

THE NECESSITY, LOGIC, AND FORMS OF REPLICATION

RICHARD A. BETTIS,¹ CONSTANCE E. HELFAT,^{2*} and J. MYLES SHAVER³

¹ Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina, U.S.A.

² Tuck School of Business at Dartmouth, Hanover, New Hampshire, U.S.A.

³ Carlson School of Management, University of Minnesota, Minneapolis, Minnesota, U.S.A.

Research summary: A replication study assesses whether the results of a particular prior study can be reproduced, including in new contexts with different data. Replication studies are critical for building a cumulative body of research knowledge. This article discusses and provides a typology of different types of replications, compares replications with other approaches to cumulating knowledge, and provides guidelines toward producing high-quality replication studies. The articles in this Special Issue provide examples of replication studies in strategic management.

Managerial summary: Research studies sometimes draw implications for managerial practice. A single empirical study, however, is specific to a particular context, relies on a particular set of data, and uses a particular research design. As a consequence, a single study cannot establish whether the findings generalize to a different context and whether the research design is robust to alternative approaches. Replication studies can help to establish the range of applicability of prior studies and better support what implications can be drawn for managerial practice. Copyright © 2016 John Wiley & Sons, Ltd.

INTRODUCTION

This Replication Special Issue represents an important and concrete step in the commitment by *The Strategic Management Journal* (SMJ) to making published research consistent with both the proper practice of the statistical theory of hypothesis testing and good science as discussed in the SMJ editorial, “Creating repeatable, cumulative knowledge in strategic management” (Bettis *et al.*, 2016).

We started this Special Issue with a degree of what we now realize was unjustified confidence. After many months of work, we now look back at our naiveté regarding real replications. It never

occurred to us how many new and different issues the studies would raise when we started receiving them. This facilitated considerable learning on our part regarding different forms of replication, best practices for replications, a myriad of technical details about replications, and how to make editorial judgments for replication studies. Our intention is to summarize the most important of this learning in this article. After reading this article, we hope you will realize how important replication is to the empirical credibility of strategic management scholarship going forward. We hope those who read the replications in this Special Issue will be motivated to pursue replication as part of their own research. It is the responsibility of all of us to make our research base as repeatable and cumulative as possible.

In recent years, there has been a growing recognition in a large number of fields of the

Keywords: replication; cumulative research; statistical evidence; guidelines; quasi-replication

*Correspondence to: Constance E. Helfat, Tuck School of Business at Dartmouth, 100 Tuck Hall, Hanover, NH 03755, U.S.A. E-mail: constance.helfat@dartmouth.edu

necessity for replication studies of statistical results. In biomedical science and psychology, a substantial percentage of articles in highly cited journals cannot be replicated successfully (e.g., Ioannidis, 2005; Open Science Collaboration, 2015). One important reason for this lack of replicability is that many highly cited journals favor publishing results with significant coefficients.¹ Some only will publish results with significant coefficients. Once a significant coefficient is found, then replications are not common in many highly cited journals. There are two severe problems with this approach.² First, it ignores the fact that statistical results only apply to the particular sample and nothing definitive can be said by *sample* statistics about the population. Because of sample variation, another sample from the same population, or even several more samples, might not find a significant result, but such “replications” are uncommon because they are often not publishable in top journals. This means that Popper’s logic of scientific discovery (i.e., falsifiability), the widely accepted definition of science, is not met. A replication with a nonsignificant coefficient testing a theoretical or predictive hypothesis³ is unlikely to be published in a highly cited journal.⁴ In parallel, it is important to note that any one study that fails to replicate the previous result does not mean that the previous result has been rejected. It does mean that the balance of evidence regarding the existing results moves toward questioning the original result.⁵ The extent to which the balance of evidence changes should depend on several factors related to the quality of the replication that we highlight below. Nevertheless, this outcome suggests further study of the issue is likely necessary. Such an outcome might also motivate theoretical advances. Generally, the aim of replication is to provide additional evidence, not to overturn prior studies.

Second, the emphasis on finding significant coefficients defined by arbitrary levels of significance (e.g., 0.05 or 0.001) does not have a scientific basis,

and furthermore, can lead to specification searches to find “significant” coefficients for which hypotheses can subsequently be formulated. If reported as hypothesis testing and not exploratory research, specification searches destroy the logic and efficacy of hypothesis tests and again violate Popperian falsifiability. Specification searches completely invalidate the *t*-statistics and the associated *p*-values. Hence, they report invalid results. (See Bettis [2012], and Goldfarb and King [2016] for extensive discussion and appropriate references.)

