Journal of International Business Studies (2025) 56:1102–1124
https://doi.org/10.1057/s41267-025-00808-9

The insights from the crowd: Drawing inferences from many approaches to key empirical questions in international business

Andrew Delios · Tianyou Hu · Shu Yu · Nan Zhou · Eric Uhlmann, et al. ***[full author details at the end of the article]***

Received: 1 August 2024 / Revised: 29 May 2025 / Accepted: 11 June 2025 / Published online: 5 November 2025

Abstract
In this crowdsourced initiative, 57 independent analysts used the same longitudinal dataset to address four major empirical questions in international business. For all four research questions, different analysts obtained substantial estimates in opposite directions, meaning that they could have drawn any conclusion at all had they conducted the project alone. Aggregating across the results obtained by different analysts pointed to an overall answer for two of the four research questions, although for one of the two questions, the evidence was more suggestive than conclusive. That said, the variability in results was not simply random, and could in some cases be meaningfully explained. Choices regarding how to operationalize variables played an important role in determining the empirical results, and expert analysts were more likely to report large positive effects. Rather than exhibiting a bias to confirm their pre-existing beliefs, analysts appeared to rationally update their beliefs considering the evidence. Overall, these findings empirically demonstrate the role of subjective researcher choices in shaping results in international business research yet also show that it is still possible to draw meaningful conclusions in science. We advocate for an open science of international business in which the consequences of subjective analytic choices are rendered as transparent as possible.

Keywords Crowdsourcing · Many analysts · Open science · Entry mode strategy · Multinational firms

Introduction

Since 2010, we have seen increasing concerns about publication incentives and researcher confirmation bias negatively affecting the reliability of the scientific literature (Aguinis et al., 2020; Nelson et al., 2018; Nosek et al., 2022). Simulations demonstrate that given sufficient choice points in a dataset, researchers could, in principle, select the path that supports whatever they are (consciously or perhaps unconsciously) biased to conclude (Murphy & Aguinis, 2019; Simmons et al., 2011). In part, because of such opportunities to "*p*-hack", *p* value distributions in published articles reflect substantial numbers of statistically implausible findings (Fanelli, 2010; Goldfarb & King, 2016; Ioannidis, 2005; Simonsohn et al., 2014). Indeed, one predictor of the results of an empirical scientific investigation is the prior intellectual commitments of the research team (Berman & Reich, 2010).

The consequent reform movement to increase the reproducibility, replicability, and robustness of scientific findings across disciplines holds value for research on international business (IB) just as it does for other fields (Aguinis et al., 2017, 2022, 2023). Although less prone to the small samples and underpowered tests that characterize many behavioral experiments (Schimmack, 2012; Weingarten et al., 2016), international business studies often rely on complex longitudinal datasets with many choice points and defensible analytic approaches. Meta-scientific investigations suggest concerning rates of scientific errors (Bergh et al., 2017), difficulties reproducing the same findings from the same

Andrew Delios, Tianyou Hu, Shu Yu, and Nan Zhou share first authorship. They, together with Eric Luis Uhlmann, formed the core project team that designed and conducted the meta-scientific project and wrote the paper. The remaining 64 co-authors served as analysts who carried out the analyses for the four research questions.

Correspondence to: Tianyou Hu, Faculty of Business Administration, University of Macau, E22, Avenida da Universidade, Taipa, Macau. Email: tianyouhu@um.edu.mo; Nan Zhou, School of Economics and Management & Advanced Institute of Business, Tongji University, 1500 Siping Road, Shanghai, China, 200092. Email: zhounan38@hotmail.com.

Accepted by Arjen van Witteloostuijn, Area Editor, 11 June 2025. This article has been with the authors for four revisions.

data (Bergh et al., 2017; Delios et al., 2022), and effect size overestimation (Bosco et al., 2016; Goldfarb & King, 2016) in strategy and international business—just as has been observed across all disciplines examined thus far (e.g., Fanelli, 2010; Ioannidis, 2005; Open Science Collaboration, 2015). In this study, we demonstrate how empirical estimates are contingent on subjective analytic choices, which we suggest represents a more significant challenge to the science of international business than *p*-hacked or suboptimal analyses.

Prior theoretical scholarship has raised concerns about epistemic uncertainty in management research (Ketokivi & Mantere, 2010; King et al., 2021; Mantere & Ketokivi, 2013), and the dependency of results on analytic choices has been directly demonstrated for a few specific claims (Berchicci & King, 2022; Goldfarb & King, 2016; Goldfarb & Yan, 2021; Nandialath & Rogmans, 2019). Building on Goldfarb and King (2016), who estimate statistically that false positives are common in strategic management research (as in many fields), we examine whether results that are true-positives under a given specification may not always be robust to the alternative specifications other capable scholars might have used. By organizing a mass collaboration, we provide a birds-eye empirical view of the importance of context in this space as theorized by Ketokivi and Mantere (2010).

