

IS THERE A CREDIBILITY CRISIS IN STRATEGIC MANAGEMENT RESEARCH?

EVIDENCE ON THE REPRODUCIBILITY OF STUDY FINDINGS

DONALD D. BERGH
Daniels College of Business
The University of Denver
2101 S. University Blvd.
Denver, CO 80208
dbergh@du.edu

BARTON M. SHARP
Department of Management
Northern Illinois University
Barsema Hall
DeKalb, IL 60115
bsharp1@niu.edu

HERMAN AGUINIS
Department of Management
School of Business
George Washington University
2201 G. Street, NW
Washington, DC 20052
haguinis@gwu.edu

MING LI
University of Liverpool Management School
Chatham Street, Liverpool L69 7ZH, United Kingdom
lilyliming@hotmail.com

Forthcoming, *Strategic Organization*, March 2017

IS THERE A CREDIBILITY CRISIS IN STRATEGIC MANAGEMENT RESEARCH?

EVIDENCE ON THE REPRODUCIBILITY OF STUDY FINDINGS

Recent studies report an inability to replicate previously published research, leading some to suggest that scientific knowledge is facing a credibility crisis. In this essay, we provide evidence on whether strategic management research may itself be vulnerable to these concerns. We conducted a study whereby we attempted to reproduce the empirical findings of 88 articles appearing in the *Strategic Management Journal* using data reported in the articles themselves. About 70 percent of the studies did not disclose enough data to permit independent tests of reproducibility of their findings. Of those that could be retested, almost one-third reported hypotheses as statistically significant which were no longer so and far more significant results were found to be non-significant in the reproductions than in the opposite direction. Collectively, incomplete reporting practices, disclosure errors, and possible opportunism limit the reproducibility of most studies. Until disclosure standards and requirements change to include more complete reporting and facilitate tests of reproducibility, the strategic management field appears vulnerable to a credibility crisis.

Keywords, Reproducibility, replication, knowledge credibility

IS THERE A CREDIBILITY CRISIS IN STRATEGIC MANAGEMENT RESEARCH?

EVIDENCE ON THE REPRODUCIBILITY OF STUDY FINDINGS

Currently, there are important concerns about the state of study findings in literatures as diverse as psychology, economics and biomedicine (e.g., Bakker and Wicherts, 2011; Bosco, Aguinis, Field, Pierce, and Dalton, 2016; Chang and Li, 2015). For example, scholars report an inability to replicate the findings of seminal studies (e.g., Begley and Ellis, 2012; Ioannidis, 2012), leading some to suggest a possible credibility crisis in science (e.g., Baker, 2012; Carpenter, 2012; Cortina, Aguinis, and DeShon, in press).

Do these concerns extend to strategic management research? In this essay, we consider whether empirical findings in strategic management research can be reproduced using their own data. Reproducibility is “the ability of other researchers to obtain the same results when they reanalyze the same data” (Kepes, Bennett and McDaniel, 2014, p. 456). If study findings cannot be reproduced from their own data, questions arise about their credibility. Can the strategic management field’s findings be reproduced? If not, why, and if so, how closely do reproduced results compare to those of the original research? In addition, what is the scope of reproducibility—can most empirical studies be reproduced, half, or is it only a few? Are differences between reported and reproduced findings random or do they suggest a systematic bias on the part of authors? Although a recent study examines whether published findings are correctly reported in terms of their coefficient statistical properties (e.g., Goldfarb and King, 2016), we know of no studies that address the congruence of statistical findings and the data upon which they are based in strategic management research. Currently, we lack evidence for estimating the reproducibility of findings and if the field is vulnerable to a credibility crisis.

Our essay reports an initial examination of the reproducibility of findings appearing in

strategic management research. Within this field, we examine articles published in the *Strategic Management Journal (SMJ)*, as this journal offers an unambiguous source of research directly addressing strategic management domains and disseminates the most articles devoted to strategy research, and has been used in other reviews as a single source for evaluating precedent-setting methodological practices in strategic management studies (e.g., Aguinis, Edwards, and Bradley, in press; Bergh and Fairbank, 2002; Ferguson and Ketchen, 1999; Shook, Ketchen, Cychota and Crockett, 2003). We do not isolate the *SMJ* for criticism; rather, we examine some of its articles because it has long been considered a leading strategic management research outlet and its practices are likely to be adopted by others.

Using data from two samples of *SMJ* articles, we attempt to reproduce empirical models and compare our results with those originally reported. These comparisons provide insights into the proportion of strategic management findings that can be reproduced, whether supported hypotheses remained so, and if there is cause for concern about the state of reproducibility – and by extension the credibility - of empirical findings in strategic management research.

Overall, our study examines some possible drivers behind a credibility crisis by documenting the ability to reproduce empirical findings and identifying barriers to reproduction. It illuminates the possible effects of the non-reproducibility of findings on the field's empirical foundation and illustrates how reproducibility (or lack thereof) affects the interpretability and meaning of replication and extension studies. Viewed generally, these results offer insight into whether conditions exist that might create a credibility crisis within the strategic management field.

METHOD

Reproducing a study's findings generally implies re-analyzing original datasets and

comparing reproduced findings with those initially reported in the study. Such a process would seem to require either the data from a focal article or an independent re-collection of the data. However, an alternative approach exists whereby a study's variable means, standard deviations (SDs), correlations, and sample sizes can serve as substitutes for the original data set, as these descriptive and correlational data are all that is necessary for many analytical tests including analysis of variance (ANOVA), linear regression, *t*-tests, discriminant analysis, and structural equation modeling (see Bergh, Sharp and Li, in press; Shaver, 2005; Boyd, Bergh and Ketchen, 2010, for illustrations). Indeed, most statistical packages, such as Stata, IBM SPSS, SAS, and R (among others), offer the capability to analyze descriptive and correlational data in lieu of raw data, and the findings from testing either form of data will be identical. For example, the manual for Stata's 'corr2data' procedure reports that it...

“allows you to perform analyses from summary statistics (correlations/covariances, means) when these summary statistics are all you know and summary statistics are sufficient to obtain results...the analysis... extracts from the summary statistics you specified, and then makes its calculation based on those statistics...**the results from the regression based on the generated [summary] data are the same as those based on the real data.**” (<http://www.stata.com/manuals13/dcorr2data.pdf>, emphasis added)

Other analytical packages offer a “Matrix” feature which provides users with selecting between a matrix of descriptive summary and correlational data or the raw data file as the data input when conducting their analyses (demonstrated in the online Appendix with IBM SPSS).

Samples

To test whether empirical findings in strategic management research can be reproduced from their own data, we drew samples of articles appearing in *SMJ* that employed linear

regression (LR) and structural equation modeling (SEM) analyses—the two most popular techniques in the strategy field (Shook et al., 2003). Both analytical techniques can be implemented using descriptive and correlational statistics and represent a range of analytical sophistication that gives insights into whether reproducing results is more possible in the most accessible basic tests (LR), the more advanced ones (SEM), or equally in both.

We assembled two samples of *SMJ* studies using different criteria so our findings might apply to a diverse set of articles. One consisted of *SMJ* articles using the LR technique and the other consisted of the top ten most cited *SMJ* articles that employed the SEM technique, based on citations in 2016 from the ISI Web of Science. The LR study sample represents a typical body of *SMJ* articles—a mix of some having high impact and others having normal and lower impact. The SEM studies likely played visible roles in shaping the development of the field.

LR studies. We identified the sample of *SMJ* articles reporting findings from LR tests using the following steps. (1) We searched for all *SMJ* articles that used the terms “OLS” or “ordinary least squares” using the full-text search function at Wiley Online (the *SMJ* publisher) through the *SMJ* website (<http://smj.strategicmanagement.net>). We restricted our search terms to avoid judgment calls and ensure conditions for meaningful replication of our own work. Also, we focused on OLS regression, as opposed to logistic, time series, or other forms, because OLS models are the most basic of regression analytical tests and involve little interpretation on our part, thereby minimizing the possibility that we might misunderstand the authors’ procedures. (2) We defined our sampling frame to articles published in two time periods, 2000–2004 and 2010–2013 to determine if the reproducibility of findings has changed over time as well as include articles likely to reflect more current methodological practices. (3) We retained only those articles that reported OLS regression models and results. This three-step process identified 79

articles (50 from 2000–2004 and 29 from 2010–2013). (4) We then examined each to identify the correlations, means, standard deviations, and sample sizes. The 79 articles are identified in an accompanying online file.