With this background, we next proceed to elaborate and discuss the many different types of replication that can be useful. We then compare the benefits of replication with other approaches to cumulating knowledge such as meta-analyses. Third, we turn to a brief discussion of each article and the contribution it makes to the Special Issue. Then, we discuss guidelines toward producing high-quality replication studies. Finally, we conclude with a few summary thoughts about the importance of replication in increasing the repeatability and cumulativeness of empirical research published in the *Strategic Management Journal* and other journals.

TYPES OF REPLICATIONS

In talking with colleagues, we have learned that conceptions of replication range from using the exact same data with the exact same analysis as in a prior study, to producing results generally aligned with prior research, with many variants in between. Strictly speaking, a study that contains a finding similar to that in previous research is not a replication. Instead, the reference point for a replication is a particular prior study (see, e.g., Hubbard, Vetter, and Little, 1998; Singh, Ang, and Leong, 2003). Beyond that, it is helpful to distinguish among different types of replications and what we can learn from each.

Tsang and Kwan (1999) proposed that replications vary according to the source of the data and the method of measuring constructs and analyzing data. We find these distinctions helpful and adopt a similar classification of replications along two dimensions depicted in Figure 1: (1) the similarity of the data and empirical setting to that in the original study, and (2) the similarity of the research design to that of the original study. Replications can vary in the closeness to the original studies on either or both of these dimensions, and variation along these

¹ This is definitely changing in psychology and biomedical science as this is being written.

² We note that the *Strategic Management Journal* has taken corrective policy action for both these two problems.

³ For distinction between theoretical and predictive hypotheses, see Bettis *et al.* (2014a).

⁴ This applies at the time of writing, but appears to be changing across a variety of fields.

⁵ In this regard, it is easy to see why well-executed and novel statistical model results that fail to show small *p*-values for the coefficient estimates being tested should be published.

	Same Research Design	Different Research Design
Same Data and Sample		
Same Population (Same Context) with Different Sample		
Different Population (Different Context)		

Figure 1. Dimensions of replication.
Row and column headings adapted from Tsang and Kwan (1999)

dimensions is associated with different benefits. First, replications may vary in whether they use the same research design as the original studies, including approach, measures, method, and model. In addition, replications may vary in the data that they use, including: the exact same data as in the original study, different data drawn from the same population as in the original study, or data drawn from a different population (Tsang and Kwan, 1999). In Figure 1, a “population” or “research context” refers to a specific setting that corresponds to a particular population of actors. For example, a particular industry, in a particular geographic area, in a particular time period, with a particular type of actor (e.g., firm, manager, etc.) would constitute a research context that would have a certain population.

The narrowest form of replication uses the exact same data as in the original study and the exact same research design. This sort of replication can tell us that the original study may contain errors or the results may be sensitive to the statistical software employed. Although there have been some high-profile studies with mistakes or deliberate misrepresentations in the data or analysis, the consensus is that this is a relatively small percentage of published studies. A slightly less narrow form of replication uses different data, but drawn from the same population, with the same research design as in the original study. This type of study can inform us about the reliability and representativeness of the data for a specific population; if studies of the same population differ only in the data sample but produce different results, the validity of the results may warrant further investigation. These two forms of narrow replication are depicted in Figure 2.

Often scholars would like to know if the results of a particular study hold in a different research

context (and population) or using different research designs. Studies that assess whether the results of an earlier study hold when using different research designs, but the same data or new data drawn from the same population, can inform us about the robustness of the original results to alternative measures, models, and methods. For example, using alternative measures (e.g., different proxy variables) that seek to capture the same concepts or constructs as in the original study can show how sensitive the results are to differences in measurement. Results may also be sensitive to alternative methods and models. We expect that as statistical methods improve, earlier findings may not always replicate, but emphasize that this is not a failing of the original study if the methods were unavailable or difficult to implement at the time.

Replications can also help us to understand how well the results of a prior study generalize to other settings. For this reason, some replications use data from a different population than in the original study, while using the same research design. Replications of this type can help us to understand whether the original results were idiosyncratic to a particular setting or not. And studies that use data on a new population as well as a different research design simultaneously provide information about the generalizability of the original study and the robustness of the measures, method, and model in the new population.