We selected four unanswered research questions that are intrinsically interesting in-and-of-themselves to exploit the insights of the crowd to clarify the nature of the relationships between key variables in international business studies, such as intangible assets, policy uncertainty, ownership, and firm performance. This first-order contribution to address the research questions themselves, some of which are the subject of dozens of prior publications that rendered mixed and conflicting results, is a key value-add beyond the second-order contribution of turning a meta-scientific lens on the scientific process itself.

Our study shows that IB as a discipline is not immune to the subjectivity challenge uncovered in other fields. At the same time, our target questions varied in the degree of theoretical consensus in the field regarding what pattern "should" emerge, allowing us to examine cross-question differences in empirical conclusions. In addition to aggregating results across disparate approaches in an attempt to draw substantive inferences (Landy et al., 2020; Nandialath & Rogmans, 2019), we examine meaningful moderators (Berchicci & King, 2022; Ketokivi & Mantere, 2010; King et al., 2021) such as theoretically laden variable operationalizations, researcher expertise, and beliefs, with more success than in any prior many-analysts initiative from any field. In the Discussion, we present a typology of interventions designed to tackle subjectivity in science while also discussing their downsides. We outline an optimistic vision for an open science of IB that considers both the benefits and costs of potential reforms as well as promising strategies for dealing with heterogeneous estimates, such as aggregation and moderator identification.

Research background

Subjectivity in science

Even absent analytic mistakes, difficulties reproducing findings, publication pressures, or an intellectual commitment to a theory, there remains some inherent subjectivity in scientific approaches to a problem (Landy et al., 2020; Silberzahn et al., 2018). For complex datasets, this is rendered transparent by a multiverse in which a solo researcher or small team conducts many analyses (Sala-i-Martin, 1997; Simonsohn et al., 2020; Steegen et al., 2016; Young & Holsteen, 2017) or a crowdsourced approach in which many scientists use the same dataset to test the same research question (Silberzahn et al., 2018). Crowdsourcing is far less efficient than a multiverse yet it reveals the naturally emerging strategies of real researchers. A many-analysts approach can directly address the question of what would have happened if another investigator had analyzed the same data and further examined which of their characteristics (degree of topic or statistical expertise, beliefs about the hypothesis, etc.), made a difference in the outcomes of their empirical tests.

In one early many-analysts initiative, 29 research teams tested the relationship between the skin tone of soccer players and whether they received red cards from referees. The estimates obtained from the 29 research teams ranged from large positive estimates, consistent with the conclusion that referees are biased against darker-skinned players, to small negative effects reflecting the opposite directional bias (Silberzahn et al., 2018). Most subsequent crowd collaborations find an even greater heterogeneity in results across independent analysts (Botvinik-Nezer et al., 2020; Breznau et al., 2022; Menkveld et al., 2024; Schweinsberg et al., 2021). Schweinsberg et al. (2021) refer to it as "radical effect size dispersion" when different researchers analyzing the same dataset to address the same research question return different effect size estimates in opposite directions.

For example, one analyst may conclude that including more women in a scientific debate increases the likelihood that each individual woman will speak, whereas another finds that the presence of more women suppresses female participation (Schweinsberg et al., 2021). Related work reveals such a wide dispersion of results across experimental studies designed by independent research teams (Baribault et al., 2018; Huber et al., 2023; Landy et al., 2020; Tierney et al., 2025). To the extent that such a pattern emerges frequently for the hypotheses tested in a scientific field, the implications regarding the products of traditional science done in small teams are considerable. When reading a typical published paper, the reader is hence left unsure whether to trust the results; if another team had pursued the same idea, they might have reached the reverse conclusion.

Two means of managing the uncertainty: Aggregation versus parsing

When different scientists using disparate approaches reach different conclusions, how can we adjudicate? The two major approaches to managing variability and uncertainty in science are *aggregation* and *parsing* (Berchicci & King, 2022; Cyrus-Lai et al., 2022). Aggregation involves mechanically averaging across the results from the many analyses or many designs, for example, via meta-analysis (Landy et al., 2020) or Bayesian model averaging (Hinne et al., 2020; Nandialath & Rogmans, 2019). Parsing means identifying meaningful empirical moderators of the different estimates across diverse approaches (Berchicci & King, 2022). Perhaps different researchers interpreted the research question differently (Auspurg & Brüderl, 2021; Breznau et al., 2022; Kummerfeld & Jones, 2023) or obtained divergent results because they operationalized variables in distinct ways (Schweinsberg et al., 2021). The associated differences in results could be non-arbitrary and hold theoretical implications that are lost via aggregation. Parsing is consistent with the perspectivist thesis that the opposite of a great truth is also true (McGuire, 1983), and that the task of the scientist is to unravel this web of theoretically rich moderation.

