how does one handle alternative, competing databases to operationalize cultural or institutional distance? As discussed earlier, alongside Hofstede (1980, 2001), the Schwartz (1994, 1999, 2006) and Globe databases (House et al. 2004) provide country scores on a set of cultural dimensions. All three frameworks measure cultural variation across countries. Similarly, institutional distance has been measured using the QoG, EFI and ICRG databases, which all measure variation in institutional quality across countries.

All these frameworks provide country measures of cultural dimensions or institutional quality. Should they be treated as substitutes such that a cultural or an institutional distance index based on Hofstede or QoG data can be used as a robustness test for a cultural or an institutional distance index, based on Schwartz/Globe or ICRG/EFI data, respectively? If the frameworks cannot be considered substitutes, how does one support the choice of a specific one? If two frameworks provide country scores for the same conceptual dimension but in slightly different ways, can they (or should they) be combined to generate one composite distance index? We address these practical questions below.

Additional Dimensions

Some composite distance constructs, such as the six-dimensional institutional distance index based on the QoG indicators, have been based on the same set of six dimensions since their inception. Moreover, as noted above, the six QoG indicators are so highly correlated that leaving one out would not have a significant impact on the index.[12] This is not necessarily the case for other databases, and that includes the most often used cultural distance index based on Hofstede's data.

Hofstede and co-authors have added "Long term orientation" and "Indulgence versus restraint" to the original four dimensions (Hofstede et al. 2010). The first four dimensions are based on data collected from 1968 to 1972, while the additional two use more recent data from the 2000s. Long-term orientation and indulgence versus restraint are moderately correlated ($r = -0.43$), and not highly correlated with the original four dimensions (see "Appendix B"). This raises the question of whether the six-dimensional index is preferable to the four-dimensional one, and whether adding two dimensions makes a difference. The correlation between the (Mahalanobis) four- and six-dimensional constructs is high at 0.81 (it is 0.82 using the Euclidean distance), so adding the two new dimensions has little impact.

Although the high correlations between the four- and six-dimensional Hofstede frameworks suggests that adding dimensions is not very likely to yield radically different results, it does raise the more fundamental

question of whether adding these two dimensions is theoretically and methodologically sound in the first place. Hofstede's four-dimensional framework has been fiercely criticized, with detractors questioning the representativeness of his sample, the face validity of the questions, the labelling of the dimensions, the treatment of individualism and power distance as two separate dimensions, and the usefulness of a framework developed more than 40 years ago (see, e.g., Ailon 2008; Baskerville 2003; Brewer and Venaik 2011; McSweeney 2002, 2009; Smith et al. 1996; Oyserman et al. 2002).

In our view, whether these additions make sense depends to a large extent on whether one considers the framework in its totality (i.e., a set of cultural dimensions shaping behavior), or whether one is interested in cross-country *distance* indices. When considering the former, one should take into account that the first four dimensions are based on survey data different from those used for the two additional dimensions. Hofstede did not use the methodology one would use today to develop cultural dimensions, nor follow item selection procedures, nor apply factor analysis to all items associated with all six dimensions. Unfortunately, it is not possible to go back and do this today, as the original raw data are no longer available.[13] Importantly, a factor analysis on the six dimensions (not the underlying items) yields three factors (see "Appendix D"). Had Hofstede used factor analysis on both IBM and WVS questions, it is highly unlikely that he would have settled on a six-dimensional framework.

From a distance perspective, however, some of the critical comments raised against Hofstede (and also against Schwartz and Globe) need not be problematic because they are not really relevant. For example, the discussion of whether the labels reflect the underlying items matters when giving substantive meaning to specific dimensions, which is not the case when all dimensions are collapsed into a cultural distance index (but, of course, labels do matter when theorizing on the distance on a specific cultural dimension). Similarly, while cultural indicator *levels* can change, this may not affect cultural *distance.* Cultures change, but the available evidence suggests that many countries tend to move in the same direction towards becoming more individualistic, less power distant, and more indulgent and emancipative (Beugelsdijk and Welzel 2018; Fernandez

et al. 1997; Inglehart 1997; Inglehart and Baker 2000; Inglehart and Welzel 2005).[14] This leaves the cultural distance between countries relatively constant. Hence, 1970s' culture scores would no longer be representative when used in terms of absolute levels, but they may still be useful as input in a cultural distance index (Beugelsdijk et al. 2015; Beugelsdijk and Welzel 2018).

Finally, whether a particular culture framework is useful also depends on whether it is deployed as an integrated and internally consistent set of cultural dimensions to analyze, explain, and predict how particular cultural values affect specific outcomes, or whether the dimensions associated with a framework are used as input in a cultural distance index. When reflecting on the Hofstede's framework, the correlated nature of the additional dimensions may be somewhat problematic when looking at the framework in its totality, but the addition of two dimensions is less problematic from a distance perspective, given the correlation of 0.8 between the four- and six-dimensional frameworks.

Recommendation

Assuming that the research question addressed calls for a composite index, as opposed to an individual distance dimension (e.g., the difference in uncertainty avoidance or difference in corruption levels), we suggest following Shenkar (2001), in that researchers should take all readily available information into account. For cultural distance and the Hofstede framework, this would imply that the six-dimensional framework is preferred as compared to the four-dimensional one, with the important caveat that users of the six-dimensional framework need to be aware of its theoretical and methodological characteristics, such as its sampling procedures, theoretical grounding, factor analytic structure, and the relationship between questionnaire items and the labelling of the dimensions.

Alternative Frameworks

As noted above, scholars have a choice between Hofstede, Schwartz and Globe for cultural distance. In addition, Berry et al. (2010) and Beugelsdijk et al. (2015) have used WVS-EVS data to develop Hofstede-

inspired cultural dimensions. To measure institutional distance, scholars have mostly used the World Bank QoG data, the Heritage Foundation EFI or the ICRG scores from the PRS group. In many studies, the preference for one of these frameworks has not been properly explained. Frequently, reference is just made to prior studies using a particular framework. The risk of not specifying explicitly why a particular framework and associated data are most suitable for a specific distance study is that it allows for *p*-hacking (Meyer et al. 2017). Furthermore, the choice of framework matters dramatically. In Table 26.4, we compare Mahalanobis distance indices using alternative data sources.

In the case of cultural distance, the various frameworks generate very different indices. The index based on all six Hofstede dimensions correlates at 0.01 with the Schwartz-based index, and at 0.11 with the Globe-based index. Schwartz and Globe only correlate at 0.18. The (very) low correlations between these three well-known cultural frameworks extend to the WVS-EVS based cultural distance indices as developed by Berry et al. (2010) and Beugelsdijk et al. (2015).[15]

Figure 26.1 visualizes the relationship between a Hofstede-based and a Schwartz-based cultural distance index using Mahalanobis' technique, with the United States as the home country. If both distance indices resulted in the same scores, the correlation would be 1 and all observations in Fig. 26.1 would lie on the 45° diagonal. This is clearly not the case and the shared variance is close to zero. Comparisons between Hofstede and Globe, and Schwartz and Globe yield similar pictures. Thus, cultural distance scores depend to a large extent on the culture framework used.

Table 26.4 Pairwise correlations between different cultural distance indices (Mahalanobis corrected)

		1	2	3	4
1	Cultural distance Hofstede 6	1			
2	Cultural distance Schwartz	0.01	1		
3	Cultural distance Globe	0.11	0.18	1	
4	Cultural distance WVS-EVS Berry et al.	0.16	0.11	0.13	1
5	Cultural distance WVS-EVS Beugelsdijk et al.	0.27	0.24	0.32	0.25

Correlations are based on unique country pairs (home–home combinations are excluded). The WVS-EVS used in Berry et al. (2010) are available for 96 countries. The WVS-EVS data used in Beugelsdijk et al. (2015) are available for 86 countries

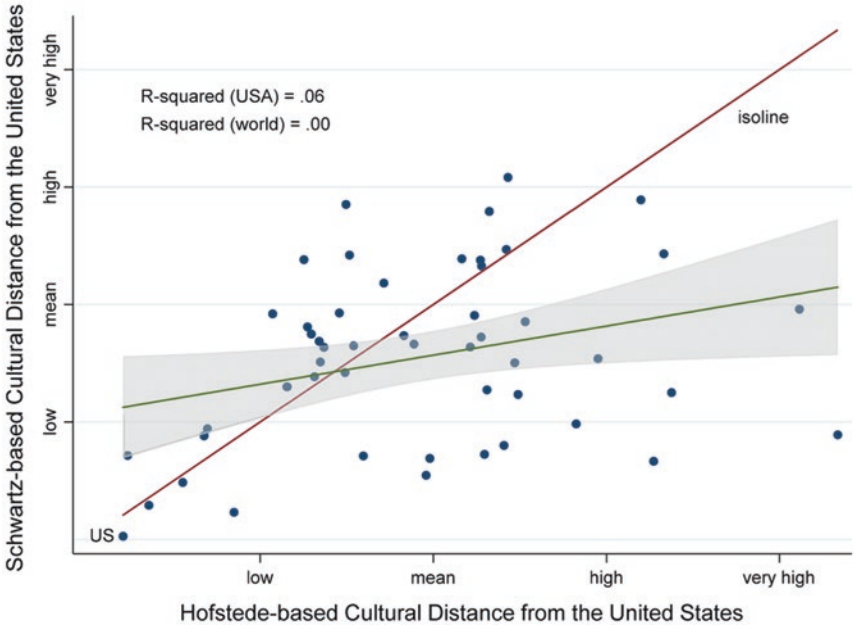

Fig. 26.1 Comparing Hofstede- and Schwartz-based cultural distance scores; cultural distance is computed using Mahalanobis' technique and scores are standardized to facilitate comparison. The *gray area* is the 95% confidence interval around the *regression line*. The United States is the home country

Figure 26.2 visualizes the relationship between an EFI-based and ICRG-based institutional distance index (using Mahalanobis' technique and again with the United States as the home country). The correlation across all country pairs in the world is 0.27 (0.34 for the United States as the home country). Although these correlations are higher than in the case of alternative cultural distance indices, they can still be considered low. Thus, institutional distance scores also depend on the data used.