We use the term *quasi-replication* to refer to studies that assess robustness and/or generalizability, as shown in Figure 3. We believe that these types of studies hold especially strong promise for the field of strategic management, because quasi-replications inform us about how well results hold up in multiple settings, using a variety of

	Same Research Design	Different Research Design
Same Data and Sample	Checking for errors and/or falsification of results	
Same Population (Same Context) with Different Sample	Reliability and representativeness of data	
Different Population (Different Context)		

Figure 2. Narrow replication

	Same Research Design	Different Research Design
Same Data and Sample		Robustness to different measures, methods, and models
Same Population (Same Context) with Different Sample		Robustness to different measures, methods, and models
Different Population (Different Context)	Generalize to new population (subjects, industry, time period, etc.)	Generalize to new population and assess robustness

Figure 3. Quasi-replication

measures and methods. However, if we are to understand why the results of a prior study do or do not hold up in a particular replication, it is important to systematically alter the data, measures, method, and model in stages. If instead a replication alters the population, the measures, the method, and the model all at once and the original results do not replicate, we will have no idea why—and therefore, will learn little.

A high-quality replication therefore moves in stages. For a quasi-replication that seeks to understand how well prior results generalize to a new population, it works best to first replicate the original study with data drawn from a population that differs from the original in only one or a few dimensions (e.g., a different time period for the same industry with the same size firms in the same geographic location), using the same or similar measures and the same method and model. This makes it possible to clearly identify any differences in the results due to composition of the sample.

Then, the replication can go on to alter one or a few measures, or alter the empirical methodology or model, again in stages holding the other elements of the study constant, so that any differences in the results can be more clearly attributed to the particular measures, method, or model. For narrow replications that use data from the same population, it is preferable to alter the measures or method or model one at a time rather than all at once, again holding the other factors constant.⁶ This makes it possible to assess the robustness of the original study to changes in specific elements of the research design.

These dimensions of replication apply to studies using observational (archival), experimental, and qualitative data. Although researchers may worry that qualitative studies could be difficult to replicate, prior research has shown that quasi-replications in

⁶ However, in some cases, such as for qualitative studies, the measures and the methods go together (e.g., for quantitative coding of cases and analysis), so varying one may entail varying the other.

different populations using the same methods as in the original studies are feasible (see e.g., Kurke and Aldrich, 1983). Replications of qualitative studies could also use alternative methods and measures for similar populations (for an overview of alternative qualitative methods, see Bettis *et al.*, 2014b). For any of these types of replications, studies that confirm the original results are as important those that do not. Moreover, as noted earlier, a single replication is only one additional piece of information. Multiple replications of a prior study may be needed in order to draw strong conclusions. Finally, replications need not, and most often will not, contain new theory. And replications need not be of research that test hypotheses. However, we expect replication studies in strategic management to help scholars gain a better understanding of the phenomena and relationships that we study, and to have implications for future research.

DEDICATED REPLICATION STUDIES COMPARED TO OTHER APPROACHES IN THE LITERATURE

As we have discussed, replication studies play an important role in advancing a cumulative body of empirically supported research. Some readers might consider that existing approaches in the literature share qualities of the dedicated replication studies that appear in this Special Issue. We draw attention to two alternate approaches that scholars have used in order to highlight the limitations of these approaches for achieving the goals of dedicated replication studies.

Meta-analyses

Meta-analyses pool findings from several studies in order to assess the significance and effect size of relationships between variables. Some scholars consider this in the spirit of replication because meta-analysis is only possible if multiple studies examine the same relationship. Moreover, meta-analysis provides an analytical technique to compile results across multiple studies. Meta-analyses can benefit scholarship in strategic management by informing us about patterns in a group of prior empirical studies, but replications have other benefits not available through meta-analysis, especially given the nature of empirical work in the strategy field.

Meta-analyses focus on assessing the underlying correlation between variables. Of course, establishing a correlation between variables does not equate to establishing whether or not the correlation is spurious. This concern is pronounced when the data in the underlying studies are nonexperimental in nature. Nonexperimental data comprise the vast majority of empirical studies in the strategy field. As empirical scholarship on a topic evolves, one form of advancement is employing research designs that better identify what mechanisms drive the correlation between variables. Therefore, one form of dedicated replication study that we see in the Special Issue examines whether previous conclusions hold when the analysis is subjected to advances in research designs, including statistical techniques. Such insights, which acknowledge the nature of the data we use and the evolution of findings in our field, are not possible through meta-analyses.