Although perspectivism offers a beautiful vision of scientific inquiry, the many-analysts projects thus far have struggled to quantitatively explain much variance in estimates. Silberzahn et al. (2018) failed to identify reliable moderators of the results across analysis teams, finding only null effects of expertise and other variables. Schweinsberg et al. (2021) demonstrated that theoretically laden operationalizations of variables (e.g., how status is conceptualized) accounted for a meaningful portion of the results, but only a small amount in absolute terms. Breznau et al. (2022) considered numerous predictor variables, which combined explained only 4% of the variability in results across analysts. Huber et al. (2023) coded features of their crowdsourced many designs, which together captured 3.1–6.4% of the variance in estimates. Baribault et al. (2018) had more success, demonstrating that supraliminal priming designs that presented stimuli above the threshold of conscious perception produced reliable priming effects, whereas subliminal designs did not.

Past crowd initiatives have proved groundbreaking, but also hold major limitations, especially with regards to cracking the parsing problem. Some prior investigations used arbitrarily selected research questions, such as possible racial bias among soccer referees (Silberzahn et al., 2018; see also Landy et al., 2020). Others generated and selected hypotheses to test leveraging a crowd of scientists (Schweinsberg et al., 2021) or chose research questions the project coordinators considered important to their respective fields (e.g., Breznau et al., 2022; Menkveld et al., 2024). We consider these major improvements in the selection process for research questions, since the extremely inefficient process of organizing and coordinating a small army of collaborators is better justified by a high-stakes outcome (Isager et al., 2021). Posing more than one research question in the same project (Landy et al., 2020; Schweinsberg et al., 2021) further affords the opportunity to explore whether the question itself matters. For example, a general research claim with a high latitude of construal could be interpreted differently by different scientists, increasing the diversity of methodological strategies used to tackle the problem (Auspurg & Brüderl, 2021; Kummerfeld & Jones, 2023).

Although the number of collaborators involved was far higher than in a small team project, the numbers of data analysts or materials designers in early crowd initiatives were small in absolute terms (e.g., 29 analysis teams in Silberzahn et al., 2018, and 23 individual analysts in Schweinsberg et al., 2021; 15 or fewer teams of materials designers per hypothesis in Landy et al., 2020). Silberzahn et al. (2018) found that team leaders' prior beliefs about referee racial bias did not predict their estimates, but with 29 units of observation, this does not provide strong evidence against researcher confirmation bias. Such samples make it difficult to assess the role of individual differences, for example, in topic expertise or statistical skills, unless the relationship is very strong. False negatives become uncomfortably likely.

More recently, Menkveld et al. (2024) observed a small correlation between expertise and analytic estimates in their sample of 164 teams, in contrast to earlier estimates that were close to zero. Despite the extraordinary coordination effort required, more crowd projects, including as many analysts as possible, are necessary to address the parsing challenge. In particular, does the analyst matter, and if so, which aspects of an analyst's profile are most important?

The present research

We crowdsourced tests of four unanswered research questions in international business studies with a group of 57 collaborators using a common complex longitudinal dataset. A reliable finding that is robust to different analytic approaches is necessary, but not sufficient, for scientific credibility. A credible finding should also be theoretically and practically relevant (RRBM, 2024). We selected these four questions because they are important ones in IB, and each has had many prior investigations targeted at them with a diversity of empirical outcomes and no unequivocal conclusion. The answers, if attainable, are both theoretically interesting to academics and of practical importance to management practitioners. As emphasized earlier, there is an intrinsic value in attempting to answer the specific research questions themselves, above and beyond the meta-scientific question of variability in estimates due to researcher choices.

Further, although these four research questions have been addressed in much empirical research, prior tests have varied in sample size, sample constitution (geography, industry, and time period) and key variables (Shen et al., 2017; Zhao et al., 2004). Thus, it is useful to explore if such variance in results is maintained when data and research questions are fixed across researchers. In our research collective of 57 collaborators, most contributors had direct topic expertise in strategic management or international business, and many had extensive statistical experience.

Epistemic uncertainty is increasingly recognized in management research (Ketokivi & Mantere, 2010; King et al., 2021; Mantere & Ketokivi, 2013), as in other fields such as psychology (Nosek et al., 2022) and medicine (Ioannidis, 2005; Musani et al., 2007). It is therefore valuable to empirically examine whether there is a dispersion in estimates when a crowd of capable researchers attempts to tackle the same questions with the same dataset. We do not claim that the problem is more severe in international business studies than elsewhere, nor do we make field-by-field comparisons. However, quantitative research in international business often relies on complex sets of observations that involve numerous analytic choices. The present meta-scientific insights are relevant to any field or subfield that likewise depends on topic experts to navigate their way through a garden of forking paths (Gelman & Loken, 2014) to reach an empirical conclusion.