These low correlations between the Hofstede, Schwartz, and Globe-based cultural distance indices (as well as the WVS-EVS-based extensions) and the EFI- and ICRG-based institutional distance indices essentially mean that these indices capture different facets of culture and institutions.[16] This raises the questions of whether one index is preferable to the other, and why. It also has implications for how to interpret and

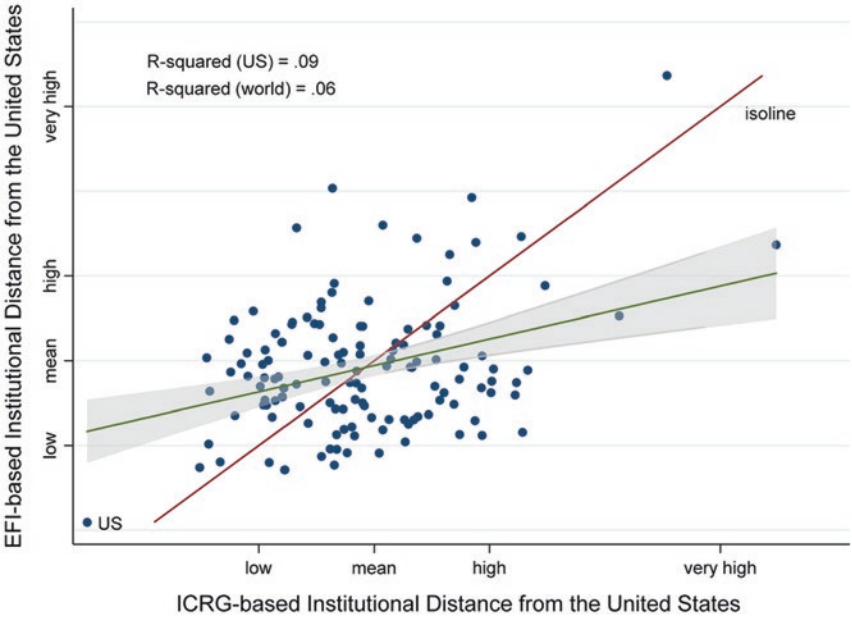

Fig. 26.2 Comparing ICRG- and EFI-based institutional distance scores; institutional distance is computed using Mahalanobis' technique and scores are standardized to facilitate comparison. The *gray area* is the 95% confidence interval around the *regression line*. The United States is the home country

compare studies which use alternative indices. If there are major differences between two indices supposedly measuring the same construct, results from different operationalizations cannot be compared in a simple fashion, thereby calling for more thoughtful reflection.

Recommendation

There are three options to address the choice among alternative frameworks. We illustrate these options in the context of cultural distance, but our reasoning can also be applied to institutional distance.

The first option is to select one of the cultural frameworks and to provide a theoretical and/or methodological justification for its use. It goes beyond the scope of this editorial to discuss all the theoretical and

methodological pros and cons of the three frameworks. As Schwartz notes when comparing his framework with Hofstede's, his "dimensions are based on different theoretical reasoning, different methods, a different set of nations, different type of respondents, data from a later historical period, a more comprehensive set of values, and value items screened to be reasonably equivalent in meaning across cultures" (Schwartz 1994: 116–117).[17] We refer to the original studies as well as discussions in cross-cultural psychology comparing these frameworks (Bond 2002; Hofstede 2010; Oyserman et al. 2002; Matsumoto and Yoo 2006; Peterson 2003, 2004; Peterson and Castro 2006; Peterson and Søndergaard 2011; Ralston et al. 2011, 2014; Schwartz 2014; Smith 2006; Smith et al. 1996).

However, as we argued before, not all of the criticism raised against these frameworks is relevant when cultural *distance* is concerned. For example, if Inglehart's thesis is correct, and cultures change but countries continue to move in the same direction, then it does not make sense, for example, to choose Globe over Hofstede because Globe data are more recent. Explicitly specifying why a particular framework is used is important, because the choice of framework is likely to affect empirical results (Beugelsdijk et al. 2018). Such explanations and justifications will improve the quality of the debate between authors and reviewers.

The second option is to argue, on theoretical and methodological grounds, that none of the differences between the three frameworks allow for a clear reason why one should be preferred over another. In this case, one might be indifferent regarding which framework to use. In addition, one could argue that all three frameworks capture part of the overall variation in cultural values and all three do so in an imperfect way. Hofstede's data give information on cultural diversity in a matched sample of IBM employees (complemented by two dimensions based on stratified representative samples). Schwartz provides similar information coming from students and teachers, while Globe does so coming from middle managers. While the frameworks partly overlap, as evidenced by the correlations between the dimensions, combined, the three sets of data arguably pick up more variation in cross-country cultural differences than when used in isolation (Steenkamp 2001). Therefore, one could argue that integrating Hofstede, Schwartz, and Globe in one overall distance index may provide a more complete picture of the overall variation in cultural values

(Beugelsdijk, Kostova, and Roth 2017). From such an "agnostic" perspective, all 22 indicators for cultural differences (six dimensions from Hofstede, seven from Schwartz, and nine from Globe) are indicative of cultural values, and the Mahalanobis technique can be used to integrate them in one overall index. The resulting "grand" cultural distance index correlates 0.25 with Hofstede's, 0.36 with Schwartz', and 0.51 with Globe's cultural distance indices. One practical disadvantage is that data for a combined "grand" index are only available for 40 countries. More work needs to be carried out to explore the usefulness, as well as the conceptual and methodological soundness, of such a "grand" index approach.

The third option is simply not to use a cultural distance index. All three frameworks have serious theoretical and methodological drawbacks, which have led some scholars to recommend that they should be avoided (McSweeney 2002; Kirkman et al. 2006). The low correlations between the cultural distance indices built upon Hofstede, Schwartz and Globe suggest that the three indices capture very different aspects of the overall variation in cultural values, but this result can also be interpreted as evidence of their lack of conceptual and methodological soundness.

This third option still allows controlling for cultural differences in empirical studies. If the variation in cultural distance is related to the variation in other – less disputed – distance dimensions, we could perhaps indirectly control for cultural distance by including those alternative distance dimensions. For example, we know that economic development affects cultural values (Beugelsdijk and Welzel 2018; Inglehart and Baker 2000), meaning that the inclusion of economic distance as a control variable in empirical studies is likely to capture a substantial part of the impact of cultural distance.[18] Empirical research leveraging index construction methods referred to earlier is required to further unpack the empirical relationship between cultural distance and other distance dimensions.

Conclusion

In this editorial, we have discussed a set of recurring questions on the theory, methods and data utilized in cross-country distance research. We have pointed to commonalities in distance indicators and have formulated

best practice guidelines on: (1) how to theorize on distance, and (2) what method and (3) what data to use when constructing a distance index. We have illustrated each of these topics with data on as many country pairs as possible. Our goal is to help authors, reviewers, and editors focus on what matters most in terms of theory, method, and data by clarifying the issues most critical to improving distance research. At the same time, we acknowledge that there are some considerations that simply require more transparency so that the nature of the models and relationships tested become clearer. Above all, we call for a more disciplined approach in distance research that is based on a better understanding of what has already been achieved in the field and what still remains to be done.

We reach three main conclusions, which collectively highlight the need to match data and method with the nature of an explicitly and carefully crafted theoretical argument. First, there is no distance theory as such. Distance is given meaning within the context of specific theoretical frameworks. Hence, it is critical that authors specify clearly the theoretical context of their arguments on distance. It is important to be explicit and precise about the exact mechanisms by which distance affects a particular outcome – especially because many decisions in international business are multi-staged. Credibly articulating assumptions and mechanisms should clarify the nature of the relationship between distance and, for example, location choice, entry mode decisions, and performance. Theoretical clarity is also required when distance is conceptualized as an aggregate construct, which requires a composite index.

Second, using three different cultural frameworks (Hofstede, Schwartz and Globe) and three different sets of indicators of institutional quality (Quality of Governance, Economic Freedom Index, and International Country Risk Guide), we have demonstrated that the choice of method can, but need not, have a major impact on the results. Scholars need to be transparent about the method employed, and ideally use the (co-) variance-corrected distance. Although the Mahalanobis technique to control for (co-)variance is very powerful, the exact way to control for co-variance is related to the theoretical nature of the construct (formative–reflective), an aspect of distance research that has so far not been sufficiently addressed in the distance literature.

Third, in contrast to the relatively minor implications of correcting for (co-)variance, the selection of the framework and the data to measure cultural or institutional distance has a major impact. The Hofstede-based cultural distance framework measures different aspects of cross-cultural distance compared with the frameworks of Schwartz or Globe. Authors should properly justify their choice of framework. It is important to note, however, that some of the criticism directed towards cultural frameworks does not apply to composite distance constructs. For institutional distance, the choice of data also matters, but the difference between what QoG and EFI measure is smaller than in the case of cultural distance.