In addition, many studies in the strategy field rely on data where the measurement of constructs is not well-refined with common scales. Authors will often use context-specific measures that capture general constructs as they apply within a particular industry, yet the specific measures may not be meaningful in other industries. For example, a measure of incremental technological innovation may depend on the technology in question, which differs by industry. To the extent that studies use very different measures for the same construct, it is not clear if there should be the same underlying relationship between operationalized constructs across settings. As a result, due to the nature of empirical strategy research, meta-analysis may be most applicable to the subset of studies in strategic management that rely on common measures of constructs.

In summary, our view is that although meta-analyses provide useful information about prior research on a topic, the nature of the questions that we ask, and the nature of the data that we collect in the strategy field limit the use of meta-analyses as replications. We believe that dedicated replication studies are better suited for this task.

Revisiting previous findings en route to testing novel hypotheses

Studies in our field sometimes re-examine existing findings as an intermediate step to advancing novel hypotheses, so as to highlight specific conditions where the existing relationship will or will not hold (e.g., Shaver and Flyer, 2000), or to highlight

additional relationships beyond the original one. Revisiting previous findings can be considered in the spirit of replication, but this type of general confirmation or disconfirmation of prior research is often not a replication in the sense we previously described. Moreover, these studies tend to provide a more limited scope than the dedicated replication studies in this volume. We reach this conclusion for the following three reasons.

First, we only observe these studies when authors believe that the existing findings provide only a partial explanation of a phenomenon or test of a theory, or apply under some conditions but not others. In the latter case, the existing findings are set up to be invalidated by more precise hypotheses. If we rely solely on such studies to assess the results of prior research, then we will only see them when the authors' focus is to falsify the existing state of empirical knowledge. However, replications that confirm the results of prior research are as important as those that call prior results into question.

Second, we only observe these studies when authors propose hypotheses that build on or modify an established theoretical rationale. However, replication need not be theoretically motivated to be useful in building a cumulative body of research knowledge. Establishing the universality of empirical relationships is often a key insight from which theory can be advanced.

Third, these studies tend to move quickly from the tests of existing findings to the tests of the novel hypotheses. Therefore, a careful replication of existing findings is at best a secondary focus. As a result, these studies often lack the discipline of a dedicated replication study in trying to match the replication to the initial study and examining a host of reasons why results replicate or not. For instance, these studies tend to focus on the novel findings rather than assess the construct or measurement issues that might affect the replication of previous findings.

In summary, our view is that the nature of these studies means that they do not play as comprehensive a role in replicating previous findings as dedicated replication studies. Moreover, there are likely systematic biases in the topics on which they focus and whether they tend to confirm or contradict existing findings.

ARTICLES IN THIS SPECIAL ISSUE

The articles in this Special Issue cover a broad spectrum of topics in strategic management. However,

the studies that are replicated do not constitute a random sample of prior work. Therefore, the findings of the research in this Special Issue cannot be generalized to draw conclusions about the ability to replicate findings in our field more generally.

In what follows, we briefly summarize each of the studies, organized by topic.

Two articles deal with inter-organizational relationships. Ranganathan, Ghosh, and Rosenkopf replicate the study by Ahuja, Polidoro, and Mitchell (2009) on the determinants of alliance formation in the chemical industry, but in a later time period. The replication takes a staged approach, first replicating the Ahuja *et al.* (2009) research design as closely as possible, then using a newer methodology (exponential random graph modeling, ERGM) not available at the time of the original study to estimate the underlying network formation process, and finally, repeating the ERGM estimation for data on the semiconductor industry in the later time period. The results for the replication using a similar research design to the original article produces similar results to those in the original study for the relationship of geographic and product-market similarity and positional embeddedness to alliance formation, but weak results for technical similarity and relational embeddedness. However, when estimated using ERGM, the results for all three types of similarity are strong, consistent with the original study, but the results for positional and relational embeddedness are not. The authors suggest that the evolution of the chemical industry, which underwent a large restructuring subsequent to the time period of the original study, may account for the differences in the results of the two studies. The extension to the semiconductor industry shows some differences between that industry and chemicals, which the authors suggest may reflect differing characteristics of the two industries.

Howard, Withers, Carnes, and Hillman replicate Beckman, Haunschild, and Phillips (2004), whose highly cited study explored factors that affect the choice of alliance and interlock partners. This article is based on a partitioning of uncertainty into firm-specific and market uncertainty. The replication both repeats the original time period of five years and then extends it to 20 years. This provides both a larger sample and the possibility of differences across time. The direct replication using the same time period found results closely matching the original study. The replication in a later time period found strikingly different results. The

authors conclude that important contextual influence on the formation of inter-organizational ties can vary across time.