Adding further value, the four focal research questions varied in the level of theoretical consensus regarding the expected outcome based on the past literature. This allows us to explore cross-question differences, such as whether subjective theoretical consensus is related to the objective degree of dispersion in empirical estimates across analysts. Even if high intellectual consensus does not necessarily translate to consistency in empirical results, aggregating across widely disparate estimates could still yield directional answers to questions where established theory points to what "should" happen. Thus, the project's value-add in potentially answering some of the research questions themselves is intertwined with the meta-scientific contribution of extending the many-analyst approach to strategic management.

Research question 1 What is the relationship between entry mode and foreign subsidiary performance?

The relationship between entry mode and foreign subsidiary performance is one of the most studied in international business research (Brouthers, 2002; Chen & Hu, 2002; Shen et al., 2017; Wu et al., 2022). Entry mode choice is an important strategic decision in international expansion because it has implications for various critical issues, such as control over foreign operations, investment risk, and resource commitment (Zhao et al., 2004). Past studies suggest it further influences managerial satisfaction regarding subsidiary performance (Brouthers et al., 2003), subsidiary survival (Shaver, 1998), and the objective profitability of the subsidiary (Chen & Hu, 2002). Different theories and perspectives could explain the relationship between entry mode and subsidiary performance, such as transaction cost theory (Brouthers et al., 2003), internationalization theory (Johanson & Vahlne, 1977; Wiedersheim-Paul & Johanson, 1975) and cross-national distance (Tihanyi et al., 2005). Transaction cost theory would predict that sole ownership leads to better performance due to reduced transaction costs. Internationalization theory would expect that shared ownership would lead to better performance due to the reduction of risk and opportunities for learning and resource sharing to either scale resources or link resources along the value chain. Therefore, we expected a high level of theoretical consensus among scholars in the field that entry mode should matter, but conflicting predictions regarding the direction of the effect.

Research question 2 What is the relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries?

The possession of intangible assets that are subject to market failure is one of the reasons that firms invest abroad (Buckley & Casson, 1976). Intangible assets are at the core of internalization theory, which has been widely accepted in the international business literature (Henisz, 2003; Morck & Yeung, 1992; Zeng et al., 2019). There is little theoretical controversy that a higher level of intangible assets should lead to a higher level of ownership because foreign subsidiaries need to protect their intangible assets from leaking to partners (Guillén, 2003; Martin & Salomon, 2003). Thus, we anticipate a high level of intellectual consensus among the collaborators that there should be a positive relationship between intangible assets and the level of ownership. This consensus should be greater than that for research question (RQ1).

Research question 3 What is the relationship between policy uncertainty and a firm's level of ownership in its foreign subsidiaries?

Institutional theory posits that firm strategy is influenced by the external institutional environment (North, 1990; Scott, 2008). Policy uncertainty in the host country increases the risk of operations, and firms could reduce their level of ownership to reduce risk (Delios & Henisz, 2000; Henisz, 2000a). However, this may not always be the case since some firms may be able to deal with policy uncertainty by other strategic means (Lee, 2018; Sun et al., 2016). Thus, our reading of the literature is that the level of academic consensus for RQ3 is not as high as for RQ2.

Research question 4 How does the level of policy uncertainty moderate the relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries?

When the level of policy uncertainty is high, the risk of expropriation by partners is high. An opportunistic local partner would use all available means, such as the manipulation of the political system to seize opportunities for expropriation (Delios & Henisz, 2000). In such a situation, a foreign subsidiary would tend toward full ownership to reduce the risk of expropriation by a local partner (Henisz, 2000a). Accordingly, we expect the relationship between intangible assets and level of ownership to be stronger when the level of policy uncertainty is high. However, this relationship is rarely tested in the literature and does not carry strong explicit assumptions on the part of most scholars (Delios & Henisz, 2000; Henisz, 2000b). Thus, even more so than for RQ3, we treat RQ4 as an open empirical question.

Finally, at a meta-scientific level, our project has several value-adds that reflect learnings and best practices gleaned from past crowd initiatives. To our knowledge, this is only the second such project, after Silberzahn et al. (2018), to assess beliefs about the research questions among our researchers both before and after carrying out the analyses. We are interested in post-analysis beliefs because scientists' views regarding the hypothesis are likely to be shaped by the results they found, and indeed normatively ought to be. If post-beliefs are more strongly correlated with results than pre-beliefs, this suggests belief updating—changing one's beliefs to fall in line with the empirical results, something a good scientist should do.

Our design thus allows us to potentially capture not only confirmation bias but also rational updating of beliefs considering the evidence. We provide a new test of whether analysts and their characteristics matter, examining the potentially dynamic interplay between beliefs and evidence with a sample size approximately double that of Silberzahn et al. (2018). Building on Schweinsberg et al. (2021), we leave variable operationalizations unconstrained and test their potential importance in explaining dispersion in results across analysts. We empirically compare the importance of research question and researcher, as in Landy et al. (2020), but with 57 units of observation rather than 15. Our aim is to provide informative tests of the four unresolved research questions in international business research, and at the same time shed unique new light on the nature of social scientific inquiries based on complex data.