When reported, the descriptive and correlational statistics were arranged into matrices that were entered as input data into the regression analyses. We then retested each LR study using the `corr2data` procedure in accordance with the instructions provided within the Stata manual for testing data reported in a matrix format (the program code, illustrative data from one of the 79 *SMJ* LR studies, the reported findings, and the results produced by the Stata analyses are also reported in the supplemental online file). In the cases where discrepancies existed between reported and reproduced findings, we double-checked the inputted data and re-ran our tests using an alternative analytical software package, IBM SPSS (illustrative syntax is included in a supplemental online file). This multi-test approach served as a reproduction test of our own data and findings and reduced the likelihood of error on our part.

SEM studies. We identified the sample of *SMJ* articles reporting SEM using the following steps. (1) We searched for all articles that used the terms “structural equation modeling,” “structural equations,” “lisrel,” “amos,” or “path analysis” using the *SMJ* website (<http://smj.strategicmanagement.net>) and its search function. (2) We retained only those that used SEM analyses. (3) We collected citation counts for each article from the ISI Web of Science, ranked them by total counts, and then selected the top 10 (note that one of these 10 articles was also in the sample of 79 OLS articles, leading to a final sample of 88 total articles). We focused on the most highly cited articles because they would have had the largest impact on subsequent research. In addition, this sample provides a different context for comparing reproducibility of study findings since it includes the most influential articles rather than ordinary articles

employing the LR study technique. (4) Finally, we examined each identified article to locate the correlations, means, standard deviations, and sample sizes when reported. The articles are listed in the supplemental online file.

We conducted the SEM analyses using the IBM SPSS Amos statistical analysis software. The summary data were configured into a matrix, entered into the program, and then tested using the same assumptions and parameter specifications as reported by authors (an example appears in the supplemental online file). Differences between reported findings and our analyses were retested through repeating each of the analyses 10 times (explained below).

RESULTS

Reproducibility of Statistical Significance Conclusions

Fifty eight of the 79 *SMJ* studies using LR (73 percent) and four of the 10 employing SEM (40 percent) did not report sufficient information to permit any reproducibility analysis whatsoever. The 58 LR studies that could not be retested (36 from 2000–2004 and 22 from 2010–2013) had incomplete disclosure of means, SDs, and correlations for some study variables, missing cells within the correlation matrices, statistical analyses using disaggregated subgroups, transformed variables for which summary statistics were not reported, or correlation matrices which were not positive semi-definite (please see Table 1 for the full list of reasons).¹ These barriers to reproduction are relatively equal across both time periods suggesting data reporting practices have consistently impeded reproduction in a large majority of LR studies and nearly half of those using SEM analyses.

[Insert Table 1 about here]

Of the 21 LR studies that could be retested, 16 could be reproduced partially (some models could be reproduced but others could not) and 5 fully (all models could be reproduced).

Four of the 10 SEM studies could not be re-analyzed due to one missing a correlation matrix, another failing to report means and standard deviations for some study variables, one that did not specify the precise model that was tested, and one that did not include the coefficients of the final model. Six provided sufficient descriptive summary data to permit re-analysis. Collectively, we were able to retest 20 *SMJ* studies employing only LR, 5 that used only SEM, and 1 study that reported both techniques. Overall, we could either partially or fully retest the results reported in only 26 of 88 articles (29.5 percent).

Following previous tests of the congruence between reported and reproduced findings in other scientific fields (e.g., Bakker and Wicherts, 2011; Berle and Starcevic, 2007; Garcia-Bethou and Alcaraz, 2004), we focused on the signs of the coefficients (indicating either positive or negative relations with the dependent variable) rather than their actual magnitude to avoid errors on our part. For example, some authors report coefficients in standardized formats, others report them as non-standardized, and it is not always clear which approach was used. For example, some articles include a table with regression coefficients that excludes the intercept, giving the impression that coefficients are standardized—because in those cases the value for the intercept is zero (Aguinis, 2004). But, upon reading the article it becomes clear that coefficients are actually unstandardized and the table simply excluded information regarding the intercept. Further, the directionality of the coefficients and their significance levels always play a role in hypothesis testing whereas the magnitude rarely if ever does.

The 21 articles using LR reported a total of 732 coefficients in the models which could be reproduced. The reproduced findings corroborated 670 of the 732 (91.5 percent) coefficient directional signs (either both coefficients were the same sign, or one of them was zero). The six

studies employing SEM reported 10 models that collectively included 91 coefficients, of which 86 directional signs were reproduced (95 percent).

Next, we compared reported with reproduced statistical significance threshold levels. Our retests focused on significance bands ($p < 0.01$; $p < 0.05$), rather than precise point estimates for the p values because most prior *SMJ* articles indicate statistical significance using the star notation (e.g., ** for $p < 0.01$; * for $p < 0.05$). The re-analyses of the 21 LR studies resulted in the reproduction of the precise reported significance band for 538 of the 732 coefficients (73.5 percent). The re-analyses of the 6 SEM studies resulted in 79 of the 91 coefficients' significance levels (87 percent) within the same bands.

Reproducibility of Results and Conclusions Regarding Hypothesized Relations

Hypotheses serve as the basis for article conclusions and are critical in terms of a study's "take-away" message, including implications for theory and practice. If the findings pertaining to the hypotheses cannot be reproduced, then causes for concern regarding the credibility of findings and recommendations would exist. Accordingly, we conducted additional analyses focused on the variables corresponding to hypotheses.

We first identified the coefficients in the 21 re-analyzable *SMJ* studies using LR that were associated with hypothesis testing. We found 144 coefficients representing 51 variables representing hypotheses. Of those 144 coefficients, 14 associated with 10 variables were reported as statistically significant in the original *SMJ* article, but were reproduced as statistically non-significant (14 of 144 coefficients, or 10 percent; 10 of 51 variables, or 20 percent). This change in statistical support occurred in six of the 21 articles (28 percent of articles). We also found that three coefficients were reported as statistically non-significant in the original *SMJ* article but

reproduced as having significance at the $p < 0.05$ level or higher (3 of 144 coefficients, or 2 percent).

We then considered a possible alternative explanation for these findings, namely that a lack of reproducibility was due to rounding. In other words, authors who found an exact p -value of 0.0499 for a given coefficient might report that finding with a single star indicating $p < 0.05$. In addition, it possible that rounding in the reported descriptive statistics might lead our reproduced findings to vary slightly from the original regressions, possibly leading to an exact p value of 0.051 which we would have to categorize in the $p < 0.10$ band. Thus, we tested for the possibility that reported and reproduced p values may have been concentrated very near the $p < 0.05$ threshold leading us to perceive a lack of reproducibility when in fact no meaningful difference existed.

The results are reported in Table 2 for the hypotheses in the six *SMJ* studies using LR that lost statistical support. Results from independent Stata and IBM SPSS analyses are also identical, suggesting that there is no effect on results due to the choice of software package. The table shows that only one of the 14 coefficients was close to the threshold; in Study #5, the reported level was $p < 0.05$ whereas the reproduced value was $p = 0.053$. None of the other reproduced coefficients was even border-line to conventional significance levels. We also note that four of the six *SMJ* studies using LR each have one result that was not reproduced, one of the six has two results that could not be reproduced, and one study has eight reportedly significant findings that we could not reproduce.

Overall, the findings indicate that in six studies hypotheses which were supported in the publication lost empirical support in our retests. Based on our re-analysis, some of these studies' substantive conclusions would not have received supporting evidence.

[Insert Table 2 about here]

We next identified the coefficients in the six SEM studies that were associated with hypotheses. Hypotheses in two studies (33 percent) could not be reproduced, one study lost support for one of 14 supported hypotheses while another lost support for 11 of 19. In addition, we examined whether the differences in significance levels between reported and reproduced are due to rounding error. The findings reported in Table 3 suggest that such an alternative explanation is unlikely.² Further, all of the significant-turned-non-significant findings pertained to hypotheses. Thus, 22 percent of the statistical significance of the hypothesis testing coefficients (12 of 55) in the *SMJ* studies using SEM were not confirmed and no cases existed where the findings for hypotheses were reported as non-significant but found to be significant in the re-analysis.