Four articles investigate different aspects of geographic location, both within countries and globally. Kalnins's article replicates Baum and Mezias' (1992) study on localized competition in the Manhattan hotel industry. The replication is motivated by four issues. First, this article has impacted many different literatures, and therefore, is worthy to replicate. Second, it is an open question of whether results would hold for hotels outside of Manhattan, which is a unique market in many dimensions. The replication considers 90 urban markets in United States. Third, the original paper did not find consistent results looking at market versus submarket levels of analyses. The replication revisits both sets of analysis. Fourth, in the original paper, some of the distance measures confounded distance with the effect of market size. The replication advances a new measure. Overall, the results from the replication are consistent with (and arguably, more compelling than) the original paper. At the market level, the replication finds evidence of localized competition in terms of size, geography, and price. The original paper did not find these effects. At the submarket level, the original paper found evidence of localized competition in terms of size, geography, and price. The replication finds the effects of size and location—but location only with the new measure.

The article by Chang, Kogut, and Yang replicates an influential article by Denis, Denis, and Yost (2002), which showed a large global diversification discount for multinational firms relative to purely domestic ones. Although published in a finance journal, this article has important implications for the strategy of multinational firms because it questions the effectiveness of global diversification. The replication first constructs a similar sample of firms to that of the original article, but in a later time period using arguably more accurate data on global diversification at the firm level. The replication then uses the same methodology and variables as in the original paper, and obtains similar results to the original study. Next, the replication corrects for self-selection in global diversification choice and finds that global diversification results in a premium rather than a discount to firm performance, contrary to the results of the original study. The study also provides evidence that this premium results from the improved operating flexibility that global

diversification affords, which suggests that investing in a global network has substantial strategic benefits.

Berry and Kaul replicate Lu and Beamish's (2004) research, a widely cited study finding an S-curve empirical relationship between multinationality and performance, using a larger, comprehensive database on the population of U.S. multinational corporations from 1989 through 2007. Using the entire sample and various subsamples, they fail to find an S-curve empirical relationship. They include various robustness tests similar to those in the original paper. It seems obvious that the differences in samples are driving the differences in results. The authors caution readers about the frailty of the S-curve relationship. Importantly, they note that their results, coupled with their finding that even the association between multinationality and performance does not replicate across countries, suggest that the search for a generalizable aggregate effect of multinationality on performance is open to debate.

Tsang and Yamanoi replicate Barkema and Vermeulen's (1998) study of international expansion through start-ups or acquisition. The authors motivate their replication by showing how the results from a sample of 23 large Dutch nonfinancial companies are often interpreted as a universal effect. For this reason, they replicate the original analysis using data on the foreign investments of Singapore firms. As the first step in their replication, the authors note that two of the original hypotheses tests were mis-specified and once re-specified lead to different conclusions. When they take the original paper's four hypotheses to the Singapore data, the results are largely not generalizable (they can partially replicate one of the four hypotheses). They comment on how differences in country characteristics might lead to the failure to replicate. They also note that firm characteristics across the two samples differ and that this might be a plausible reason for the difference in findings.

The next two articles examine knowledge and product expansion through corporate venture capital and product-market diversification. Shafi and Colombo replicate the Dushnitsky and Shaver (2009) study of the limitations of interorganizational knowledge acquisition in the corporate venture capital context of the United States. The replication is done in the European context where the institutional and legal environments are different. It also examines whether timing and social

defenses encourage EU new ventures to form same industry CVC ties. Their findings indicate that the institutional setting substantially influences the relation between protection from knowledge misappropriation offered by legal defenses and the likelihood of same industry CVC tie formation. The authors suggest that the difference between the replication and the original study may arise because EU timing defenses are ineffective, and social defenses are a complement to, rather than a substitute for, legal defenses. Overall, this study is suggestive of the crucial role that may be played by boundary conditions based on different country/regional contexts.

Miller and Yang replicate Chang's (1996) study on the dynamics of diversification. The authors note that this study was the first to consider both market entry and exit over time while using a resource-based measure of related diversification. Whereas Chang's sample included only publicly-held manufacturing companies, the replication sample includes public manufacturing and nonmanufacturing firms, and in a later time period. Extensions to the analysis include additional controls introduced to the literature since the original study as well as a limited analysis of market entry and exit by private firms. The findings largely accord with Chang's original results that firms are more likely to enter industries that have similar human resource profiles to those in their existing businesses, particularly those entered more recently, and that firms are more likely to exit markets with human resource profiles that are dissimilar to those of existing businesses. The authors find some differences in the types of human resources that formed the basis for their results and Chang's, and find that private firms sometimes entered businesses with resource profiles dissimilar rather than similar to those of current businesses. The authors suggest that differences in firm characteristics in the two samples may account for the results.