Methods

Data and code availability

All data and code used in this paper are publicly available on the Open Science Framework (OSF) at https://osf.io/ew3vz/?view_only=52ab5518ada34e0ca1046aeab08a5122. Our primary supplementary document that contains Supplements 1–9 is accessible on the JIBS webpage of this paper (https://doi.org/10.1057/s41267-025-00808-9).

Pre-registration

Our methods and analyses were pre-registered at https://osf.io/4euaj. See our original plan in Material 1 in "Further Results and Materials", which is posted on the OSF repository. Supplement 1, located in the supplementary document (on the JIBS webpage of this paper), summarizes deviations from and additions to the originally planned analyses.

Dataset

The "Overseas Japanese Companies" dataset, provided by Toyo Keizai Inc, is widely used in international business research. The version of the dataset we used in this study includes information on Japanese firms' foreign subsidiaries operating in 155 countries during the period from 1991 to 2009. This version of the dataset covers over 2100 firms with more than 26,000 subsidiaries operating in various industrial sectors. According to a report from Western University in October 2022, this dataset has been used in at least 161 peer-reviewed publications across fields such as management and organization research, international business, political science, accounting, finance, and other social sciences (Ivey, 2015).

Analysts

Leveraging our personal networks and email advertisements (Material 10 on OSF), we recruited a crowd of 57 researchers based in 20 different countries and territories. Their ages ranged from 24 to 57, with a mean of 34.88 years (SD = 8.16). Twenty-two (38.60%) self-identified as women and 35 (61.40%) as men. The regions with the most analysts were mainland China (13 analysts, 22.81%) and the United States (12 analysts, 21.05%), followed by Australia (five analysts, 8.77%), Singapore (four analysts, 7.02%), India (three analysts, 5.26%), New Zealand (three analysts, 5.26%), the Czech Republic (two analysts, 3.51%), Hong Kong (two analysts, 3.51%), and Taiwan (two analysts, 3.51%). Austria, Brazil, Colombia, Denmark, Germany, Greece, Ireland, Japan, Malaysia, Thailand, and the United Kingdom had one analyst each. See Supplement 8 for a summary of the analysts' academic profiles and their demographics.

Although analysts came from a variety of disciplinary backgrounds, such as economics, organizational behavior, and finance, the majority (82.46%) of them were from strategic management or international business. They reported an average of 7.01 years of experience with data analysis (SD = 4.39). Forty (70.18%) of them had five or more years of experience with data analysis and 34 (59.65%) analysts performed data analyses at least once per week. The 57 analysts included three Full Professors (5.26%), eight Associate Professors (14.04%), 22 Assistant Professors (38.6%), 20 doctoral students (35.09%), one post-doctoral student, two Masters students and one teaching fellow. Our sample of analysts was skewed towards junior academics. Senior representation also dropped from 24 to 19% over the course of the project, perhaps in part due to the higher opportunity costs of time for Full and Associate Professors. The statistics (Supplement 8) obtained from the pre-survey (Material 2 on OSF) show that 37 (64.91%) analysts had published at least one scientific paper, and 12 (21.05%) analysts had published five or more papers in strategic management or international business. A total of 11 (19.30%) analysts had published at least one paper with a primary contribution in methodology or statistics, and eight (14.04%) had taught at least one statistics class.

Project website

The project website provided detailed information about the project to colleagues potentially interested in taking part (Material 9 on OSF). In the background section, we introduced the purpose of the project and the database that would be employed to test the research questions. The data description portion included an overview and details regarding each variable in the database. In the data analysis section, we provided data in three alternative formats, specifically STATA, SPSS, and Excel. In the FAQ section, we provided answers to common questions analysts might have in mind, alongside one author's contact information if they had any further queries.

Protocol

We asked the data analysts to perform their responsibilities independently by completing five steps: (1) complete the pre-survey, (2) access the research questions, the data and the data descriptor, (3) undertake the analyses to address the four research questions, (4) complete the post-survey, and (5) upload their results and statistical code from their analyses using an online portal.

In the first step, we received 158 responses for the pre-survey, which asked for an analyst's beliefs about each research question as well as their demographic characteristics (Supplement 8 and Material 2 on OSF). Our review of these 158 responses showed that 35 responses had an accomplishment percentage below 50%, which meant at least half of the questions were unanswered in their responses. In addition, in the completed answers, there were four respondents who submitted the pre-survey twice. Ultimately, 119 unique analysts successfully completed the pre-survey.