[Insert Table 2 about here]

Averaging across the LR and SEM samples, about 30 percent re-analyzable studies (8 of 27 total) reported significant hypotheses that lost statistical support in the reproduction tests. By contrast, less than five percent have hypotheses reported as statistically non-significant but meet conventional statistical levels in the retests.

DISCUSSION

A high profile debate on the replicability of study findings is emerging across multiple disciplines (e.g., Baker, 2012; Bissell, 2013; Bosco et al., 2016; Carpenter, 2012) raising concerns about whether science is facing a credibility crisis (Butler, Delaney, and Spoelstra, in press; Schwab and Starbuck, in press). Our essay considers this topic within strategic management research. We sought to document the reproducibility of study findings, identify barriers that impede reproducibility, and when possible, compare whether reported and

reproduced findings were the same. Drawing from two samples of articles appearing in the *SMJ*, we were unable to conduct reproducibility analysis for more than 70 percent of studies employing linear regression (consistent over the two time periods) and for four of the top ten cited articles using structural equation modeling. Furthermore, of the studies that could be reproduced, nearly one of three reported hypotheses as statistically significant which were no longer so in re-testing, and far more significant results were found to be non-significant in the reproductions than in the opposite direction. In some cases, multiple hypotheses within a single article lost support even though most of the corresponding coefficient directional signs were reproduced. Primary conclusions in those articles were based on statistical results that could not be reproduced. These findings exist for articles having low as well as high impact.

Overall, based on our sample of 88 *SMJ* articles, the strategic management literature appears vulnerable to credibility problems for two main reasons. One, the majority of the articles did not report their data sufficiently to permit reproduction, leaving us in the dark with regards to the accuracy of their reported results. Two, among those articles where reproduction analyses were possible, a significant number of discrepancies existed between reported and reproduced significance levels.

Implications for Substantive Conclusions and Research Evaluation Practices

The study findings suggest some initial insights into whether findings in strategic management empirical research can be confirmed from their own reported data. For the most part, we simply cannot tell, although some indications point to “no.” In those cases where we were able to re-run the authors’ analytical models, we found a troubling proportion of discrepancies between the reported and reproduced results; in some cases those discrepancies could alter our conclusions about hypothesized relations and their underlying theories. In the

majority of cases using LR and nearly the majority of those using SEM, reproduction was simply not possible, rendering us unable to confirm findings and conclusions. At best, this lack of reproducibility represents deficient reporting practices. At worst, it means that we as researchers are attempting to build on results that do not accurately and fully represent the underlying data, and that we as teachers are passing on to practitioners conclusions that may not be sound reflections of the true underlying phenomena. To bolster the value of and confidence in empirical research, we call for the field to recognize the role of reproducibility in the scientific process.

In addition, our findings suggest that replication studies may have challenges: if we cannot re-test 70 percent or more of the studies' findings, and some 30 percent of the articles that could be re-tested contained significant coefficients used to test hypotheses that could not be confirmed, then we have a potentially perilous basis for offering conclusions about the meaning of findings from replications. Indeed, authors of replications that have different conclusions might attempt to attribute them to the generalizability of the focal research while all along the reasons for the discrepancies could be unknowable undisclosed data decisions, errors, or possibly even malfeasance. A case in point is an article in the 2016 *SMJ* Special Issue on Replication in Strategic Management by Park, Borah and Kotha. These authors attempt to replicate three articles on signaling theory, finding no support for original results, concluding that the reasons for the differences in their replicated results included sampling periods, measures, geographical context, extraneous factors and omitted variables (2016: 2374-2375).

However, we posit that replications need to first test whether the focal study is reproducible from its own data. If reproducibility is unsuccessful or not possible, the ability to draw conclusions from a replication could be compromised, as any differences between the findings of an original study and a replication could be due to unobservable issues. For example,

the authors of the original study may not have disclosed their decisions about outliers (Aguinis, Gottfredson, and Joo, 2013), cherry-picked their findings from a larger set of models (Bosco et al., 2016), tweaked and altered the analyses (Banks et al., 2016), or state that they test one type of effect such as full mediation (Holland, Shore, and Cortina, in press) whereas in actuality they do not. In addition, some have suggested that some articles experience a metamorphosis during the review process whereby authors may engage in post hoc alterations of hypotheses and data as well as questionable research practices (Bosco et al., in press; Butler et al., in press; O'Boyle, Banks and Gonzales-Mule, 2014). If such decisions are employed, then descriptive and correlational data may not reflect the data that are ultimately used for testing the models. A reproduction test would uncover these incongruencies while a replication would not. In the case of a replication study, the author would likely produce a study that would mirror the decisions that produced the reported descriptive and correlational data of the focal study although might be unable to replicate the findings due to the unobservable and unreported actions that led to the final results. A reproduction test could detect this possible problem before the replication was even attempted.

Thus, it is critical that authors of replications first reproduce the focal study's findings from its own data. If the reproduction yields findings that differ from those reported in the original study's, replication researchers are faced with the conundrum of deciphering whether differences in a replication are due to context or because the results in the original study do not reflect its own underlying data. Thus, we recommend that all replications first reproduce focal study findings. If replication studies do not reproduce focal study findings, and they are unable to replicate their results, we may not know why the observed differences exist.

Thus the reproducibility of a focal study's findings is a vital and essential prelude to

meeting the conditions for repeatable cumulative knowledge development. Until a study's findings are reproduced, they cannot be assumed as a trustworthy benchmark for replication, comparison, and extension. In those instances where reproduction is not possible, we recommend that authors adopt a cautious interpretation of differences in their results, emphasizing more on what they know (their study decisions), less on what they do not know (the decisions in the focal study), and calling for additional research to reconcile and understand differences in conclusions. We propose that authors employ a 'verify then trust' approach before attempting any type of replication.

The findings also have implications for the field's peer review process. We posit that the presence of non-reproducible results limits the confirmation of received results and that the peer review process needs to remove the obstacles that stand in the way of reproduction tests. We suggest the following remedies would increase the reproducibility of findings: (1) reporting significance levels using precise p values rather than cutoffs such as 0.01 and 0.05; (2) disclosing all methodological decisions that have a bearing on the reported findings including the handling of missing data and outliers (perhaps in a supplemental online file); (3) reporting all descriptive, correlational and corresponding sample size data for *all* variables in each corresponding test, including controls, dependent, independent, mediators and moderators, and transformed variables; (4) reporting verification of linear models as a preliminary stage in all models, including those that also test more complex structures; (5) including figures that specify all of the variables (indicators, latent factors, controls, and error terms); (6) in the case of path models, indicating which covariances among exogenous variables were allowed to vary (alternatively, a footnote to the figure should clarify which covariances were allowed to vary, and which were fixed to zero); and (7) reproducibility tests, or at least evidence that reproducibility can be

achieved through the reported data. Such disclosures would close a disclosure loophole in current peer review processes, permit independent reproduction, allow for identification of reporting mistakes, discourage some forms of questionable research practices and scientific misconduct, strengthen the field's empirical literatures and contribute to replication and meta-analytical research (e.g., Aytug et al., 2012). Further, these recommendations go beyond recent changes in submission requirements at journals like the *SMJ* (Bettis et al., 2016), *Journal of Management* (Wright, 2016), and others, but would more likely produce the kind of disclosures that these editors expect for their published articles.

Potential Limitations and Suggestions for Future Research

Our analyses and results need to be considered relative to our study's limitations. First, we retested studies using linear modeling techniques which, although the most popular, do not represent the universe of approaches used in strategic management research. We cannot make any conclusions regarding the reproducibility of empirical findings of studies that did not use LR or SEM. However, we have no reason to expect that those other approaches would be any more reproducible than the studies that we examined. Indeed, most members of the strategic management research community are trained in the general linear model, so if problems are found in the application of this most basic model, then it would seem likely that problems could exist elsewhere. Thus, if we cannot reproduce most studies using techniques that are diffused widely, then studies employing more complex approaches may be even less amenable to reproduction and our study's results may actually understate the verification problems within the management field.

Two, our findings could be influenced by author reporting. Authors are required to review their findings before their articles are published, check proofs, are held responsible for

their study's disclosures, and generally make few reporting errors (Bakker and Wicherts, 2011). But, to the extent that authors make mistakes or behave opportunistically with their reporting, then reproduction findings could be affected by such decisions (Schwab and Starbuck, in press). However, given that (a) most coefficient signs were reproduced and (b) the discrepancies in the re-tested results greatly favored the support of hypotheses, then the presence of errors may be less likely than the "chase of statistical significance...and the strong tendency to obtain confirmed hypotheses" (Kepes and McDaniel, 2013, p. 254). So author error alone may not be driving the observed differences in our findings. Although intentional opportunism is a potential cause, it would be premature to make such a claim based on our results. Such evidence instead underscores the need for reproduction in the review process and further research into the articles whose results could not be confirmed using their own reported data.