Two other articles focus on IPOs (initial public offerings). The article by Chadwick, Guthrie, and Xing replicates the study by Welbourne and Cyr (1999) on the relationship between the presence of a human resource (HR) executive in the top management team and firm performance subsequent to initial public offerings (IPOs). The authors motivate the study by observing that strategic human capital and human resource management has been attracting increased attention in our field. The replication uses a larger sample of firms in a later time

period, includes data from multiple years rather than a single year as in the original study, uses more appropriate performance measures and associated control variables in addition to the original ones, and includes checks for endogeneity. The results using almost the identical model specification and very similar variable measurement to the original study fail to show a direct relationship between having an HR executive and post-IPO firm performance, as did the original study. However, the analysis using alternative performance measures related to post-IPO survival produced the opposite result to that in the original study.

The article by Park, Borah, and Kotha replicates three studies about signaling and IPO underpricing. They replicate Certo *et al.* (2001), who study how the existence of a founding manager affects IPO pricing; Filatotchev and Bishop (2002), who examine how governance indicators affect IPO pricing; and Cohen and Dean (2005), who examine how TMT legitimacy affects IPO pricing. They use a similar data source to Certo *et al.* (2001), and Cohen and Dean (2005), but in a later time period (2010–2013)—a key rationale being that IPOs after 2009 “marked a new era in terms of how firms, consumers, and investors share and gather information.” The data source differs from Filatotchev and Bishop (2002), whose sample is drawn from the United Kingdom. In addition to making efforts to match sampling criteria and measures, the replication adds controls introduced to the literature since these studies were published. The authors do not replicate the major findings of any of the three studies. They speculate that changes in regulation and information technology post-2009 affect the importance of prior signaling sources and that this might lead to the lack of replication.

The last article deals with corporate social responsibility. Zhao and Murrell replicate Waddock and Graves (1997), a highly cited study of the relationship between corporate social performance (CSP) and corporate financial performance (CFP). The original paper suggests that CSP is likely to lead to improved CFP and vice versa. The replication uses the same specification and tests, but with a much larger sample of firms (more than nine times larger) across a longer time period. The replication finds that indicators of accounting financial performance are positively related to CSP as a dependent variable, but the magnitudes of the effects are much smaller than in the original study. Furthermore, the replication does not find

a relationship between CSP as an independent variable and two of three indicators of accounting financial performance, whereas the original study found a positive relationship for two of three performance indicators. A new, different, and much larger sample across a longer time period seems to be making a difference in the results.

These articles replicate a selected sample of studies, and the findings cannot be considered representative of replications in general. Nevertheless, they span the types of results that one might expect from replication studies, even though the distribution of these types of results would likely differ for a broader set of replications. When you read the articles, you will see that some successfully replicate the results of the original studies, and others do not. Of the latter, some of these studies attribute the differences to an altered context (e.g., different time period or industry), which is hardly surprising. This does not invalidate the original results, but suggests that boundary conditions may apply to the findings. Some articles also include extensions to the original studies, such as through inclusion of additional control variables or improved methods. These extensions lead to results that sometimes accord with the results in the original studies and sometimes do not, suggesting that future research on these topics is warranted. More generally, as we noted above, regardless of whether or not a replication study confirms a prior study, a single replication is not the last word on a subject, and instead, provides a basis for future research, including additional replications.

GUIDELINES TOWARD HIGH-QUALITY REPLICATION STUDIES

Having received approximately 90 submissions to this Special Issue, we had considerable opportunity to assess what led to successful replication (i.e., studies in which we and the reviewers found the results helpful in building a cumulative body of research knowledge). We highlight the following five dimensions in which there was often a difference in the research we found to be successful replications compared to studies where there was less consensus about their success. We hope that highlighting these five dimensions helps identify what leads to a successful replication and provides a conceptual roadmap of how a dedicated replication study can be structured.

Is the focus of the replication an important finding?

Strategy is a broad field with many research traditions. As a result, not all findings are equally influential in the impact that they have had on the field. Moreover, some findings are more central in the current research discourse. In light of this, replicating research findings that are widely disseminated and central in the current research discourse weighs heavily in assessing if the replication will be beneficial to building a cumulative body of empirical research. Many of the articles in the Special Issue are explicit in justifying the importance of the result that they replicate.