Our discussion of distance is not without limitations. First, we have not addressed the stability of the effect of distance on a particular outcome variable. If firms learn how to deal with contextual differences, their effect should become smaller over time, even though the distance itself stays the same. Yet the seven meta-analyses on the impact of cultural distance effects show no consistent evidence of its reduced impact over time (Beugelsdijk et al. 2018; Magnusson et al. 2008; Morschett et al. 2010; Stahl and Voigt 2008; Tihanyi et al. 2005; Zhao et al. 2004). The lack of evidence of a temporal effect of cultural distance contrasts with firm-level studies showing that firms can learn to deal with cross-country differences (Johanson and Vahlne 2009). Addressing the temporal stability of the *effect* of distance on international business outcomes would be a fruitful avenue for further research (Friedman 2005; Ghemawat 2017).

Second, in samples of only one home or one host country, distance effects are conflated with level effects (Brouthers et al. 2016; Harzing and Pudelko 2016; Van Hoorn and Maseland 2016; Sivakumar and Nakata 2001). In such a sample structure, there can be a high negative or positive correlation between the distance from a home country to other countries and the level score of the construct on which the distance is calculated. For example, the QoG-based institutional distance between the United States and all 196 host countries for which we have data correlates – 0.97 with the institutional quality in these host countries. The reverse obtains for poor home countries, with the correlation being + 0.98 for Zimbabwe. These high correlations for extreme countries (United States on the one hand and Zimbabwe on the other) illustrate that studies with single home or host countries cannot disentangle distance from country-level effects.

Clearly, distance effects can be asymmetric. All the correlations we have reported are based on all home and host countries for which data are available. We do so to make sure that our conclusions are not affected by this conflation of distance and level effects. In addition to using multiple home and host countries, as recommended by Brouthers et al. (2016), we suggest that scholars report the correlation between the distance variable and the host country-level score of the variable for which distance is calculated. A high correlation is reason for concern, as it affects the interpretation of the distance argument tested.

Lastly, we have argued that there is no grand theory of distance, and that distance only has meaning within the context of a specific theory. Here, researchers should carefully reflect about the spatial mechanisms relevant to the research question they try to address. From a conceptual perspective, continuous distance effects can be found at all spatial levels, both within and between countries (Dheer et al. 2015; Lenartowicz and Roth 2001), while national border effects only occur between countries. Unlike distance effects, border effects are discrete because borders often serve as a qualitative disjuncture in space (Anderson 1991; Beugelsdijk and Mudambi 2013), and because many contextual characteristics are nation-specific (this holds especially for formal institutions) (Peterson et al. 2018). Here, we see an exciting research agenda unfolding on cross-country distance and on the rising meaning of national borders in the face of anti-globalization movements.

Acknowledgements We thank Heather Berry, Johannes Kleinhempel, Thomas Lindner, Laura Spierdijk, Alain Verbeke, Tom Wansbeek, Juliette de Wit, participants of the AIB 2018 professional development workshop, and our anonymous reviewers, for comments and suggestions.

Appendix A

Table 26.5 Overview of cross-cultural frameworks

	Hofstede framework	Schwartz value survey	Globe	WVS-EVS
Key references	Hofstede (1980, 2001), Hofstede et al. (2010)	Schwartz (1994, 1999, 2006)	House et al. (2004)	Inglehart (1990, 1997), Inglehart and Baker (2000)
Mostly used in	Cross-cultural psychology, management	Cross cultural psychology, management	Management	Sociology, political science, economics
Respondents	117,000	75,148	7794[a]	495,000
Sample	IBM employees (4 dimensions) and stratified samples of adults (2 dimensions)	Teachers and students	Middle managers	Stratified samples of adults
Country coverage	70 for first 4 dimensions 62 for all 6 dimensions	71	58	110
Year coverage	1968–1972 for IBM data 2000s for 5th and 6th dimensions	1988–2005	1995–1997	1981–2014 with irregular intervals
Availability of individual "raw" data	No	Yes, publicly available from Israeli science foundation	No, not made available	Yes, publicly available from the WVS and EVS website
Number of dimensions	4 + 2 Two dimensions were added 40 years later	7[b]	9(×2) Globe distinguishes between values and practices	Not predefined; Inglehart (1990) defined two

(*continued*)

Table 26.5 (continued)

	Hofstede framework	Schwartz value survey	Globe	WVS-EVS
Nature of questions	1–5 scale	1–7 scale	1–7 scale	A mix of 1–10; 0–1; 1–4
Cultural dimensions	1. Collectivism-individualism 2. Power distance 3. Masculinity 4. Uncertainty avoidance 5. Long-term orientation 6. Indulgence versus restraint Dimensions 5 and 6 were added later.	1. Harmony 2. Embeddedness 3. Hierarchy 4. Mastery 5. Affective autonomy 6. Intellectual autonomy 7. Egalitarianism	1. Uncertainty avoidance 2. Future orientation 3. Power distance 4. Institutional collectivism 5. Humane orientation 6. Performance orientation 7. In-group collectivism 8. Gender egalitarianism 9. Assertiveness	Inglehart defined two: 1. Traditional-secular/rational 2. Survival-selfexpression

[a]House et al. (2004) mention a total of 15,000, but it should be noted that approximately half of this sample has been used to collect data on leadership, and 7794 respondents for the survey on national cultural values. Of these 7794 respondents, Sweden stands out with a total of 895 respondents (based on personal communication from Paul Hanges)

[b]Note that Schwartz (1994, 1999, 2006) distinguishes between personal values and national cultural (societal) orientations. The number of dimensions and the meaning of those dimensions are not the same at the individual and societal level. At the societal level, Schwartz conceptualizes cultural values as "the normative value emphases that underlie societal functioning" (Schwartz 2011: 314). Here, we refer to the seven national cultural orientations (Schwartz 2006). For a recent analysis discussing Schwartz' two frameworks in the context of internationalization strategy, see Verbeke et al. (in press). Hofstede has stated explicitly that his framework is a national cultural framework and thus cannot be used at the individual level (Hofstede 2001)

Appendix B

Table 26.6 Correlation table of Hofstede's dimensions, Schwartz's dimensions and Globe's value dimensions

	Cultural dimensions		1	2	3	4	5	6	7	8	9	10
1	Collectivism–individualism	H	1									
2	Power distance	H	– 0.62	1								
3	Masculinity	H	0.07	0.13	1							
4	Uncertainty avoidance	H	– 0.22	0.23	– 0.07	1						
5	Long-term orientation	H	0.07	0.05	0.01	– 0.02	1					
6	Indulgence versus restraint	H	0.16	– 0.31	0.09	– 0.07	– 0.43	1				
7	Harmony	S	0.21	– 0.12	– 0.06	0.37	0.23	0.04	1			
8	Embeddedness	S	– 0.55	0.62	– 0.06	– 0.03	– 0.34	– 0.21	– 0.43	1		
9	Hierarchy	S	– 0.48	0.43	0.16	– 0.23	– 0.01	– 0.23	– 0.62	0.50	1	
10	Mastery	S	– 0.22	0.11	0.14	– 0.06	0.00	– 0.12	– 0.48	– 0.13	0.46	1
11	Affective autonomy	S	0.67	– 0.61	0.00	– 0.13	0.35	0.13	0.22	– 0.87	– 0.40	0.21
12	Intellectual autonomy	S	0.43	– 0.42	0.07	0.14	0.30	0.13	0.57	– 0.85	– 0.57	– 0.11
13	Egalitarianism	S	0.43	– 0.47	– 0.10	0.05	– 0.25	0.49	0.43	– 0.43	– 0.63	– 0.45
14	Uncertainty avoidance	G	– 0.72	0.72	0.07	0.32	– 0.27	– 0.25	– 0.24	0.76	0.47	0.13
15	Future orientation	G	– 0.49	0.58	0.13	0.29	– 0.42	– 0.06	– 0.32	0.61	0.32	0.13
16	Power distance	G	0.15	– 0.01	0.15	– 0.46	0.05	– 0.29	– 0.31	0.38	0.42	0.08
17	Institutional collectivism	G	– 0.52	0.36	0.01	0.37	– 0.34	0.16	0.14	0.16	0.02	0.08
18	Humane orientation	G	0.20	– 0.10	0.01	– 0.06	0.01	0.08	– 0.09	– 0.02	– 0.01	– 0.18
19	Performance orientation	G	– 0.15	0.15	0.04	0.16	– 0.59	0.38	0.06	0.12	– 0.24	– 0.09
20	In–group collectivism	G	– 0.26	0.20	– 0.16	0.24	– 0.61	0.41	– 0.06	0.23	– 0.04	– 0.02
21	Gender egalitarianism	G	0.42	– 0.46	– 0.02	0.06	– 0.05	0.55	0.30	– 0.69	– 0.52	0.01
22	Assertiveness	G	– 0.17	0.29	0.15	– 0.38	0.09	– 0.29	– 0.39	0.27	0.51	0.23

(*continued*)

Table 26.6 (continued)