Finally, we examined articles appearing in one journal only, the *SMJ*, a high-quality outlet devoted to the field of strategic management studies, which has among the highest review standards. However, like any study, our findings may not apply to other journals. We hope that future research will extend our reproducibility approach to an examination of other journals in strategic management research and other management subfields such as organizational behavior, entrepreneurship, human resource management, and international business. Such extensions will provide insights into the scope and scale of the reproducibility challenges in empirical research.

Concluding Remarks

We suggest that limitations in studies' reproducibility present a threat to the credibility of some study findings within the strategic management literature. If we cannot reproduce a study using its own data, then can we have confidence in its findings?

Our findings represent a first assessment of the congruence between findings and the data upon which they are based and more research into the scope and scale of reproducibility is needed to fully comprehend the size of the matter and its implications. We therefore call for extensions of our research methodology to include seminal studies that shaped the field, studies published in different time periods, those using other analytical approaches (the data matrix approach can be used to test other forms of regression and linear and non-linear models such as factor analysis), and broadening the assessments to different literatures beyond strategic management. A comprehensive assessment of reproducibility is needed to fully comprehend its pervasiveness and effects.

Overall, the credibility of the strategic management field's body of knowledge seems at risk until disclosure and peer review requirements are changed to increase the reproducibility of all empirical studies. We hope that our study's findings motivate our fellow researchers to further assess empirical work, to educate all scholars about basic reporting requirements for enhancing reproducibility, and that the field's gatekeepers revise disclosure requirements and include reproducibility in the review process going forward.

Endnotes

¹ When a matrix is not positive semi-definite, then a mismatch likely exists within the relations and some of the values (i.e., bivariate correlations) are likely to be out of range. For example, if variables A and B are positively and highly correlated, and B and C are also highly and positively correlated, then the correlation between A and C must be high and positively as well (Aguinis and Whitehead, 1997). But, for example, if the correlation between A and C is zero, then the statistical properties of the matrix are not positive semi-definite. Possible sources of a non-positive semi-definite matrix are mistakes in the published correlation table, rounding errors that create the appearance of negative variance, and pairwise correlations within the same matrix that have different sample sizes.

² We retested all SEM studies where differences existed between reported and reproduced conclusions as undisclosed reporting decisions with respect to estimator approaches, correlations between error terms, and model fitting procedures might lead to differences in results (Landis, Edwards, and Cortina, 2009). One possible source for such differences is that the maximum likelihood estimator is converging on a local maximum which might have occurred in either the *SMJ* article or in our re-analysis, an outcome that we attempted to address through conducting multiple analyses. Similar to Shaver (2005), we used maximum likelihood estimation (MLE). MLE attempts to maximize the likelihood function—the function that represents the likelihood of the data that are observed. However, depending on the initial parameters chosen, the algorithm might stop and return premature estimates, which are called ‘local maxima.’ There are no known solutions for this problem, but one way to address it is to use a range of randomly generated initial parameters (Myung, 2003), which is the procedure implemented by IBM SPSS Amos, LISREL, and many other programs typically used by strategic management researchers. Nevertheless, even if random parameters are used, there may be a difference between results reported in published articles and those reproduced in our study, which may be a reason for the observed discrepancies between results. To assess the possibility that local maxima may have occurred in our reproductions, we re-ran each of our analyses 10 times. As expected, and because the initial parameter values are random, substantive results remained unchanged. So, there is a possibility that initial parameter estimates in the published studies were not random, which may have led to local maxima, and authors did not disclose this information. The failure to report such information inhibits verification and replication of the findings.

References

- Aguinis H, Boyd BK, Pierce CA and Short JC (2011) Walking new avenues in management research methods and theories, Bridging micro and macro domains. *Journal of Management* 37: 395-403.
- Aguinis H, Gottfredson RK and Joo H (2013) Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods* 16: 270-301.
- Aguinis H and Whitehead R (1997) Sampling variance in the correlation coefficient under indirect range restriction, implications for validity generalization. *Journal of Applied Psychology* 82: 528–538.
- Baker R (2012) Independent labs to verify high-profile articles. *Nature*. DOI: 10.1038.2012.1/176.
- Bakker M and Wicherts J (2011) The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods* 43: 666–678.
- Barrett LF (2015) Psychology is not in crisis. *New York Times*. Available at: <http://www.nytimes.com/2015/09/01/opinion/psychology-is-not-in-crisis.html> (accessed 1 September 2015).
- Begley CG and Ellis LM (2012) Drug development, raise standards for pre-clinical cancer research. *Nature* 483: 531–532.
- Bergh DD and Fairbank JF (2002) Measuring and testing change in strategic management research. *Strategic Management Journal* 23: 359-366.
- Bergh DD, Sharp B, and Li, M. (in press) Tests for identifying “Red Flags” in empirical findings: Demonstration and recommendations for authors, reviewers, and editors. *Academy of Management Learning and Education* doi.org/10.5465/amle2015.0406.
- Berle D and Starcevic V (2007) Inconsistencies between reported test statistics and *p*-values in two psychiatry journals. *International Journal of Methods in Psychiatric Research* 16: 202-207.
- Bettis RA, Ethiraj S, Gambardella A, Helfat C, and Mitchell W (2016). Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal* 37: 257-261.
- Bissell M (2013) Reproducibility, the risks of the replication drive. *Nature* 503: 333–334.
- Bosco FA, Aguinis H, Field JG, Pierce CA and Dalton DR (in press). HARKing’s threat to organizational research, evidence from primary and meta-analytic sources. *Personnel Psychology* doi: 10.1111/peps.1211.
- Boyd BK, Bergh DD, Ireland RD and Ketchen DJ Jr (2013) Constructs in strategic management. *Organizational Research Methods* 16: 3-14.

- Boyd BK, Bergh DD and Ketchen DJ Jr (2010) Reconsidering the reputation-performance relationship, a resource-based view. *Journal of Management* 36: 588–609.
- Boyd BK, Ketchen DJ Jr and Bergh DD (2008) Research methodology in strategic management. Past accomplishments and future challenges. *Organizational Research Methods* 11: 643-658.
- Butler N, Delaney H and Spoelstra S (in press). The grey zone: Questionable research practices in the business school. *Academy of Management Learning and Education* doi: 10.5465/amle2015.0201.
- Carey B (2015) Psychology's fears confirmed: Rechecked studies do not hold up. *New York Times*, 28 August, 2015.
- Carpenter S (2012) Psychology's bold initiative. *Science* 335: 1558-1561.
- Chang AC and Li P (2015) Is Economics Research Replicable? Sixty Published Articles from Thirteen Journals Say "Usually Not", *Finance and Economics Discussion Series 2015-083*. Washington: Board of Governors of the Federal Reserve System, <http://dx.doi.org/10.17016/FEDS.2015.083>. Available at <https://www.federalreserve.gov/econresdata/feds/2015/files/2015083pap.pdf>
- Cortina JM, Aguinis H and DeShon RP (2016) *Twilight of dawn or of evening? A century of research methods in the Journal of Applied Psychology*. Working Paper. Available at <http://kelley.iu.edu/facultyglobal/directory/FacultyProfile.cfm?netID=haguinis>
- Ferguson TD and Ketchen DJ Jr (1999) Organizational configuration and performance: The role of statistical power in extant research. *Strategic Management Journal* 20: 385-396.
- Garcia-Berthou E and Alcaraz C (2004) Incongruence between test statistics and *p* values in medical articles. *BMC Medical Research Methodology* 4: 13.
- Goldfarb BD and King AA (2016) Scientific apophenia in strategic management research: significance tests & mistaken inference. *Strategic Management Journal* 37: 167-176.
- Ioannidis JPA (2012) Why science is not necessarily self-correcting. *Perspectives on Psychological Science* 7: 645–654.
- Kepes S, Bennett A and McDaniel M (2014) Evidence-based management and the trustworthiness of cumulative scientific knowledge, Implications for teaching, research and practice. *Academy of Management Learning and Education* 13: 446-466.
- Kepes S and McDaniel MA (2013) How trustworthy is the scientific literature in industrial and organizational psychology? *Industrial and Organizational Psychology* 6: 252–268.
- Koole SL and Lakens D (2012) Rewarding replications, a sure and simple way to improve psychological science. *Perspectives on Psychological Science* 7: 608–614.