Our assessment of whether or not a result was important to replicate did not revolve around whether or not there were conflicting findings of the result in the literature. We see value in replicating well-established results because they often form the foundation for further research. We also see value replicating research where the literature presents conflicting results because careful replications of such findings have the potential to make sense of the equivocal pattern of results and help advance the literature.

Is the goal a narrow replication or a quasi-replication?

As we described previously, narrow replications and quasi-replications provide distinct contributions to the literature. Because they have distinct contributions, the demands of each type of study also differ. In the case of narrow replications, it is important that the study be grounded very closely in the existing study in order to draw conclusions about the robustness of a finding within the population from where the original sample was drawn. In the case of quasi-replications, it is often important to justify the benefit of assessing if a result is generalizable to the sample that is the focus of the replication study, or if the use of improved methods is warranted. There can be more or less compelling reasons for replicating to certain populations. Building the appropriate case for a replication study requires understanding the goal of the replication.

How closely does the replication match the original study?

As we described previously, in order to draw the most informed conclusions that help build a body

of knowledge, it is important that any extension of existing findings be done in stages. Regardless of the motivation of the replication, the first stage is to as closely as possible match the research design of the existing study. This is important to calibrate the replication with the original study. If the replication cannot be calibrated, then it is difficult for the replication to build a cumulative body of empirical knowledge. Our experience with the Special Issue is that replicating the precise research methodology was often not perfectly attainable, for example due to differences in the data. Therefore, we do not see this as an absolute standard. Rather, we see the importance of judgement when making this assessment.

Is it beneficial or important to incorporate advances in the literature since the publication of the focal study?

In the quest to satisfy the first criterion that we list, we received replications of studies that were published years ago. Although this often aids justifying the replication effort, it also can lead to the complication that the related literature has advanced since the publication of the focal study. We found it most compelling when replication studies incorporated advances in the literature after replicating the original findings as closely as possible. For example, had factors omitted from the original study become important in the development of the research findings? Incorporating such advances helped us understand why the original study might or might not replicate. Moreover, we found that when carefully done, this was very useful in building a cumulative body of research. In essence, it meant that replications assessed the foundational findings and the important extensions in the literature.

We also found it beneficial when replications incorporated advances in the interpretation of results or advances in empirical methodologies. For example, studies that considered endogeneity concerns where relevant, either methodically or in their interpretation and discussion. Likewise, studies that incorporated new estimation techniques for particular types of data, such as for social networks.

Are the results interpreted so as to help build the literature?

We found replication results most compelling when they were presented in such a way that they helped

build a cumulative research understanding. In the cases where results replicated, this meant further assessing the effect size to aid in determining not only the existence of an effect, but also the magnitude of the effect.

When results did not replicate, this often meant assessing why the replication failed and ways that future studies might help ascertain why the results diverged. One element that we found important was that authors not over-conclude when results did not replicate. Just as findings from one study should not constitute a definitive result, one analysis that fails to replicate does not eradicate a finding. It is for this reason that we found it useful for the authors to try and provide insight into how the field should advance in light of the conflicting findings.

We found these five dimensions to be important in assessing the strength of the research we considered in the Special Issue. We wish to note that many of studies we accepted varied in the extent to which they were able to forcefully address all five dimensions. It is for this reason that we consider these as guidelines rather than strict requirements for high-quality replication studies in strategy.

CONCLUSION

We undertook this Special Issue driven by the belief that replication and reporting of nonresults are keys to cumulative repeatable science. Seeing the Special Issue take shape has reinforced and strengthened this belief. Many of the insights provided by the articles in this volume stem directly from their focus as dedicated replication studies, and would not be disseminated by other approaches. We encourage you to read these replication studies to see the nature of the contributions that they make and the insights that they provide on the specific questions that they revisit.

In addition, we found that engaging in writing, reviewing, or editing replication studies requires that one reflect on the underlying statistical theory of hypothesis testing—especially what a single coefficient estimate can and cannot tell us. We believe that such reflection and refinement in our thinking is important for advancing and fostering high-quality empirical research in strategic management. Therefore, replication studies not only advance our understanding of the specific results that they revisit; but also advance the rigor of scholarship in our field. This serves as a call for

all of us to be involved in replicating important results.