	Cultural dimensions		11	12	13	14	15	16	17	18	19	20	21
1	Collectivism–individualism	H											
2	Power distance	H											
3	Masculinity	H											
4	Uncertainty avoidance	H											
5	Long-term orientation	H											
6	Indulgence versus restraint	H											
7	Harmony	S											
8	Embeddedness	S											
9	Hierarchy	S											
10	Mastery	S											
11	Affective autonomy	S	1										
12	Intellectual autonomy	S	0.68	1									
13	Egalitarianism	S	0.19	0.46	1								
14	Uncertainty avoidance	G	– 0.76	– 0.65	– 0.58	1							
15	Future orientation	G	– 0.59	– 0.56	– 0.36	0.66	1						
16	Power distance	G	– 0.34	– 0.28	– 0.38	0.16	– 0.07	1					
17	Institutional collectivism	G	– 0.36	– 0.11	0.15	0.43	0.47	– 0.33	1				
18	Humane orientation	G	0.09	– 0.02	0.14	– 0.17	– 0.12	– 0.38	– 0.13	1			
19	Performance orientation	G	– 0.14	– 0.11	0.14	0.18	0.42	– 0.33	0.46	– 0.02	1		
20	In-group collectivism	G	– 0.17	– 0.23	0.02	0.27	0.49	– 0.18	0.30	– 0.21	0.56	1	
21	Gender egalitarianism	G	0.59	0.52	0.59	– 0.56	– 0.34	– 0.47	– 0.01	0.19	0.22	0.16	1
22	Assertiveness	G	– 0.16	– 0.20	– 0.54	0.18	0.08	0.31	– 0.22	– 0.11	– 0.01	– 0.02	– 0.28

All correlations larger than |0.02| are significant at 0.05
H Hofstede; *S* Schwartz; *G* Globe

Appendix C

Table 26.7 Correlation table of QoG dimensions, EFI dimensions and ICRG dimensions

	Institutional dimensions		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Control of corruption	QoG	1													
2	Government effectiveness	QoG	0.93	1												
3	Political stability	QoG	0.75	0.69	1											
4	Regulatory quality	QoG	0.86	0.94	0.61	1										
5	Rule of law	QoG	0.94	0.94	0.78	0.88	1									
6	Voice and accountability	QoG	0.77	0.77	0.71	0.76	0.84	1								
7	Property rights	EFI	0.92	0.91	0.64	0.88	0.94	0.78	1							
8	Freedom from corruption	EFI	0.97	0.93	0.71	0.86	0.94	0.75	0.92	1						
9	Fiscal freedom	EFI	− 0.35	− 0.30	− 0.21	− 0.25	− 0.32	− 0.41	− 0.31	− 0.35	1					
10	Government spending	EFI	− 0.41	− 0.37	− 0.41	− 0.26	− 0.43	− 0.44	− 0.33	− 0.39	0.45	1				
11	Business freedom	EFI	0.70	0.77	0.48	0.76	0.73	0.61	0.71	0.69	− 0.08	− 0.27	1			
12	Labor freedom	EFI	0.35	0.36	0.27	0.35	0.38	0.21	0.33	0.31	0.13	− 0.07	0.44	1		
13	Monetary freedom	EFI	0.51	0.52	0.44	0.61	0.55	0.54	0.54	0.49	− 0.18	− 0.16	0.42	0.22	1	
14	Trade freedom	EFI	0.51	0.61	0.40	0.71	0.58	0.53	0.51	0.51	− 0.10	− 0.25	0.54	0.21	0.45	1
15	Investment freedom	EFI	0.63	0.69	0.46	0.85	0.70	0.68	0.71	0.64	− 0.22	− 0.15	0.56	0.26	0.64	0.63
16	Financial freedom	EFI	0.65	0.74	0.43	0.86	0.72	0.64	0.75	0.66	− 0.14	− 0.17	− 0.60	0.26	0.56	0.62
17	Government stability	ICRG	0.02	0.00	0.16	0.01	0.00	− 0.19	− 0.02	0.07	0.15	0.16	− 0.06	0.18	0.15	− 0.02
18	Socio-economic conditions	ICRG	0.81	0.86	0.65	0.81	0.82	0.60	0.76	0.81	− 0.21	− 0.35	0.72	0.37	0.44	0.59
19	Investment profile	ICRG	0.67	0.72	0.57	0.76	0.71	0.54	0.69	0.69	0.02	− 0.05	0.61	0.44	0.49	0.52
20	Internal conflict	ICRG	0.61	0.59	0.89	0.58	0.61	0.59	0.47	0.61	− 0.19	− 0.31	0.44	0.17	0.33	0.47
21	External conflict	ICRG	0.40	0.42	0.58	0.47	0.42	0.52	0.37	0.42	− 0.16	− 0.26	0.25	0.11	0.36	0.31
22	Corruption	ICRG	0.95	0.89	0.72	0.83	0.91	0.78	0.90	0.95	− 0.47	− 0.47	0.71	0.30	0.51	0.54
23	Military in politics	ICRG	0.72	0.77	0.73	0.76	0.77	0.73	0.68	0.71	− 0.24	− 0.45	0.65	0.34	0.43	0.60
24	Religious tensions	ICRG	0.38	0.36	0.60	0.38	0.36	0.46	0.30	0.37	− 0.21	− 0.25	0.34	0.09	0.37	0.43
25	Law and order	ICRG	0.76	0.73	0.61	0.65	0.80	0.51	0.69	0.75	− 0.31	− 0.48	0.63	0.33	0.43	0.49
26	Ethnic tension	ICRG	0.31	0.29	0.49	0.27	0.30	0.19	0.15	0.30	0.06	− 0.18	0.24	0.11	0.13	0.28
27	Democratic accountability	ICRG	0.55	0.63	0.44	0.65	0.61	0.85	0.62	0.55	− 0.44	− 0.38	0.43	0.02	0.47	0.48
28	Bureaucratic accountability	ICRG	0.83	0.89	0.62	0.81	0.84	0.74	0.78	0.82	− 0.41	− 0.48	0.65	0.37	0.42	0.52

(*continued*)

Table 26.7 (continued)

	Institutional dimensions		15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	Control of corruption	QoG														
2	Government effectiveness	QoG														
3	Political stability	QoG														
4	Regulatory quality	QoG														
5	Rule of law	QoG														
6	Voice and accountability	QoG														
7	Property rights	EFI														
8	Freedom from corruption	EFI														
9	Fiscal freedom	EFI														
10	Government spending	EFI														
11	Business freedom	EFI														
12	Labor freedom	EFI														
13	Monetary freedom	EFI														
14	Trade freedom	EFI														
15	Investment freedom	EFI	1													
16	Financial freedom	EFI	0.83	1												
17	Government stability	ICRG	– 0.01	0.02	1											
18	Socio-economic conditions	ICRG	0.53	0.62	0.13	1										
19	Investment profile	ICRG	0.62	0.68	0.27	0.72	1									
20	Internal conflict	ICRG	0.42	0.42	0.20	0.54	0.47	1								
21	External conflict	ICRG	0.37	0.37	0.13	0.31	0.40	0.56	1							
22	Corruption	ICRG	0.61	0.64	0.02	0.76	0.62	0.55	0.42							
23	Military in politics	ICRG	0.56	0.58	0.00	0.66	0.53	0.65	0.55	0.67	1					
24	Religious tensions	ICRG	0.35	0.29	0.07	0.30	0.26	0.61	0.33	0.32	0.47	1				
25	Law and order	ICRG	0.46	0.49	0.05	0.69	0.48	0.47	0.22	0.72	0.65	0.25	1			
26	Ethnic tension	ICRG	0.23	0.20	0.07	0.34	0.26	0.43	0.21	0.23	0.36	0.44	0.33	1		
27	Democratic accountability	ICRG	0.60	0.58	– 0.26	0.39	0.37	0.38	0.41	0.56	0.58	0.31	0.33	0.02	1	
28	Bureaucratic accountability	ICRG	0.54	0.63	– 0.06	0.76	0.58	0.49	0.40	0.81	0.69	0.29	0.64	0.22	0.59	1

All correlations larger than |0.17| are significant at 0.05

QoG Quality of Governance; *ER* Economic Freedom Index; *ICRG* International Country Risk Guide

Appendix D

Table 26.8 Factor analysis of Hofstede's six dimensional model

	(Rotated) Factor loadings Three-factor solution		
Hofstede dimensions	Factor 1 (explains 30%)	Factor 2 (explains 25%)	Factor 3 (explains 17%)
Power distance	0.85	0.15	0.20
Individualism	– 0.87	0.03	0.08
Masculinity	0.04	– 0.02	0.98
Uncertainty avoidance	0.47	– 0.07	– 0.07
Long-term orientation	– 0.12	0.88	0.07
Indulgence versus restraint	– 0.26	– 0.84	0.11

n =62 countries. The analysis results in three factors with eigenvalues larger than 1, explaining 72% of the variation across all six dimensions

Notes

1. In one of the earlier explanations of this notion, Goodnow and Hansz (1972) state that, as firms enter markets further away from the United States, "government becomes less stable, the markets become poorer, the economy becomes less stable, cultural homogeneity declines, legal and geographical barriers go up and cultures become different" (Goodnow and Hansz 1972: 37). Johanson and Vahlne (1977) generalize this argument by stating that the further from the home country, the more uncertainty and the higher the costs of acquiring information.