In Steps 2 and 3, analysts were provided with pooled longitudinal data, which they used to estimate various fitting models, including ordinary least squares, logistic models, tobit models, and generalized linear models, among others. Members of our crowd worked individually—consistent with investigations that are solo-authored and scientific collaborations that include a single data analyst—but in contrast to the broad overall trend across fields towards large teams of collaborators (Wuchty et al., 2007), presumably with multiple members contributing to the analyses. Past many-analyst projects have recruited both teams and individuals, with comparable overall results (e.g., Menkveld et al., 2024; Schweinsberg et al., 2021; Silberzahn et al., 2018), but the use of solo analysts does stand as a limitation of this research.

In Step 4, we received 93 responses to the post-survey. Within these responses, four respondents provided an invalid ID number and name; eight respondents submitted their answers twice, and three respondents did not finish their survey. As such, we had 78 analysts with complete information inclusive of the post-survey, which collected individual analysts' beliefs regarding each research question, and details on how they conducted their analyses. For example, we asked analysts on their methods and theoretical rationale to operationalize important variables (such as Entry Mode, Intangible Assets, Policy Uncertainty, Performance, and Level of Ownership), as well as their reasons for the selection of their statistical technique and control variables (Material 3 on OSF).

In Step 5, we found that 72 of these 78 analysts submitted both results and coding files. We checked each submission and carefully reproduced all the analyses and results. Amongst these 72 analysts, two analysts did not submit complete statistical software codes, and 13 had submissions that are irreproducible due to unrecognizable codes, missing regressions, or ambiguous variables (see Material 14 on OSF for details). As such, our focal analyses are based on the 57 analysts whose submissions are complete and replicable. Among these analysts, four used SPSS to process their analysis while the remaining 53 used Stata. These analysts often provided multiple analyses for one or more research questions; hence, we had 227, 126, 98, and 135 specifications for RQ1–4, respectively. Instituting quality control measures, such as excluding results than cannot be reproduced from the same data and code, helps ensure that the heterogeneity in estimates does not emerge as an artifact of substandard work by some members of the crowd. This presents a conservative test of the "many-analysts" phenomenon in which crowds of analysts obtain disparate estimates using the same data to test the same hypothesis. That said, it is important to be transparent about what specifications and results were excluded and why. We have posted on the OSF all analysts' submissions, including their model specifications and results, along with the data they used and our overall project analysis code. The reasons for exclusions are clearly stated (either "incomplete submission" or "analysis not reproducible") in Material 14. In the spirit of open science, we welcome further perspectives from additional colleagues.

Results

Effect size dispersion across analysts

Analysts used a wide variety of specifications to conduct their analyses. No two analysts adopted the same approach when we consider the variables, method, and sample selection simultaneously. Material 4 on OSF provides a detailed summary table describing the specifications employed by each analyst for each research question. Supplement 2 summarizes the control variables used by the analysts.

To achieve a standardized measure of the effect sizes of the independent variables on dependent variables across all analyses, we computed the marginal effect sizes (Breznau et al., 2022; Fey et al., 2023), which represent the increase in a dependent variable for a unit increase in an independent variable. In Fig. 1, we present the distribution graphs of the effect size estimates found in the separate analyses for each research question. For all four research questions, the range of estimates encompasses both negative and positive values and crosses zero.

The most used threshold for considering an individual estimate statistically significant is the arbitrary value of $p <$ 0.05. Whether or not an estimate is associated with a p value below or above 0.05 gives a sense of what an individual