Makel MC, Plucker JA and Hegarty B (2012) Replications in psychology research, how often do they really occur? *Perspectives on Psychological Science* 7: 537–542.

Molina-Azorin JF (2012) Mixed methods research in strategic management: Impact and applications. *Organizational Research Methods* 15: 33-56.

Myung IJ (2003) Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology* 47: 90–100.

O’Boyle EH, Banks GC and Gonzalez-Mulé E (2014) The Chrysalis effect, How ugly initial results metamorphosize into beautiful articles. *Journal of Management* 43: 376-399.

Park UD, Borah A and Kotha S. 2016. Signaling revisited: The use of signals in the market for IPOs. *Strategic Management Journal* 37: 2362-2377.

Schwab A and Starbuck W. (in press). A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning and Education* doi: 10.5465/amle.2016.0039

Shaver JM (2005) Testing for mediating variables in management research, concerns, implications, and alternative strategies. *Journal of Management* 31: 330–353.

Shook CL, Ketchen DJ, Cycyota CS and Crockett D (2003) Data analytic trends and training in strategic management. *Strategic Management Journal* 24: 1231–1237.

Wright PM (2016). Ensuring research integrity: An editor’s perspective. *Journal of Management* 42: 1037-1043.

Wright PM, Coff R and Moliterno TP (2014) Strategic human capital, Crossing the great divide. *Journal of Management* 40: 353-370.

Table 1

Results of Reproducibility Study Findings that were based on Ordinary Least Squares Regression

	Number of articles (% of sample)	
	2000–2004	2011–2013
Reason regression results could not be reproduced		
Descriptive statistics not reported	10 (20%)	5 (17%)
Descriptive statistics missing for key variables	7 (14%)	1 (3%)
Industry, firm, or time dummies not reported in descriptive statistics or regression results	5 (10%)	13 (45%)
Descriptive statistics given for raw variables, transformed variables used in regressions	5 (10%)	2 (7%)
Reported tables of correlations not positive semi-definite	4 (8%)	1 (3%)
Descriptive statistics given for full sample, regressions conducted on sub-samples	3 (6%)	
Other	2 (4%)	
Total articles for which reproducibility analysis was not possible	36 (72%)	22 (76%)
Problems in reproduced studies that limited complete retesting		
Descriptive statistics not reported for interaction terms	5 (10%)	5 (17%)
Descriptive statistics missing for other key variables	3 (6%)	1 (3%)
Descriptive statistics given for raw variables, transformed variables used in regressions	1 (2%)	
Reported tables of correlations not positive semi-definite	1 (2%)	
Total articles for which some but not all models could be reproduced	10 (20%)	6 (21%)

Table 2

Reproducibility of Ordinary Linear Regression Hypothesis Findings:
Reported and Reproduced Statistical Significance Levels

Study Identifier	Reported p value	Reproduced p value by Stata	Reproduced p value by IBM SPSS
1	<0.01	0.087	0.087
2	<0.05	0.704	0.704
3	<0.01	0.244	0.244
4	<0.01	0.109	0.109
4	<0.001	0.179	0.179
4	<0.05	0.241	0.241
4	<0.05	0.386	0.386
4	<0.001	0.172	0.172
4	<0.01	0.093	0.093
4	<0.001	0.115	0.115
4	<0.05	0.909	0.909
5	<0.05	0.053	0.053
6	<0.05	0.174	0.174
6	<0.05	0.213	0.213

Note. p : observed probability for the null hypothesis that the coefficient is zero in the population. Reported p values are those reported in the published studies and reproduced results are those obtained using the reproducibility procedures described in text.

Table 3

Reproducibility of Structural Equation Modeling Hypothesis Findings:
Reported and Reproduced Statistical Significance Levels

Study Identifier	Reported p value	Reproduced p value
1	<0.05	0.07
2	<0.05	0.211
2	<0.05	0.749
2	<0.05	0.511
2	<0.05	0.763
2	<0.05	0.505
2	<0.05	0.912
2	<0.05	0.898
2	<0.05	0.912
2	<0.05	0.822
2	<0.05	0.822
2	<0.05	0.053

Note. p : observed probability for the null hypothesis that the coefficient is zero in the population. Reported p values are those reported in the published studies and reproduced results are those obtained using the reproducibility procedures described in text.

Biographies

Donald D. Bergh is the Louis D. Beaumont Chair of Business Administration and Professor of Management at the University of Denver. He received a PhD from the University of Colorado at Boulder. His research in corporate strategy and research methods has appeared in the *Academy of Management Journal*, *Strategic Management Journal*, *Organization Science*, *Journal of Management*, *Journal of Management Studies* and *Organizational Research Methods*. In addition to serving on the editorial review boards of these journals, Dr. Bergh has been an associate editor of the *Academy of Management Journal*, *Organizational Research Methods*, *Journal of Management Studies* and in July of 2017, will become Chair of the Scientific Integrity and Rigor Task Force of the *Journal of Management*.

Barton M. Sharp is the Mike and Kristina McGrath Professor of Entrepreneurship in the Department of Management at Northern Illinois University. He received a PhD from Purdue University. His research interests include the precedents and consequences of innovation within established organizations, corporate strategy, and research methodologies. He has published work in the *Journal of Management*, *Academy of Management Learning & Education*, and *Organizational Research Methods*, among others.

Herman Aguinis is the Avram Tucker Distinguished Scholar and Professor of Management in the School of Business, George Washington University. His research interests span human resource management, organizational behavior, and research methods and analysis topics. He published five books and about 140 articles in refereed journals, delivered about 240 presentations and keynote addresses at professional conferences, and delivered more than 120 invited presentations in all seven continents except for Antarctica. He is a Fellow of the Academy of Management, American Psychological Association, Association for Psychological Science, and Society for Industrial and Organizational Psychology. He served as editor-in-chief of *Organizational Research Methods* and received numerous awards including the Academy of Management Research Methods Division Distinguished Career Award for lifetime scientific contributions, Academy of Management Entrepreneurship Division IDEA Thought Leader Award, and Best Article of Year Awards from *Personnel Psychology*, *Journal of Organizational Behavior*, *Academy of Management Perspectives*, and *Organizational Research Methods*.

Ming Li is a senior lecturer at University of Liverpool Management School. She received a PhD in management from University College Dublin. Her current research interests include cross-cultural management, international HRM, global leadership, Chinese management, and application of research methods. She has published work in journals such as

the Academy of Management Learning & Education, Personality and Individual Differences, and Organizational Research Methods.

Appendix A: Articles that were re-analyzed

Afuah A (2002) Mapping technological capabilities into product markets and competitive advantage, the case of cholesterol drugs. *Strategic Management Journal* 23: 171–179.

Agarwal R, Anand J, Bercovitz J and Croson R (2012) Spillovers across organizational architectures, the role of prior resource allocation and communication in post-acquisition coordination outcomes. *Strategic Management Journal* 33: 710–733.

Aggarwal VA, Siggelkow N and Singh H (2011) Governing collaborative activity, interdependence and the impact of coordination and exploration. *Strategic Management Journal* 32: 705–730.

Anand BN and Khanna T (2000) Do firms learn to create value? The case of alliances. *Strategic Management Journal* 21: 295–315.

Barnett ML and Salomon RM (2012) Does it pay to be really good? Addressing the shape of the relationship between social and financial performance. *Strategic Management Journal* 33: 1304–1320.

Barroso A and Giarratana MS (2013) Product proliferation strategies and firm performance, the moderating role of product space complexity. *Strategic Management Journal* 34: 1435–1452.

Bloom N, Kretschmer T and Van Reenen J (2011) Are family-friendly workplace practices a valuable firm resource? *Strategic Management Journal* 32: 343–367.

Brews PJ and Tucci CL (2004) Exploring the structural effects of internetworking. *Strategic Management Journal* 25: 429–451.

Bruton GD, Ahlstrom D and Wan JCC (2003) Turnaround in East Asian firms, evidence from ethnic overseas Chinese communities. *Strategic Management Journal* 24: 519–540.

Capron L and Pistre N (2002) When do acquirers earn abnormal returns? *Strategic Management Journal* 23: 781–794.