2. It is interesting to note that trade, the dependent variable in gravity models, is one of the explanatory variables in the Uppsala model.
3. There is no grand theory in the sense of a single framework with a set of assumptions, mechanisms, and boundary conditions. That does not mean that there is no theory of national cultural differences (Adler 1983).
4. Theoretically, such discrete changes can also be found at other levels. For example, Ronen and Shenkar (2013) have shown that countries can be grouped into a limited number of culturally homogeneous supranational zones.
5. In light of the popularity of their distance measure, it is interesting to point out that Kogut and Singh's (1988) findings regarding the effect of uncertainty avoidance on entry mode were stronger than the cultural distance effect (see Kogut and Singh 1988: 424).
6. Occasionally, this is referred to as Cartesian distance (e.g., Buchner et al. 2017).
7. In addition to the Kogut and Singh index and the Euclidean index, some researchers have used a mix of the two. For example, Barkema and Vermeulen (1997) use a Euclidean distance index but correct for differences in the variances of each of Hofstede's cultural dimension by multiplying the individual dimension distances by $1/v_i$ where v_i is the variance of each cultural dimension. The Kogut and Singh index adapts the scales in a similar way (see Eq. 26.1), but does not take the square root. In this case, the difference between the Kogut and Singh index and this third approach is the aggregation procedure (square root vs. dividing by the number of dimensions). Although this is not the same type of transformation, they are effectively highly correlated.
8. In the case of South Africa, the WVS-EVS data provide a score for the fifth and sixth dimensions of Hofstede's framework. The four original Hofstede dimensions refer to Caucasian South Africans only. We have decided not to mix both samples, and hence exclude South Africa from the analysis.
9. Inglehart (1997) used the WVS-EVS data to develop two dimensions of differences in national cultural values related to degree of materialism. However, his framework is rarely used in management (it is mostly used in sociology and political science). One reason for the limited use of the WVS-EVS data in management is the fact that, despite the use of stratified nationally representative samples and the richness of the WVS-EVS

data (they contain more than 200 value-related questions), the user-friendliness of the database is limited. Note also that in its fifth and sixth rounds, the WVS has included a condensed ten-item version of the Schwartz values.

10. These correlations do not depend on the choice of home country. Drogendijk & Slangen (2006, 372) report a correlation of.97 between the Euclidean and Kogut and Singh versions of the four-dimensional Hofstede index when using the Netherlands as a home country. For the most used home country in cultural distance research, the United States, this correlation is also. 97.
11. We use data for 2013, but as the QoG index is highly correlated over time (0.99 between t and $t + 1$), the selected year does not affect the outcome.
12. The Mahalanobis institutional distance based on six dimensions correlates 0.96 with the five dimensional Mahalanobis institutional distance index (using QoG data).
13. As the European manager of personnel research at IBM, Geert Hofstede had privileged access to the confidential IBM employee data used to develop his cross-cultural framework. The original data stayed at IBM after Geert Hofstede left IBM in 1973 (based on personal communication from Geert Hofstede).
14. See Ralston et al. (1997) and Ralston (2008) for an alternative view.
15. Although both Berry and Beugelsdijk use WVS-EVS to develop Hofstede-inspired cultural dimensions, there is only limited overlap in the WVS-EVS questions used by Berry et al. (2010) and Beugelsdijk et al. (2015). Both studies use the question on trust (WVS code a165). The correlations between the original Hofstede dimensions, and the ones developed by Berry et al. (2010) and Beugelsdijk et al. (2015) differ. Berry et al.'s (2010) individualism correlates 0.24 with Hofstede's original individualism, Berry et al's power distance correlates 0.25 with Hofstede's, uncertainty avoidance correlates 0.52 with Hofstede's and Berry et al's masculinity correlates 0.16 with Hofstede's. For Beugelsdijk et al. (2015), these correlations with the original Hofstede dimensions are: 0.77 for individualism, 0.74 for uncertainty avoidance, 0.72 for power distance, and 0.92 for indulgence versus restraint.
16. Note that the correlations shown in Table 26.3 are country-specific, because each cultural distance is calculated relative to a different home country. Hence, the correlation between two distance indices can vary

depending on home country. A similar country specificity holds for the correlation between the ICRG- and EFI-based institutional distance indices. We have calculated all the cultural distance correlations shown in Table 26.3 as well as the correlation between the ICRG- and EFI-based institutional distance indices for each home country. Calculating the country-specific correlations does not change our overall conclusion.

17. Schwartz has described the difference between his approach and those of others (i.e., Hofstede) in the following way: "(a) It [the approach] derived the cultural orientations from a priori theorizing rather than post hoc examination of data. (b) It designated a priori the value items that serve as markers for each orientation. (c) It used as measures only items tested for cross-cultural equivalence of meaning. (d) It included a set of items demonstrated to cover the range of values recognized cross-culturally, a step toward ensuring relative comprehensiveness of cultural value dimensions. (e) It specified how the cultural orientations are organized into a coherent system of related dimensions and verified this organization, rather than assuming that orthogonal dimensions best capture cultural reality. (f) It brought empirical evidence that the order of national cultures on each of the orientations is robust across different types of samples from each of a large number of nations around the world [and using different instruments]. These distinctive features increase the promise of this approach for future research." (Schwartz 2004: 73, and reprinted in Schwartz 2006: 179).
18. A similar relationship exists between institutional distance and economic distance, because high-quality institutions are generally associated with high levels of economic development (Acemoglu et al. 2005).

References

Abdi, M., and P.S. Aulakh. 2012. Do country level institutional frameworks and interfirm governance arrangements substitute or complement in international business relations. *Journal of International Business Studies* 43 (5): 477–497.

Acemoglu, D., S. Johnson, and J.A. Robinson. 2005. Chapter 6: Institutions as a fundamental cause of long run growth. In *Handbook of economic growth*, ed. P. Aghion and S.N. Durlauf, 387–472. Amsterdam: Elsevier Holland.

Adler, N.J. 1983. Cross-cultural management research—The ostrich and the trend. *Academy of Management Review* 8 (2): 226–232.

Ailon, G. 2008. Mirror, Mirror on the wall: Culture's consequences in a value test of its own design. *Academy of Management Review* 33 (4): 885–904.

Ambos, B., and L. Håkanson. 2014. The concept of distance in international management research. *Journal of International Management* 20 (1): 1–7.

Anderson, B. 1991. *Imagined communities*. London: Verso.

Anderson, J., and E. van Wincoop. 2003. Gravity with gravitas: A solution to the border puzzle. *American Economic Review* 93: 17–192.

Ang, S.H., M.H. Benischke, and J.P. Doh. 2015. The interactions of institutions on foreign market entry mode. *Strategic Management Journal* 36 (10): 1536–1553.

Barkema, H.G., and F. Vermeulen. 1997. What differences in cultural backgrounds of partners are detrimental for international joint ventures? *Journal of International Business Studies* 28 (4): 854–864.

Baskerville, R.F. 2003. Hofstede never studied culture. *Accounting, Organizations and Society* 28: 1–14.

Beckerman, W. 1956. Distance and the pattern of intra-European trade. *The Review of Economics and Statistics* 38 (1): 31–40.

Bergstrand, J.H. 1985. The gravity equation in international trade: Some micro economic foundations and empirical evidence. *The Review of Economics and Statistics* 67 (3): 474–481.

Berry, H., M. Guillen, and N. Zhou. 2010. An institutional approach to cross-national distance. *Journal of International Business Studies* 41 (9): 1460–1480.

Beugelsdijk, S., and R. Mudambi. 2013. MNEs as border-crossing multi-location enterprises: The role of discontinuities in geographic space. *Journal of International Business Studies* 44 (5): 413–426.

Beugelsdijk, S., and C. Welzel. 2018. Dimensions and dynamics of national culture: Synthesizing Hofstede with Inglehart. *Journal of Cross-Cultural Psychology*. https://doi.org/10.1177/0022022118798505.

Beugelsdijk, S., R. Maseland, and A. van Hoorn. 2015. Are scores on Hofstede's dimension of national culture stable over time? A cohort analysis. *Global Strategy Journal* 5 (3): 223–240.

Beugelsdijk, S., T. Kostova, and K. Roth. 2017. An overview of Hofstede-inspired country level culture research in international business since 2006. *Journal of International Business Studies* 48 (1): 30–47.

Beugelsdijk, S., P. Nell, and P. Ambos. 2017. When do distance effects become empirically observable? An investigation in the context of headquarters value creation for subsidiaries. *Journal of International Management* 23 (3): 255–267.

Beugelsdijk, S., T. Kostova, V. Kunst, E. Spadafora, and M. van Essen. 2018. Cultural distance and firm internationalization. *Journal of Management* 44 (1): 89–130.

Bollen, K.A., and A. Diamantopoulos. 2017. In defense of causal-formative indicators: A minority report. *Psychological Methods* 22 (3): 581–596.

Bond, M.H. 2002. Reclaiming the Individual from Hofstede's ecological analysis—A 20 year odyssey: Comment on Oyserman et al. (2002). *Psychological Bulletin* 128 (1): 73–77.

Brereton, R.G., and G.R. Lloyd. 2016. Re-evaluating the role of the Mahalanobis distance measure. *Journal of Chemometrics* 30: 134–143.

Brewer, P., and S. Venaik. 2011. Individualism-collectivism in Hofstede and GLOBE. *Journal of International Business Studies* 42 (3): 436–445.

Brouthers, K.D. 2002. Institutional, cultural and transaction costs influences on entry mode choice and performance. *Journal of International Business Studies* 33 (2): 203–221.

Brouthers, L.E., V.B. Marshall, and D.L. Keig. 2016. Solving the single-country sample problem in cultural distance studies. *Journal of International Business Studies* 47 (4): 471–479.

Buchner, A., S. Espenlaub, A. Khurhsed, and A. Mohammed. 2017. Cross-border venture capital investments: The impact of foreignness on returns. *Journal of International Business Studies* 49 (5): 575–604.

Campbell, J.T., L. Eden, and S. Miller. 2012. Multinationals and corporate social responsibility on host countries: Does distance matter? *Journal of International Business Studies* 43 (1): 84–106.

Coltman, T., T.M. Devinney, D.F. Midgley, and S. Venaik. 2008. Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research* 61: 1250–1262.

Cuypers, I.P.R., G. Ertug, P.M.A.R. Heugens, B. Kogut, and T. Zou. 2018. The making of a construct: Lessons from 30 years of the Kogut and Singh cultural distance index. *Journal of International Business Studies*. https://doi.org/10.1057/s41267-018-0181-5.