Transforming the original call for studies into a finished volume required that we move beyond our conceptual arguments for why we should engage in replications and address the pragmatic issues of how to actually do them. We encountered issues that we did not expect and that could not be simply resolved by reflecting on our initial motivations. Working through these issues with the authors and reviewers of the Special Issue allowed us to enhance our understanding of these issues and formulate our guidelines toward high-quality replication studies. We wish to thank the authors and reviewers for their willingness to engage in this process. We especially appreciate their willingness to be open-minded in how they approached these studies, and not narrow-mindedly approach the studies as if they were standard hypotheses-testing exercises. We could not have reached this outcome otherwise.

REFERENCES

- Ahuja G, Polidoro F, Mitchell W. 2009. Structural homophily or social asymmetry? The formation of alliances by poorly embedded firms. *Strategic Management Journal* **30**(9): 941–958.
- Barkema HG, Vermeulen F. 1998. International expansion through start-up or acquisition: a learning perspective. *Academy of Management Journal* **41**(1): 7–26.
- Baum JAC, Mezias SJ. 1992. Localized competition and organizational failure in the Manhattan hotel industry, 1898–1990. *Administrative Science Quarterly* **37**(4): 580.
- Beckman CM, Haunschild PR, Phillips DJ. 2004. Friends or strangers? Firm-specific uncertainty, market uncertainty, and network partner selection. *Organization Science* **15**(3): 259–275.
- Bettis RA. 2012. The search for asterisks: compromised statistical tests and flawed theories. *Strategic Management Journal* **33**(1): 108–113.
- Bettis RA, Ethiraj S, Gambardella A, Helfat C, Mitchell W. 2016. Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal* **37**(2): 257–261.
- Bettis RA, Gambardella A, Helfat C, Mitchell W. 2014a. Theory in strategic management. *Strategic Management Journal* **35**(10): 1411–1413.
- Bettis RA, Gambardella A, Helfat C, Mitchell W. 2014b. Qualitative empirical research in strategic management. *Strategic Management Journal* **36**(5): 637–639.
- Certo ST, Covin JG, Daily CM, Dalton DR. 2001. Wealth and the effects of founder management among IPO-stage new ventures. *Strategic Management Journal* **22**(6–7): 641–658.
- Chang SJ. 1996. An evolutionary perspective on diversification and corporate restructuring: entry, exit, and economic performance during 1981–89. *Strategic Management Journal* **17**(8): 587–611.
- Cohen BD, Dean TJ. 2005. Information asymmetry and investor valuation of IPOs: top management team legitimacy as a capital market signal. *Strategic Management Journal* **26**(7): 683–690.
- Denis DJ, Denis DK, Yost K. 2002. Global diversification, industrial diversification, and firm value. *Journal of Finance* **57**(5): 1951–1979.
- Dushnitsky G, Shaver JM. 2009. Limitations to interorganizational knowledge acquisition: the paradox of corporate venture capital. *Strategic Management Journal* **30**(10): 1045–1064.
- Filatotchev I, Bishop K. 2002. Board composition, share ownership, and “underpricing” of U.K. IPO firms. *Strategic Management Journal* **23**(10): 941–955.
- Goldfarb B, King AA. 2016. Scientific apophenia in strategic management research: significance tests and mistaken inference. *Strategic Management Journal* **37**(1): 167–176.
- Hubbard R, Vetter DE, Little EL. 1998. Replication in strategic management: scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal* **19**(3): 243–254.
- Ioannidis JPA. 2005. Why most published research findings are false. *PLoS Medicine* **2**(8): 696–701.
- Kurke LB, Aldrich HE. 1983. Note—Mintzberg was right!: A replication and extension of the nature of managerial work. *Management Science* **29**(8): 975–984.
- Lu JW, Beamish PW. 2004. International diversification and firm performance: the S-curve hypothesis. *Academy of Management Journal* **47**(4): 598–609.
- Open Science Collaboration. 2015. Estimating the reproducibility of psychological science. *Science* **349**(6251): aac4716.
- Shaver JM, Flyer F. 2000. Agglomeration economies, firm heterogeneity, and foreign direct investment in the United States. *Strategic Management Journal* **21**(12): 1175–1193.
- Singh K, Ang SH, Leong SM. 2003. Increasing replication for knowledge accumulation in strategy research. *Journal of Management* **29**(4): 533–549.
- Tsang EWK, Kwan K-M. 1999. Replication and theory development in organizational science: a critical realist perspective. *Academy of Management Review* **24**(4): 759–780.
- Waddock SA, Graves SB. 1997. The corporate social performance-financial performance link. *Strategic Management Journal* **18**(4): 303–319.
- Welbourne TM, Cyr LA. 1999. The human resource executive effect in initial public offering firms. *Academy of Management Journal* **42**(6): 616–629.