Carow K, Heron R and Saxton T (2004) Do early birds get the returns? An empirical investigation of early-mover advantages in acquisitions. *Strategic Management Journal* 25: 563–585.

Carpenter MA (2002) The implications of strategy and social context for the relationship between top management team heterogeneity and firm performance. *Strategic Management Journal* 23: 275–284.

Carpenter MA, Pollock TG and Leary MM (2003) Testing a model of reasoned risk-taking, governance, the experience of principals and agents, and global strategy in high-technology IPO firms. *Strategic Management Journal* 24: 803–820.

Carpenter MA and Sanders WMG (2002) Top management team compensation, the missing link between CEO pay and firm performance? *Strategic Management Journal* 23: 367–375.

Chadwick C, Hunter LW and Walston SL (2004) Effects of downsizing practices on the performance of hospitals. *Strategic Management Journal* 25: 405–427.

Chung W and Kalnins A (2001) Agglomeration effects and performance, a test of the Texas lodging industry. *Strategic Management Journal* 22: 969–988.

De Figueiredo RJP, Meyer-Doyle P and Rawley E (2013) Inherited agglomeration effects in hedge fund spawns. *Strategic Management Journal* 34: 843–862.

Desender KA, Aguilera RV, Crespi R and García-cestona M (2013) When does ownership matter? Board characteristics and behavior. *Strategic Management Journal* 34: 823–842.

Dezsö CL and Ross DG (2012) Does female representation in top management improve firm performance? A panel data investigation. *Strategic Management Journal* 33: 1072–1089.

El Akremi A, Mignonac K and Perrigot R (2011) Opportunistic behaviors in franchise chains, the role of cohesion among franchisees. *Strategic Management Journal* 32: 930–948.

Fabrizio KR and Thomas LG (2012) The Impact of Local Demand on Innovation in a Global Industry. *Strategic Management Journal* 33: 42-64.

Filatotchev I and Bishop K (2002) Board composition, share ownership and ‘underpricing’ of U.K. IPO firms. *Strategic Management Journal* 23: 941–955.

Fleming L and Sorenson O (2004) Science as a map in technological search. *Strategic Management Journal* 25: 909–928.

Frost TS, Birkinshaw JM and Ensign PC (2002) Centers of excellence in multinational corporations. *Strategic Management Journal* 23: 997–1018.

Gary MS and Wood RE (2011) Mental models, decision rules, and performance heterogeneity. *Strategic Management Journal* 32: 569–594.

Geringer JM, Tallman S and Olsen DM (2000) Product and international diversification among Japanese multinational firms. *Strategic Management Journal* 21: 51–80.

Golden BR and Zajac EJ (2001) When will boards influence strategy? Inclination \times power = strategic change. *Strategic Management Journal* 22: 1087–1111.

Gore AK, Matsunaga S and Yeung PE (2011) The role of technical expertise in firm governance structure, evidence from chief financial officer contractual incentives. *Strategic Management Journal* 32: 771–786.

- Gupta AK and Govindarajan V (2000) Knowledge flows within multinational corporations. *Strategic Management Journal* 21: 473–496.
- Hambrick DC and Cannella AA (2004) CEOs who have COOs, contingency analysis of an unexplored structural form. *Strategic Management Journal* 25: 959–979.
- Hasan I, Kobeissi N and Wang H (2011) Global equity offerings, corporate valuation, and subsequent international diversification. *Strategic Management Journal* 32: 787–796.
- Hayward MLA (2002) When do firms learn from their acquisition experience? Evidence from 1990 to 1995. *Strategic Management Journal* 23: 21–39.
- Hayward MLA (2003) Professional influence, the effects of investment banks on clients' acquisition financing and performance. *Strategic Management Journal* 24: 783–801.
- Hsu DH and Ziedonis RH (2013) Resources as dual sources of advantage, implications for valuing entrepreneurial-firm patents. *Strategic Management Journal* 34: 761–781.
- Hult GTM and Ketchen DJ (2001) Does market orientation matter? A test of the relationship between positional advantage and performance. *Strategic Management Journal* 22: 899–906.
- Kale P, Singh H and Perlmutter H (2000) Learning and protection of proprietary assets in strategic alliances, building relational capital. *Strategic Management Journal* 21: 217–237.
- Khanna T and Rivkin JW (2001) Estimating the performance effects of business groups in emerging markets. *Strategic Management Journal* 22: 45–74.
- Klepper S and Simons KL (2000) Dominance by birthright, entry of prior radio producers and competitive ramifications in the U.S. television receiver industry. *Strategic Management Journal* 21: 997–1016.
- Knight D, Pearce CL, Smith KG, Olian JD, Sims HP, Smith KA and Flood P (1999) Top management team diversity, group process, and strategic consensus. *Strategic Management Journal* 20: 445–465.
- Knott AM (2003) Persistent heterogeneity and sustainable innovation. *Strategic Management Journal* 24: 687–705.
- Kotabe M, Martin X and Domoto H (2003) Gaining from vertical partnerships, knowledge transfer, relationship duration, and supplier performance improvement in the U.S. and Japanese automotive industries. *Strategic Management Journal* 24: 293–316.
- Larsen MM, Manning S and Pedersen T (2013) Uncovering the hidden costs of offshoring, The interplay of complexity, organizational design, and experience. *Strategic Management Journal* 34: 533–552.

Lee C, Lee K and Pennings JM (2001) Internal capabilities, external networks, and performance, a study on technology-based ventures. *Strategic Management Journal* 22: 615–640.

Lee SH and Weng DH (2013) Does bribery in the home country promote or dampen firm exports? *Strategic Management Journal* 34: 1472–1487.

Li H and Atuahene-Gima K (2002) The adoption of agency business activity, product innovation, and performance in Chinese technology ventures. *Strategic Management Journal* 23: 469–490.

Lim DSK, Celly N, Morse EA and Rowe WG (2013) Rethinking the effectiveness of asset and cost retrenchment, the contingency effects of a firm's rent creation mechanism. *Strategic Management Journal* 34: 42–61.

Lu JW and Beamish PW (2001) The internationalization and performance of SMEs. *Strategic Management Journal* 22: 565–586.

Matusik SF (2002) An empirical investigation of firm public and private knowledge. *Strategic Management Journal* 23: 457–467.

McEvily B and Zaheer A (1999) Bridging ties, a source of firm heterogeneity in competitive capabilities. *Strategic Management Journal* 20: 1133–1156.

McEvily SK and Chakravarthy B (2002) The persistence of knowledge-based advantage, an empirical test for product performance and technological knowledge. *Strategic Management Journal* 23: 285–305.

McNamara GM, Luce RA and Tompson GH (2002) Examining the effect of complexity in strategic group knowledge structures on firm performance. *Strategic Management Journal* 23: 153–170.

Mihalache OR, Jansen JJJP, Van Den Bosch FAJ and Volberda HW (2012) Offshoring and firm innovation, the moderating role of top management team attributes. *Strategic Management Journal* 33: 1480–1498.

Miller DJ (2004) Firms' technological resources and the performance effects of diversification, a longitudinal study. *Strategic Management Journal* 25: 1097–1119.

Muller A and Kräussl R (2011) Doing good deeds in times of need, a strategic perspective on corporate disaster donations. *Strategic Management Journal* 32: 911–929.

Nell PC and Ambos B (2013) Parenting advantage in the MNC, An embeddedness perspective on the value added by headquarters. *Strategic Management Journal* 34: 1086–1103.