De Maesschalck, R., D. Jouan-Rimbaud, and D.L. Massart. 2000. The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems* 50 (1): 1–18.

Demirbag, M., M. Apaydin, and E. Tatoglu. 2011. Survival of Japanese subsidiaries in the Middle East and North Africa. *Journal of World Business* 46 (4): 411–425.

Dheer, R., T. Lenartowicz, and M.F. Peterson. 2015. Mapping Indiás regional subcultures: Implications for international management. *Journal of International Business Studies* 46 (4): 443–444.

Diamantopoulos, A., and H.M. Winklhofer. 2001. Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research* 38 (2): 269–277.

Diamantopoulos, A., P. Riefler, and K.P. Roth. 2008. Advancing formative measurement models. *Journal of Business Research* 61 (12): 1203–1218.

Dikova, D., P.R. Sahib, and A. van Witteloostuijn. 2010. Cross border acquisitions abandonment and completion: The effect of institutional differences and organizational learning in the international business service industry, 1981–2001. *Journal of International Business Studies* 41 (2): 223–245.

Dow, D., and A. Karunaratna. 2006. Developing a multidimensional instrument to measure psychic distance stimuli. *Journal of International Business Studies* 37 (5): 1–25.

Drogendijk, R., and A. Slangen. 2006. Hofstede, Schwartz, or managerial perceptions? The effects of different cultural distance measures on establishment mode choices by multinational enterprises. *International Business Review* 15 (4): 361–380.

Eden, L., & Miller, S. R. 2004. Distance matters: Liability of foreignness, institutional distance and ownership strategy. *Advances in International Management,* 16 *(Theories of the Multinational Enterprise: Diversity, Complexity and Relevance),* 187–221.

Feenstra, R., J. Markusen, and A.K. Rose. 2001. Using the gravity equation to differentiate among alternative theories of trade. *The Canadian Journal of Economics* 34 (2): 430–447.

Fernandez, D.R., D.S. Carlson, L.P. Stepina, and J.D. Nicholson. 1997. Hofstede's country classification 25 years later. *The Journal of Applied Psychology* 137 (1): 43–54.

Flury, B.K., and H. Riedwyl. 1986. Standard distance in univariate and multivariate analysis. *American Statistician* 40 (3): 249–251.

Friedman, T.L. 2005. *The world is flat: A brief history of the twenty-first century*. London: Macmillan.

Ghemawat, P. 2001. Distance still matters. The hard reality of global expansion. *Harvard Business Review* 79 (8): 137–147.

———. 2017. *The laws of globalization and business applications*. Cambridge: Cambridge University Press.

Goodnow, J.D., and J.E. Hansz. 1972. Environmental determinants of overseas market entry strategies. *Journal of International Business Studies* 3 (1): 33–50.

Gubbi, S.R., P.S. Aulakh, S. Ray, M.B. Sarkar, and R. Chittoor. 2010. Do international acquisitions by emerging-economy firms create shareholder value? The case of Indian firms. *Journal of International Business Studies* 41 (3): 397–418.

Guiso, L., P. Sapienza, and L. Zingales. 2009. Cultural biases in economic exchange. *Quarterly Journal of Economics* 124 (3): 1095–1131.

Håkanson, L. 2014. The role of psychic distance in international trade: A longitudinal analysis. *International Marketing Review* 31 (3): 210–236.

Håkanson, L., and B. Ambos. 2010. The antecedents of psychic distance. *Journal of International Management* 16 (3): 195–210.

Håkanson, L., and P. Kappen. 2017. The 'casino model' of internationalization: An alternative Uppsala paradigm. *Journal of International Business Studies* 48 (9): 1103–1113.

Håkanson, L., B. Ambos, A. Schuster, and U. Leicht-Deobald. 2016. The psychology of psychic distance: Antecedents of asymmetric perceptions. *Journal of World Business* 51 (2): 308–318.

Harzing, A.W., and M. Pudelko. 2016. Do we need to distance ourselves from the distance concept? Why home and host country context might matter more than (cultural) distance. *Management International Review* 56 (1): 1–34.

He, X., K.E. Brouthers, and I. Filatotchev. 2013. Resource-based and institutional perspectives on export channel selection and export performance. *Journal of Management* 39 (1): 27–47.

Hendriks, G., A.H.L. Slangen, and P.M.A.R. Heugens. 2017. How a firm's domestic footprint and domestic environmental uncertainties jointly shape added cultural distances: The roles of resources dependence and headquarters attention. *Journal of Management Studies* 55 (6): 883–909.

Hennart, J.F., and J. Larimo. 1988. The impact of culture on the strategy of multinational enterprises: Does national origin affect ownership decisions? *Journal of International Business Studies* 29 (3): 515–538.

Hofstede, G. 1980. *Culture's consequences: International differences in work-related values*. Beverly Hills: Sage.

———. 2001. *Culture's consequences; comparing values, behaviors, institutions, and organizations across nations*. Thousand Oaks: Sage.

———. 2010. The Globe debate: Back to relevance. *Journal of International Business Studies* 41 (8): 1339–1346.

Hofstede, G., G.J. Hofstede, and M. Minkov. 2010. *Cultures and organizations: Software of the mind*. 3rd ed. McGraw-Hill: New York.

House, R.J., P.J. Hanges, M. Javidan, P.W. Dorfman, and V. Gupta, eds. 2004. *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks: Sage.

Hutzschenreuter, T., J.C. Voll, and A. Verbeke. 2011. The impact of added cultural distance and cultural diversity on international expansion patterns: A Penrosean perspective. *Journal of Management Studies* 48 (2): 305–329.

Hutzschenreuter, T., I. Kleindienst, and S. Lange. 2014. Added psychic distance stimuli and MNE performance: Performance effects of added cultural, governance, geographic, and economic distance in MNEs' international expansion. *Journal of International Management* 20 (1): 38–54.

Inglehart, R. 1990. *Culture shift in advanced industrial society*. Princeton: Princeton University Press.

———. 1997. *Modernization and postmodernization: Cultural, economic, and political change in 43 societies*. Princeton: Princeton University Press.

Inglehart, R., and W.E. Baker. 2000. Modernization, cultural change and the persistence of traditional values. *American Sociological Review* 65 (1): 19–51.

Inglehart, R., and C. Welzel. 2005. *Modernization, cultural change and democracy*. New York: Cambridge University Press.

Johanson, J., and J.E. Vahlne. 1977. The internationalization process of the firm—A model of knowledge development and increasing foreign market commitments. *Journal of International Business Studies* 8 (1): 23–32.

———. 2009. The Uppsala internationalization process model revisited: From liability of foreignness to liability of outsidership. *Journal of International Business Studies* 40 (9): 1411–1431.

Johanson, J., and F. Wiedersheim-Paul. 1975. The internalization of the firm: Four Swedish cases. *Journal of Management Studies* 12 (3): 305–322.

Kaufmann, D., A. Kraay, and M. Mastruzzi. 2008. Governance matters VII: Aggregate and individual governance indicators 1996–2007. Policy research working paper 4654. Washington, DC: The World Bank.

Kirkman, B.L., K.B. Lowe, and C.B. Gibson. 2006. A quarter century of culture's consequences: A review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies* 37 (3): 285–320.

———. 2017. A retrospective on Culture's consequences: The 35-year journey. *Journal of International Business Studies* 48 (1): 12–29.

Klopf, P., and P.C. Nell. 2018. How "space" and "place" influence subsidiary host country political embeddedness. *International Business Review* 27 (1): 186–197.

Kogut, B., and H. Singh. 1988. The effect of national culture on the choice of entry mode. *Journal of International Business Studies* 19 (3): 411–432.

Kostova, T. 1996. Success of the transnational transfer of organizational practices within multinational companies. Doctoral dissertation, University of Minnesota.

Kostova, T., P.C. Nell, and A.K. Hoenen. 2017. Understanding agency problems in headquarters-subsidiary relationships in multinational corporations: A contextualized model. *Journal of Management* 3 (1): 57–81.

Lavie, D., and S.R. Miller. 2008. Alliance portfolio internationalization and firm performance. *Organization Science* 19 (4): 623–646.

Lenartowicz, T., and K. Roth. 2001. Does subculture within a country matter? A cross-culture study of motivational domains and business performance in Brazil. *Journal of International Business Studies* 32 (2): 305–325.

Li, J., X. Liu, M. Wright, and I. Filatotchev. 2014. International experience and FDI location choices of Chinese firms: The moderating effects of home country government support and host country institutions. *Journal of International Business Studies* 45 (4): 428–449.

Magnusson, P., D.W. Baack, S. Zdravkovic, K.M. Staub, and L.S. Amine. 2008. Meta-analysis of cultural differences: Another slice at the apple. *International Business Review* 17 (5): 520–532.

Mahalanobis, P.C. 1937. Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. *Sankhya* 3: 35–40.

Makino, S., and E.W.K. Tsang. 2011. Historical ties and foreign direct investment: An exploratory study. *Journal of International Business Studies* 42 (4): 545–557.

Malhotra, S., and A. Gaur. 2014. Spatial geography and control if foreign acquisitions. *Journal of International Business Studies* 45 (2): 191–210.

Maseland, R., D. Dow, and P. Steel. 2018. The Kogutand Singh national cultural distance index: Time to start using it as a springboard rather than a crutch. *Journal of International Business Studies*. https://doi.org/10.1057/s41267-018-0183-3.

Matsumoto, D., and L.H. Yoo. 2006. Toward a new generation of cross-cultural research. *Perspectives in Psychological Science* 1 (3): 234–250.

McSweeney, B. 2002. Hofstede's model of national cultural differences and their consequences: A triumph of faith—A failure of analysis. *Human Relations* 55: 89–118.