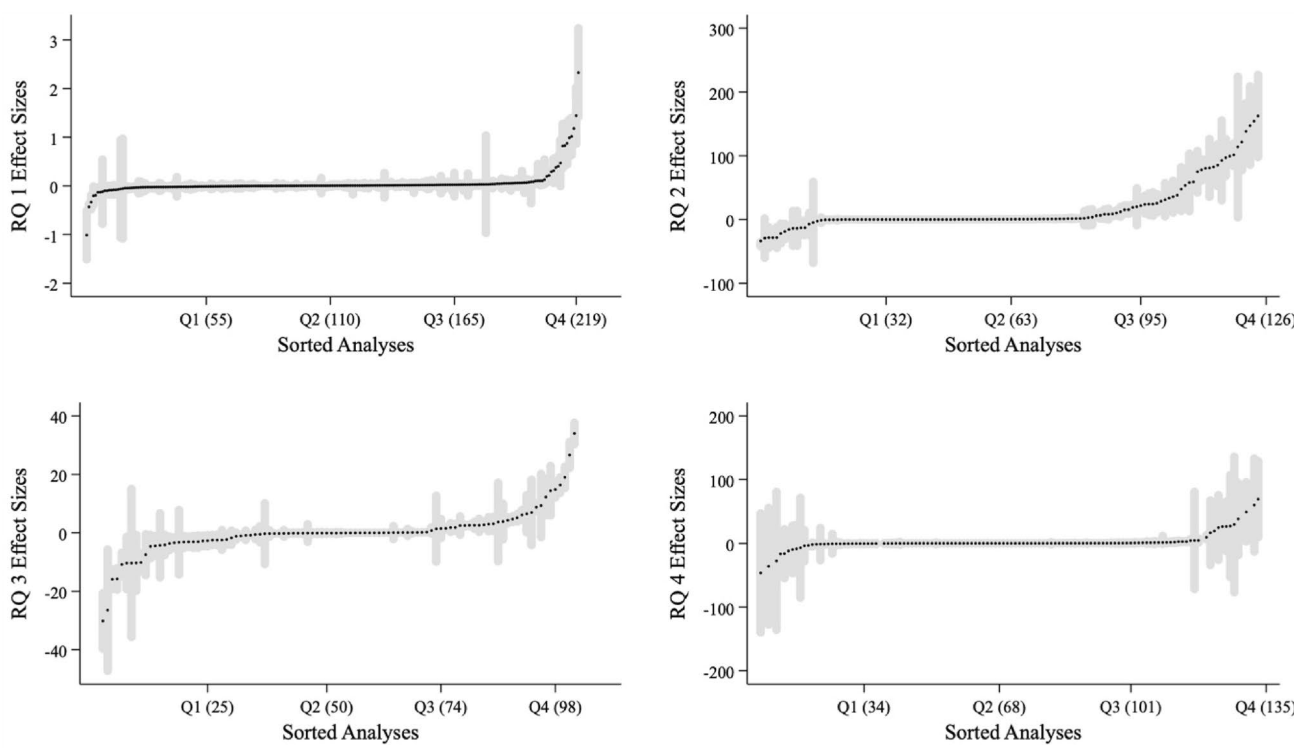

Fig. 1 Analysts' reported marginal effect sizes from their tests of four international business research questions. *Notes*: Quartiles of the number of analyses and 95% confidence intervals of the effect sizes are as indicated in the figure

stand-alone analysis or a paper built around one primary analysis might have concluded regarding each research question. As seen in Table 1, the crowd of analysts produced at least a few positive estimates that crossed the $p < 0.05$ threshold and at least a few negative estimates that crossed the $p < 0.05$ threshold for all four research questions. Thus, different researchers, given the same research question and the same dataset came to directly opposite conclusions for each of the four questions posed to the crowd. As shown in Table 2, the aggregation of analyses exhibits a high level of heterogeneity in estimates that is already apparent visually. All RQs are associated with a high value of Cochran's Q and I^2.

Aggregating across diverse approaches and results

Aggregating across all the specifications revealed an overall directional effect with a 95% confidence interval excluding zero for RQ1 (mean = 0.044, 95% CI = [0.011, 0.077]) and RQ2 (mean = 23.817, 95% CI = [11.967, 35.668]), but not for RQ3 and RQ4 (Table 2). Note that the confidence intervals for the aggregated estimate for RQ1 almost includes zero, which means we draw a cautious conclusion on a directional effect (Benjamin et al., 2018). It is important to note that this inference that we draw from aggregating effect sizes is not based on double-counting, as can happen in a meta-analysis that includes multiple results from the same

Table 1 Results for research questions 1–4 based on direction and whether they cross the conventional $p < .05$ threshold

Research question	Positive effect size & $p < 0.05$	Positive effect size & $p > 0.05$	Negative effect size & $p > 0.05$	Negative effect size & $p < 0.05$
1. What is the relationship between entry mode and foreign subsidiary performance?	27.31% ($n = 62$)	28.19% ($n = 64$)	24.23% ($n = 55$)	20.26% ($n = 46$)
2. What is the relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries?	67.72% ($n = 86$)	5.51% ($n = 7$)	8.66% ($n = 11$)	18.11% ($n = 23$)
3. What is the relationship between policy uncertainty and a firm's level of ownership in its foreign subsidiaries?	32.65% ($n = 32$)	12.24% ($n = 12$)	25.51% ($n = 25$)	29.59% ($n = 29$)
4. How does the level of policy uncertainty moderate the relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries?	20.59% ($n = 28$)	37.50% ($n = 51$)	32.35% ($n = 44$)	9.56% ($n = 13$)

All four research questions were stated to analysts in a non-directional manner. For RQ1, a positive effect size means the analyst estimates a positive relationship between wholly-owned subsidiary (WOS) and foreign subsidiary performance, rather than joint venture (JV) or acquisition. A wholly-owned subsidiary is entirely owned and managed by a parent foreign company. A joint venture is a firm created by local firms and foreign firms, generally with shared ownership, returns, risks, and governance. An acquisition is a transaction in which a foreign firm purchases most or all of another company's shares to gain control of that company. For RQ2, a positive effect size means the analyst estimates a positive relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries. For RQ3, a positive effect means the analyst estimates a positive relationship between policy uncertainty and a firm's level of ownership in its foreign subsidiaries. For RQ4, a positive effect means the analyst finds that high policy uncertainty makes the relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries more positive.