Nelson T (2003) The persistence of founder influence, management, ownership, and performance effects at initial public offering. *Strategic Management Journal* 24: 707–72

- O'Brien JP (2003) The capital structure implications of pursuing a strategy of innovation. *Strategic Management Journal* 24: 415–431.
- Obloj T and Capron L (2011) Role of resource gap and value appropriation, effect of reputation gap on price premium in online auctions. *Strategic Management Journal* 32: 447–456.
- Poppo L and Zenger T (2002) Do formal contracts and relational governance function as substitutes or complements? *Strategic Management Journal* 23: 707–725.
- Qian C, Cao Q and Takeuchi R (2013) Top management team functional diversity and organizational innovation in China, the moderating effects of environment. *Strategic Management Journal* 34: 110–120.
- Ragozzino R and Reuer JJ (2011) Geographic distance and corporate acquisitions, signals from IPO firms. *Strategic Management Journal* 32: 876–894.
- Ramaswamy K, Li M and Veliyath R (2002) Variations in ownership behavior and propensity to diversify, a study of the Indian corporate context. *Strategic Management Journal* 23: 345–358.
- Ray G, Barney JB and Muhanna WA (2004) Capabilities, business processes, and competitive advantage, choosing the dependent variable in empirical tests of the resource-based view. *Strategic Management Journal* 25: 23–37.
- Reitzig M and Sorenson O (2013) Biases in the selection stage of bottom-up strategy formulation. *Strategic Management Journal* 34: 782–799.
- Reuer JJ (2001) From hybrids to hierarchies, shareholder wealth effects of joint venture partner buyouts. *Strategic Management Journal* 22: 27–44.
- Robertson CJ and Watson A (2004) Corruption and change, the impact of foreign direct investment. *Strategic Management Journal* 25: 385–396.
- Robinson KC and McDougall PP (2001) Entry barriers and new venture performance, a comparison of universal and contingency approaches. *Strategic Management Journal* 22: 659–685.
- Rothaermel FT (2001) Incumbent's advantage through exploiting complementary assets via interfirm cooperation. *Strategic Management Journal* 22: 687–699.
- Rothaermel FT and Deeds DL (2004) Exploration and exploitation alliances in biotechnology, a system of new product development. *Strategic Management Journal* 25: 201–221.
- Rowley T, Behrens D and Krackhardt D (2000) Redundant governance structures, an analysis of structural and relational embeddedness in the steel and semiconductor industries. *Strategic Management Journal* 21: 369–386.

- Seamans RC (2013) Threat of entry, asymmetric information, and pricing. *Strategic Management Journal* 34: 426–444.
- Seth A, Song KP and Pettit RR (2002) Value creation and destruction in cross-border acquisitions, an empirical analysis of foreign acquisitions of U.S. firms. *Strategic Management Journal* 23: 921–940.
- Shen W and Cannella AA Jr (2003) Will succession planning increase shareholder wealth? Evidence from investor reactions to relay CEO successions. *Strategic Management Journal* 24: 191–198.
- Shinkle GA and Kriauciunas AP (2012) The impact of current and founding institutions on strength of competitive aspirations in transition economies. *Strategic Management Journal* 33: 448–458.
- Simonin BL (1999) Ambiguity and the process of knowledge transfer in strategic alliances. *Strategic Management Journal* 20: 595–623.
- Skaggs BC and Youndt M (2004) Strategic positioning, human capital, and performance in service organizations, a customer interaction approach. *Strategic Management Journal* 25: 85–99.
- Slater SF and Olson EM (2000) Strategy type and performance, the influence of sales force management. *Strategic Management Journal* 21: 813–829.
- Spanos YE, Zaralis G and Lioukas S (2004) Strategy and industry effects on profitability, evidence from Greece. *Strategic Management Journal* 25: 139–165.
- Thomas LG (2004) Are we all global now? Local vs. foreign sources of corporate competence, the case of the Japanese pharmaceutical industry. *Strategic Management Journal* 25: 865–886.
- Tian J, Haleblan J and Rajagopalan N (2011) The effects of board human and social capital on investor reactions to new CEO selection. *Strategic Management Journal* 32: 731–747.
- Tippins MJ and Sohi RS (2003) IT competency and firm performance, is organizational learning a missing link? *Strategic Management Journal* 24: 745–761.
- Vermeulen F and Barkema H (2002) Pace, rhythm, and scope, process dependence in building a profitable multinational corporation. *Strategic Management Journal* 23: 637–653.
- Vicent–Lorente JD (2001) Specificity and opacity as resource-based determinants of capital structure, evidence for Spanish manufacturing firms. *Strategic Management Journal* 22: 157–177.

Wales WJ, Parida V and Patel PC (2013) Too much of a good thing? Absorptive capacity, firm performance, and the moderating role of entrepreneurial orientation. *Strategic Management Journal* 34: 622–633.

Walker G, Madsen TL and Carini G (2002) How does institutional change affect heterogeneity among firms? *Strategic Management Journal* 23: 89–104.

Yli-Renko H, Autio E and Sapienza HJ (2001) Social capital, knowledge acquisition, and knowledge exploitation in young technology-based firms. *Strategic Management Journal* 22: 587–613.

Zatzick CD, Moliterno TP and Fang T (2012) Strategic (MIS)FIT, The implementation of TQM in manufacturing organizations. *Strategic Management Journal* 33: 1321–1330.

Zollo M and Singh H (2004) Deliberate learning in corporate acquisitions, post-acquisition strategies and integration capability in U.S. bank mergers. *Strategic Management Journal* 25: 1233–1256

Appendix B: Three Illustrations of Retests using *SMJ* Study Data

Illustration 1. Stata code used to reproduce the linear regression models reported by Lee, Lee, and Pennings (2001)

```
clear
matrix drop _all
matrix input Mean =
    (32.715,0.00,0.00,1.577,2.350,0.504,1.810,1.190,0.00,0.00,30.584,4.182,82.917,9.891,13.975,0.5
    62,0.146)
matrix input SD =
    (77.115,0.657,0.917,3.958,2.342,1.138,1.784,2.724,0.7550,0.6049,43.885,2.784,272.066,17.940,6.
    586,0.498,0.354)
matrix input Corr = (1.000,0.47,0.45,0.76,0.02,0.62,0.11,-
    0.01,0.15,0.18,0.53,0.18,0.07,0.06,0.06,0.09,-0.12\0.47,1.000,0.55,0.42,0.03,0.37,0.22,-
    0.02,0.29,0.25,0.34,0.16,0.24,-0.02,0.27,-0.05,-
    0.10\0.45,0.55,1.00,0.42,0.08,0.38,0.24,0.03,0.28,0.33,0.44,0.34,0.05,0.03,0.30,-0.12,-
    0.04\0.76,0.42,0.42,1.00,0.07,0.67,0.07,-0.01,0.18,0.25,0.76,0.29,0.04,0.16,0.13,0.12,-
    0.11\0.02,0.03,0.08,0.07,1.00,0.07,0.31,0.13,0.17,0.25,0.08,0.09,-0.13,0.10,0.12,0.16,-
    0.15\0.62,0.37,0.38,0.67,0.07,1.00,0.19,0.05,0.24,0.34,0.64,0.26,-0.03,0.13,0.07,0.04,-
    0.01\0.11,0.22,0.24,0.07,0.31,0.19,1.00,0.12,0.20,0.30,-0.03,0.02,0.17,-0.05,0.18,-0.20,0.06\
    0.01,-0.02,0.03,-0.01,0.13,0.05,0.12,1.00,0.07,0.12,-0.01,-0.02,-0.05,0.09,0.13,-0.05,-
    0.01\0.15,0.29,0.28,0.18,0.17,0.24,0.20,0.07,1.00,0.38,0.30,0.28,-0.01,0.08,0.13,-0.03,-
    0.05\0.18,0.25,0.33,0.25,0.25,0.34,0.30,0.12,0.38,1.00,0.29,0.30,-0.09,0.12,0.17,-0.01,-
    0.08\0.53,0.34,0.44,0.76,0.08,0.64,-0.03,-0.01,0.30,0.29,1.00,0.43,-0.04,0.13,0.18,0.14,-
    0.08\0.18,0.16,0.34,0.29,0.09,0.26,0.02,-0.02,0.28,0.30,0.43,1.00,-0.04,0.08,0.30,-
    0.01,0.03\0.07,0.24,0.05,0.04,-0.13,-0.03,0.17,-0.05,-0.01,-0.09,-0.04,-0.04,1.00,0.01,0.23,-
    0.04,-0.04\0.06,-0.02,0.03,0.16,0.10,0.13,-0.05,0.09,0.08,0.12,0.13,0.08,0.01,1.00,0.17,0.14,-
    0.10\0.06,0.27,0.30,0.13,0.12,0.07,0.18,0.13,0.13,0.17,0.18,0.30,0.23,0.17,1.00,-0.03,-
    0.05\0.09,-0.05,-0.12,0.12,0.16,0.04,-0.20,-0.05,-0.03,-0.01,0.14,-0.01,-0.04,0.14,-
    0.03,1.00,-0.47\0.12,-0.10,-0.04,-0.11,-0.15,-0.01,0.06,-0.01,-0.05,-0.08,-0.08,0.03,-0.04,-
    0.10,-0.05,-0.47,1.00)
corr2data sales ent tech fin linkent linkventcap linkuniv linkventnet linkfin linkgov size age
    mktgrowth comp entexp ind bio, n(137) means(Mean) corr(Corr) sds(SD)
reg sales size age mktgrowth comp entexp ind bio
reg sales size age mktgrowth comp entexp ind bio ent tech fin
reg sales size age mktgrowth comp entexp ind bio ent tech fin linkent linkventcap linkuniv
    linkventnet linkfin linkgov
```