———. 2009. Dynamic diversity: Variety and variation within countries. *Organization Studies* 30 (9): 933–957.

Meyer, K., A. van Witteloostuijn, and S. Beugelsdijk. 2017. What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. *Journal of International Business Studies* 48 (5): 535–551.

Mingo, S., F. Morales, and L. Dau. 2018. The interplay of national distances and regional networks: Private equity investments in emerging markets. *Journal of International Business Studies* 49 (3): 371–386.

Morschett, D., H. Schramm-Klein, and B. Swoboda. 2010. Decades of research on market entry modes: What do we really know about external antecedents of entry mode choice? *Journal of International Management* 16 (1): 60–77.

Nell, P.C., and B. Ambos. 2013. Parenting advantage in the MNC: An embeddedness perspective on the value added by headquarters. *Strategic Management Journal* 34 (9): 1086–1103.

Nell, P.C., P. Kappen, and T. Laamanen. 2017. Reconceptualising hierarchies: The disaggregation and dispersion of headquarters in multinational corporations. *Journal of Management Studies* 54 (8): 1121–1143.

Oyserman, D., H.M. Coon, and M. Kemmelmeier. 2002. Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin* 128 (1): 3–72.

Peterson, M. 2003. Reviewed work: Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations by Geert Hofstede. *Administrative Science Quarterly* 48 (1): 127–131.

———. 2004. Reviewed work: Culture, leadership and organizations: The Globe study of 62 societies. *Administrative Science Quarterly* 49 (4): 641–647.

Peterson, M., and S.L. Castro. 2006. Measurement metrics at aggregate levels of analysis: Implications for organizational culture research and the Globe project. *The Leadership Quarterly* 17: 506–521.

Peterson, M., and M. Søndergaard. 2011. Traditions and transitions on quantitative societal culture research in organization studies. *Organization Studies* 32 (11): 1539–1558.

Peterson, M., M. Søndergaard, and A. Kara. 2018. Traversing cultural boundaries in IB: The complex relationships between explicit country and implicit cultural group boundaries at multiple levels. *Journal of International Business Studies*. https://doi.org/10.1057/s41267-017-0082-z.

Ralston, D.A. 2008. The crossvergence perspective: Reflections and projections. *Journal of International Business Studies* 39 (1): 27–40.

Ralston, D.A., D.H. Holt, R.H. Terpstra, and Y. Kai-Cheng. 1997. The impact of national culture and economic ideology on managerial work values: A study of the United States, Russia, Japan, and China. *Journal of International Business Studies* 28 (1): 177–208.

Ralston, D.A., et al. 2011. A twenty first century assessment of values across the global workforce. *Journal of Business Ethics* 104: 10–31.

———. 2014. Societal level versus individual level predictions of ethical behavior: A 48-society study of collectivism and individualism. *Journal of Business Ethics* 122: 283–306.

Ronen, S., and O. Shenkar. 2013. Mapping world cultures: Cluster formation, sources and implications. *Journal of International Business Studies* 44 (9): 867–897.

Salomon, R. 2016. *Global vision: How companies can overcome the pitfalls of globalization*. Basingstoke: Palgrave Macmillan.

Salomon, R., and Z. Wu. 2012. Institutional distance and local isomorphism strategy. *Journal of International Business Studies* 43 (4): 343–367.

Schwartz, S.H. 1994. Beyond individualism/collectivism; new cultural dimensions of values. In *Individualism and collectivism. Theory, method, and applications*, ed. U. Kim, H.C. Triandis, C. Kâğitçibasi, S.C. Choi, and G. Yoon, 85–119. Thousand Oaks: Sage.

———. 1999. Cultural value differences: Some implications for work. *Applied Psychology. An International Review* 48 (1): 23–47.

———. 2004. Mapping and interpreting cultural differences around the world. In *Comparing cultures: Dimensions of culture in a comparative perspective*, ed. H. Vinken, J. Soeters, and P. Ester, 43–73. Leiden: Brill.

———. 2006. A theory of cultural value orientations: Explication and applications. *Comparative Sociology* 5 (2): 137–182.

———. 2011. Studying personal values: Personal adventure, future directions. *Journal of Cross-Cultural Psychology* 42 (2): 307–319.

———. 2014. Rethinking the concept and measurement of societal culture in light of empirical findings. *Journal of Cross-Cultural Psychology* 45 (1): 5–13.

Shenkar, O. 2001. Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences. *Journal of International Business Studies* 32 (3): 519–535.

———. 2012. Beyond cultural distance: Switching to a friction lens in the study of cultural differences. *Journal of International Business Studies* 43 (1): 12–17.

Shenkar, O., Y. Luo, and O. Yeheskel. 2008. From "distance" to "friction": Substituting metaphors and redirecting intercultural research. *Academy of Management Review* 33 (4): 905–923.

Sivakumar, K., and C. Nakata. 2001. The stampede towards Hofstede's framework; Avoiding the sample design pit in cross-cultural research. *Journal of International Business Studies* 32 (3): 555–574.

Slangen, A.H.L., and S. Beugelsdijk. 2010. The impact of institutional hazards on foreign multinational activity: A contingency perspective. *Journal of International Business Studies* 41 (6): 980–995.

Smith, P.B. 2006. When elephants fight, the grass gets trampled: The Globe and Hofstede projects. *Journal of International Business Studies* 37 (6): 915–921.

Smith, P.B., S. Dugan, and F. Trompenaars. 1996. National culture and the values of organizational employees. A dimensional analysis across 43 nations. *Journal of Cross-Cultural Psychology* 27 (2): 231–264.

Stahl, G., and R. Tung. 2015. Towards a more balanced treatment of culture in international Business studies: The need for positive cross-cultural scholarship. *Journal of International Business Studies* 46 (4): 391.

Stahl, G., and A. Voigt. 2008. Do cultural differences matter in mergers and acquisitions? A tentative model and examination. *Organization Science* 19 (1): 160–176.

Steenkamp, J.B. 2001. The role of national culture in international marketing research. *International Marketing Review* 18 (1): 30–44.

Tihanyi, L., D.A. Griffith, and C.J. Russell. 2005. The effect of cultural distance on entry mode choice, international diversification, and MNE performance: A meta-analysis. *Journal of International Business Studies* 36 (3): 270–283.

Tung, R.L., and A. Verbeke. 2010. Beyond Hofstede and GLOBE: Improving the quality of cross-cultural research. *Journal of International Business Studies* 41 (8): 1259–1274.

Vahlne, J.E., and F. Wiedersheim-Paul. 1973. Economic distance: Model and empirical investigation. In *Export and foreign establishments*, ed. E. Hörnell, J.E. Vahlne, and F. Wiedersheim-Paul. Uppsala: University of Uppsala.

Valentino, A., J. Schmitt, B. Koch, and P.C. Nell. 2018. Leaving home: An institutional perspective on intermediary HQ relocations. *Journal of World Business*. https://doi.org/10.1016/j.jwb.2018.08.004.

Van Hoorn, A.A.J., and R. Maseland. 2016. How institutions matter for international business: Institutional distance effects vs institutional profile effects. *Journal of International Business Studies* 47 (3): 374–381.

Verbeke, A., W. Yuan, and L. Kano. in press. A values-based analysis of bifurcation bias and its impact on family firm internationalization. *Asia Pacific Journal of Management*. https://doi.org/10.1007/s10490-018-9598-4.

Wu, Z., and R. Salomon. 2016. Does imitation reduce the liability of foreignness? Linking distance, isomorphism and performance. *Strategic Management Journal* 37 (12): 2441–2462.

Xie, Z., and J. Li. 2017. Exporting and innovating among emerging market firms: The moderating role of institutional development. *Journal of International Business Studies* 49 (2): 222–245.

Xu, D., and O. Shenkar. 2002. Institutional distance and the multinational enterprise. *Academy of Management Review* 27 (4): 608–618.

Yu, S., S. Beugelsdijk, and J. de Haan. 2015. Trade, trust and the rule of law. *European Journal of Political Economy* 37: 102–115.

Zaheer, A., and E. Hernandez. 2011. The geographic scope of the MNC and its alliance portfolio: Resolving the paradox of distance. *Global Strategy Journal* 1 (1–2): 109–126.

Zaheer, S., M. Schomaker, and L. Nachum. 2012. Distance without direction: Restoring credibility to a much-loved construct. *Journal of International Business Studies* 43 (1): 18–27.

Zhao, H., Y. Luo, and T. Suh. 2004. Transaction cost determinants and ownership-based entry mode choice: A meta-analytical review. *Journal of International Business Studies* 35 (6): 524–544.

27

Distance in International Business Research: A Commentary

Mark F. Peterson and Yulia Muratova

Introduction

Distance is ordinarily theorized as representing uncertainty, trust and transaction costs, and learning potential. Drawing theoretical conclusions from distance, however, is threatened by the very different estimates of these constructs that are produced when using different reference countries and different dimension schemes. Beugelsdijk, Ambos and Nell (2018) provide insightful recommendations about how to conduct distance research in international business (IB).

We offer two additions to their analysis. One is to assess whether their recommendations based on distance between both multiple home and multiple host countries apply in studies of distances between a single home

Beugelsdijk, S., B. Ambos, and P.C. Nell. 2018. Conceptualizing and measuring distance in international business research: Recurring questions and best practice guidelines. *Journal of International Business Studies* 49 (9): 1113–1137.

M. F. Peterson • Y. Muratova (✉)
Aarhus University, Aarhus, Denmark
e-mail: yulia.muratova@mgmt.au.dk

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3_27

or host country and multiple other countries. We also consider distance based on cultural dimension models besides those that they analysed.