Table A1

Comparison of Reported and Reproduced Linear Regression Results for Lee, Lee, and Pennings (2001) Predicting Sales Growth

Variable	Model 1		Model 2		Model 3	
	Reported <i>p</i> Value	Reproduced <i>p</i> Value	Reported <i>p</i> value	Reproduced <i>p</i> value	Reported <i>p</i> value	Reproduced <i>p</i> value
Org size	<0.01	<0.01	>0.1	>0.1	>0.1	<0.05
Org age	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
Market growth rate	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
Number of competitors	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
Entrepreneur's experience	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
Computer/software industry	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
Biotech industry	>0.1	>0.1	>0.1	>0.1	>0.1	>0.1
Entrepreneurial orientation			<0.1	<0.1	<0.1	>0.1
Technological capabilities			<0.05	<0.05	<0.05	<0.05
Financial resources			<0.01	<0.01	<0.01	<0.01
Linkage to other enterprise					>0.1	>0.1
Linkage to venture capital					<0.01	<0.01
Linkage to universities					>0.1	>0.1
Linkage to venture networks					>0.1	>0.1
Linkage to financial institutions					>0.1	>0.1
Linkage to government					>0.1	>0.1

Note. *p*, observed probability for the null hypothesis that the coefficient is zero in the population. Reported *p* values are those reported by Lee et al. (2001) and reproduced results are those obtained using the verifiability procedures described in text.

Illustration 2. Illustration of IBM SPSS syntax used to reproduce a regression model reported by Kale, Singh, and Permutter (2000)

```
MATRIX DATA VARIABLES=PP CM RC DUR LER PC
  /CONTENTS=MEAN SD N CORR /FORMAT=UPPER NODIAGONAL.
BEGIN DATA
4.33 4.16 4 3.7 4.13 3.98
1.76 1.68 1.63 3.88 1.89 1.58
200 200 200 200 200 200
0.51 0.49 0.13 0.41 0.39
0.67 0.1 0.64 0.56
0.18 0.68 0.45
0.07 -0.02
0.39
END DATA.
REGRESSION MATRIX=IN(*) /VARIABLES=RC CM PC
/DEP=PC /ENTER.
```

Testing procedure in IBM SPSS, open a new syntax by selecting ‘File–New–Syntax’; paste above syntax in the new syntax editor, select ‘Run–All’. For more details, refer to MATRIX DATA with REGRESSION in *IBM SPSS Statistics 22 Command Syntax Reference*.

Illustration 3. Illustration of IBM SPSS Amos syntax used to reproduce structural equation modeling results reported by Yli-Renko, Autio, and Sapienza (2001)

```
#Region "Header"
Imports System
Imports System.Diagnostics
Imports Microsoft.VisualBasic
Imports AmosEngineLib
Imports AmosGraphics
Imports AmosEngineLib.AmosEngine.TMatrixID
Imports PBayes
#End Region
Module MainModule
    Public Sub Main()
        Dim Sem As AmosEngine
        Sem = New AmosEngine
        Sem.TextOutput
        AnalysisProperties(Sem)
        ModelSpecification(Sem)
        Sem.FitAllModels()
        Sem.Dispose()
    End Sub

    Sub ModelSpecification(Sem As AmosEngine)
        Sem.GenerateDefaultCovariances(False)

        Sem.BeginGroup("C:\Users\haguinis\WP51\DOC\Bergh\Verifiability\Yli-
Renko Table 2 page 603.sav" , "Yli-Renko Table 2 page 603" )
        Sem.GroupName("Group number 1")
        Sem.Path("Knowledg", "Social_I")
        Sem.Path("Knowledg", "Relation")
        Sem.Path("Knowledg", "customer")
        Sem.Path("Knowledg", "Economic")
        Sem.Path("numbe", "Economic")
        Sem.Path("Technolo", "Economic")
        Sem.Path("Sales_co", "Economic")
        Sem.Path("Knowledg", "Firm_age")
        Sem.Path("numbe", "Firm_age")
        Sem.Path("Technolo", "Firm_age")
        Sem.Path("Sales_co", "Firm_age")
        Sem.Path("Knowledg", "Firm_siz")
        Sem.Path("numbe", "Firm_siz")
        Sem.Path("Technolo", "Firm_siz")
        Sem.Path("Sales_co", "Firm_siz")
        Sem.Path("Knowledg", "Internat")
        Sem.Path("numbe", "Internat")
        Sem.Path("Technolo", "Internat")
        Sem.Path("Sales_co", "Internat")
        Sem.Path("Knowledg", "e1", 1)
        Sem.Path("numbe", "e2", 1)
        Sem.Path("Technolo", "e3", 1)
    End Sub
End Module
```

```

Sem.Path("Sales_co", "e4", 1)
Sem.Path("numbe", "Knowledg")
Sem.Path("Technolo", "Knowledg")
Sem.Path("Sales_co", "Knowledg")

Sem.Model("Default model", "")
End Sub

Sub AnalysisProperties(Sem As AmosEngine)
Sem.Iterations(50)
Sem.InputUnbiasedMoments
Sem.FitMLMoments
Sem.Standardized
Sem.TotalEffects
Sem.Mods(4)
Sem.Seed(1)
End Sub
End Module

```

Table A2
Correlations, Means, and Standard Deviations Used as Input for Reproducing Results
(from Yli-Renko et al, 2001, Table 2)

	rowtype_	varname_	Social_I	Relation	customer	Knowledg	Economic	Firm_age	Firm_siz	Internat	numbe	Technolo	Sales_co
1	n		180.00	180.00	180.00	180.00	180.00	180.00	180.00	180.00	180.00	180.00	180.00
2	corr	Social_I	1.00
3	corr	Relation	.15	1.00
4	corr	customer	.14	.17	1.00
5	corr	Knowledg	.27	-.04	.38	1.00
6	corr	Economic	.15	.14	-.15	.09	1.00
7	corr	Firm_age	-.10	.05	-.14	-.15	.09	1.00
8	corr	Firm_siz	.01	-.19	-.04	.04	-.11	.16	1.00
9	corr	Internat	.00	-.06	-.16	-.02	.13	.11	.19	1.00	.	.	.
10	corr	numbe	.13	-.08	.19	.28	.08	-.11	.17	-.05	1.00	.	.
11	corr	Technolo	.08	-.02	.18	.23	.04	-.04	.08	.28	.05	1.00	.
12	corr	Sales_co	.05	-.05	.03	-.15	-.03	-.06	-.09	-.03	-.25	.04	1.00
13	stddev		1.71	1.32	1.88	1.46	23.92	2.79	46.94	33.45	4.58	1.57	54.44
14	mean		4.28	4.92	3.99	3.39	28.92	5.96	23.81	34.03	2.25	4.83	33.34

Note. Social_I, social interaction, Relation, relationship quality, customer, customer network ties, Knowledg, knowledge acquisition, Economic, economic exchange, Firm_age, firm age, Firm_siz, firm size, Internat, internationalization, numbe, number of new products developed, Technolo, technological distinctiveness, and Sales_co, sales cost. n = sample size, corr, correlation, stddev, standard deviation.

Table A3
Comparison of Reported and Reproduced Results for Yli-Renko et al. (2001)

Path coefficient	Reported <i>p</i>	Reproduced <i>p</i>
Social interaction → Knowledge acquisition	< 0.20	0.002
Relationship quality → Knowledge acquisition	< 0.20	0.029
Customer network ties → Knowledge acquisition	< 0.002	< 0.001
Knowledge acquisition → New product development	< 0.002	< 0.001
Knowledge acquisition → Technological distinctiveness	< 0.002	0.001
Knowledge acquisition → Sales cost	< 0.20	0.036

Note. p , observed probability for the null hypothesis that the coefficient is zero in the population. Reported p values are those reported by Yli-Renko et al. (2001) and reproduced results are those obtained using the verifiability procedures described in text.