Distances from One Home or Host Reference Country

Most IB distance studies compare one home or host country with multiple other countries (Harzing and Pudelko 2016). For example, Kogut and Singh (1988) compare US culture scores with scores for the countries of businesses that had entered the United States. In contrast, Beugelsdijk, Ambos and Nell (2018) analyse the rarer situation of studying distances between both multiple host and multiple home countries.

Studying distances from a single home or host country requires the reference country's scores to be extremely accurate (Peterson 2003). Otherwise, the reliability of all distance scores based on the reference country is threatened. Studying pairs of multiple home and host countries, as do Beugelsdijk, Ambos and Nell (2018), faces the much less demanding requirement that the country values are random variables. We illustrate our points using four different reference countries by providing tables that parallel tables in Beugelsdijk, Ambos and Nell (2018). The countries China, South Korea, the United States and the Netherlands, are among the most frequently studied in distance research (Harzing and Pudelko 2016; with South Korea taking the place of Japan for which no data were available for Ralston, Egri, et al.'s (2011) project).

We also use somewhat different culture dimension schemes for distance estimates than do Beugelsdijk, Ambos and Nell (2018). We include two pairs of societal value dimensions that are often used as alternatives to one another. One pair is the Hofstede, Hofstede and Minkov (2010) dimensions and the GLOBE "as is" dimensions (House et al. 2004). The other is the Schwartz Value Survey (SVS) domains based on data from high school teachers and university students (Schwartz 2006, 2008), and from the managers and professionals studied by Ralston, Egri and colleagues (2011). A fifth culture framework is based on *sources of guidance* (SOGs) that are modified from role theory categories to assess managers'

reliance of eight SOGs (Peterson et al. 2016; Smith et al. 2002; Smith et al. 1996).

Comparing Distance Formulas

Beugelsdijk, Ambos and Nell (2018) compare how the Kogut-Singh Index (KSI) and Euclidean distance (ED) formulas handle variance differences and explain how Mahalanobis distance (MD) adjusts for covariances among the dimensions used in a distance metric. We compare correlations for distance using the KSI, ED and MD formulas with those provided by Beugelsdijk, Ambos and Nell (2018) to see if the results are similar when a single reference country is used. For the dimensions and reference countries in our analysis, the correlations between the KSI- and ED-based formulas range from 0.88 to 0.98 (not shown) and do not differ markedly among dimensions used or reference countries. These correlations are very consistent with those of 0.89 to 1.00 reported by Beugelsdijk, Ambos and Nell (2018).

As shown in Table 27.1, the correlations between the ED- and MD-based formulas differ substantially by reference country and dimension framework. The range from 0.55 to 0.89 is similar to the range of 0.61 to 0.88 in the Beugelsdijk, Ambos and Nell (2018) analysis. The median correlation (the average of the second and third ranked correlations) between the ED and MD formulas for distance based on our five culture-dimension schemes ranges from 0.61 (GLOBE "as is") to

Table 27.1 Pairwise correlations between Euclidean distance (*ED*) and Mahalanobis distance (*MD*) for four reference countries

Distance dimensions	China	South Korea	United States	Netherlands	N of country pairs
Sources of guidance	0.75	0.72	0.56	0.78	60
Hofstede 6 dimensions	0.89	0.83	0.88	0.83	81
GLOBE "as is"	0.59	0.67	0.63	0.55	61
Schwartz's SVS sample	0.64	0.57	0.65	0.75	79
Ralston/Egri's SVS sample	0.60	0.71	0.70	0.84	49

0.85 (Hofstede). For each reference country, the median correlations are China, 0.64, Korea 0.71, the United States 0.65, and the Netherlands 0.78. The correlations between the KSI, ED and MD formulas support Beugelsdijk, Ambos and Nell's (2018) contention that the Mahalanobis distance correction is often helpful.

Other Distance Dimensions

In their fourth table, Beugelsdijk, Ambos and Nell (2018) raise the question of whether different dimension schemes offer roughly comparable conclusions about distance. Similarly, we show in Tables 27.2a and 27.2b correlations for MD-corrected distance measures using our different set of dimensions and reference countries. Beugelsdijk, Ambos and Nell (2018) show correlations between distance measures ranging from 0.01 to 0.32. We find correlations ranging even more substantially from −0.29 to 0.68. We find the greatest similarity between distance estimates based on the original SVS scores and on the SVS scores developed by Ralston, Egri and colleagues (2011). For our four reference countries, these correlations range from 0.38 to 0.68 with a median of 0.50. For distances

Table 27.2a Pairwise correlations between different cultural distance indices (Mahalanobis corrected) for China (below the diagonal) and South Korea (above the diagonal)

	Sources of guidance	Hofstede	GLOBE "as is"	Schwartz	Ralston/ Egri
Sources of guidance		−0.29*	0.10	−0.12	0.13
Hofstede 6 dimensions	0.13		0.07	−0.12	0.11
GLOBE "as is"	0.19	0.02		0.15	−0.19
Schwartz's SVS sample	0.09	0.27*	−0.20		0.45**
Ralston/Egri's SVS sample	−0.06	0.42**	−0.13	0.38*	

Notes: *N* ranges from 33 to 55 depending on matching country pairs
*$p<0.05$; **$p<0.01$

Table 27.2b Pairwise correlations between different cultural distance indices (Mahalanobis corrected) for the United States (below the diagonal) and the Netherlands (above the diagonal)

	Sources of guidance	Hofstede	GLOBE "as is"	Schwartz	Ralston/ Egri
Sources of guidance		0.21	0.29	0.46**	0.29
Hofstede	0.20		0.12	0.16	0.14
GLOBE "as is"	0.40*	0.47**		0.28	0.26
Schwartz's SVS samples	0.06	0.19	0.34*		0.55**
Ralston/Egri's SVS samples	−0.07	0.13	0.19	0.68**	

Notes: *N* ranges from 33 to 55 depending on matching country pairs
*$p<0.05$; **$p<0.01$

based on our other culture dimensions, the correlations range from −0.29 to 0.47 depending on the dimension scheme and reference country. No dimension scheme provides distance estimates having consistent positive associations with the others. No one reference country shows more consistently positive correlations between the distance estimates based on different dimension schemes.

Conclusion

The recommendations that Beugelsdijk, Ambos and Nell (2018) make about when to use a Mahalanobis distance correction are found to be just as sound for studies referencing distance against a single home or host country as in their analysis of multiple pairs of home and host countries. Their cautions about selecting and interpreting distance based on any one culture framework apply even more strongly to distance studies using a single reference home or host country. We make two additions to their recommendations:

- Occasionally, distance from a single reference country may be meaningfully estimated for some of the country's cultural or institutional characteristics that can be very confidently demonstrated.

- Otherwise, when estimates for a reference country ask more than our data can provide, our best course may be to study cultural and institutional characteristics as variables representing other countries with which it does business (Harzing and Pudelko 2016).

References

Beugelsdijk, S., B. Ambos, and P.C. Nell. 2018. Conceptualizing and measuring distance in international business research: Recurring questions and best practice guidelines. *Journal of International Business Studies* 49 (9): 1113–1137.

Harzing, A.W., and M. Pudelko. 2016. Do we need to distance ourselves from the distance concept? Why home and host country context might matter more than (cultural) distance. *Management International Review* 56 (1): 1–34.

Hofstede, G., G.J. Hofstede, and M. Minkov. 2010. *Cultures and organizations: Software of the mind*. 3rd ed. New York: McGraw-Hill.

House, R.J., P.J. Hanges, M. Javidan, P.W. Dorfman, and V. Gupta, eds. 2004. *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Thousand Oaks, CA: Sage.

Kogut, B., and H. Singh. 1988. The effect of national culture on the choice of entry mode. *Journal of International Business Studies* 19 (3): 411–432.

Peterson, M.F. 2003. Book review of "Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations, 2d ed." by Geert Hofstede. Thousand Oaks, CA: Sage, 2001. *Administrative Science Quarterly* 48 (1): 127–131.

Peterson, M.F., T.S. Barreto, and P.B. Smith. 2016. Revised sources of guidance measures: Six events and demographic controls. In *Unity, diversity and culture*, ed. C. Roland-Lévy, P. Denoux, B. Voyer, P. Boski, and W.K. Gabrenya Jr., 213–217. Melbourne: International Association for Cross-Cultural Psychology. www.iaccp.org. Accessed 4 Feb 2019.

Ralston, D.A., C. Egri, et al. 2011. A 21st century assessment of values across the global workplace. *Journal of Business Ethics* 104 (1): 1–31.

Schwartz, S.H. 2006. A theory of cultural value orientations: Explication and applications. *Comparative Sociology* 5 (2): 137–182.

———. 2008. The 7 Schwartz cultural value orientation scores for 80 countries. https://www.researchgate.net/publication/304715744_The_7_Schwartz_cultural_value_orientation_scores_for_80_countries

Smith, P.B., M.F. Peterson, and S. Schwartz. 2002. Cultural values, sources of guidance and their relevance to managerial behavior: A 47 nation study. *Journal of Cross Cultural Psychology* 33 (1): 188–208.

Smith, P.B., M.F. Peterson, and Z.M. Wang. 1996. The manager as mediator of alternative meanings: A pilot study from China, the U.S.A. and U.K. *Journal of International Business Studies* 27 (1): 115–137.

Index[1]

[1] Note: Page numbers followed by 'n' refer to notes.

L. Eden et al. (eds.), *Research Methods in International Business*, JIBS Special Collections, https://doi.org/10.1007/978-3-030-22113-3

D