Table 2 Heterogeneity statistics for the four research questions

Research questions	Number of analysts	Number of analyses	Reported effect size means	Q	I^2	Bayes factor bound means	Partial correlation coefficient means
RQ1	52	227	0.044 [0.011, 0.077]	22,835.295	99.05%	138.544 [50.670, 226.417]	0.018 [0.010, 0.026] (222)
RQ2	54	127	23.817 [11.967, 35.668]	3,159.866	96.01%	819.416 [309.150, 1329.681]	0.028 [0.018, 0.038] (72)
RQ3	52	98	− 0.246 [− 1.692, 1.201]	5,976.930	98.38%	88.484 [15.739, 161.230]	0.009 [− 0.008, 0.025] (61)
RQ4	52	136	− 3.808 [− 12.829, 5.213]	578.095	76.65%	13.178 [7.667, 18.690]	0.002 [− 0.005, 0.009] (60)

Heterogeneity tests are conducted with random-effects models. Reported effect sizes are aggregated in means. The brackets contain the 95% confidence intervals of the respective statistic. Given the nature of the diverse choice of variables and measures, means provide an overview of collective results. They suggest that the focal subsidiary will have, on average, a chance of 4.4% to be "profitable" when it is a wholly owned venture (RQ1), a 1% increase in the ratio of R&D expenditure to total sales of a parent firm will lead to a 23.77% increase in its ownership in its subsidiaries (RQ2), and so forth. Cochran's Q is the weighted sum of squared differences between the effect sizes of individual analyses and the pooled effect sizes across analyses. All the Q values in this table have a p value that is less than 0.001 and reflect a high level of heterogeneity. I^2 is the percentage of variance across analyses that is due to heterogeneity rather than chance. I^2 values higher than 75% are usually considered to indicate high heterogeneity. Statistics for Bayes Factor Bound are generated using p values (Benjamin & Berger, 2019). The means of the partial correlation coefficients between focal independent and dependent variables are stated as a reference for a standardization of effect sizes (Fitzgerald, 2024; Stanley & Doucouliagos, 2012). The numbers of analyses used for this standardization are in parentheses.

database as reported in different studies. It is a conclusion drawn from aggregating across the results of analysts that made different empirical choices but tested the same data.

Parsing the variance in estimates: Do variable operationalizations matter?

Supplement 3 summarizes the frequency of different IV and DV operationalizations. We employed ANOVA to examine the relationship between analysts' choices regarding variable operationalizations and the variance in effect sizes. The choice of dependent variables had some consistency within each research question, with the same variables appearing in 78.6%, 91.2%, 89.7%, and 89.6% of the analyses for RQ1, RQ2, RQ3, and RQ4, respectively, albeit with some variations in coding. This homogeneity of approaches results in limited variation in effect sizes across research questions that is attributable to DV operationalizations, specifically 8.5%, 16.8%, 7.9%, and 0.1%, respectively, for RQs 1–4.

In contrast, choices of independent variables showed substantial heterogeneity (Supplement 3), and explained 74% and 38% of the variance in effect size estimates for RQ1 and RQ2, respectively (measured using R-squared). At the same time, IV choices had no major apparent impact on estimates for RQ3 and RQ4, explaining 8.7% and 4.2% of the variance, respectively. When considering unique pairings of IVs and DV, we observed that the IV-DV pairing can account for an impressive 73.8%, 51.4% and 38.9% of the variance in effect sizes for RQ1, RQ2, and RQ3, while explaining only limited variance in estimates for RQ4 (4.7%). Overall, variable operationalizations (either choice of IV, IV-DV pairing, or both) mattered for three of the four research questions.

As RQs 2–4 all involve the Japanese firms' equity ownership in their foreign subsidiaries, it becomes the most common dependent variable used. In an exploratory analysis, we constrained variable operationalizations to just this single DV for these three RQs, but this still resulted in estimates with wide confidence intervals, including zero for RQ3 and RQ4 (see Supplement 4, especially Figure S4-2). However, an exploratory analysis within IV-DV pairs for RQ2 did yield some exciting evidence of convergence (see Table S4-7 of Supplement 4). Specifically, 25% of the expert analysts for RQ2 chose the equity share owned by the Japanese parent firm as their DV and advertising expenses as the proxy for the intangible assets of a firm (the IV). This subset of specifications all returned individual estimates in a single direction that crossed the $p < 0.05$ threshold or effects close to zero, with no individual analyst reporting an estimate with a p value of less than 0.05 in the opposite direction. Thus, different small teams or solo science investigators all examining RQ2 using equity share and advertising expenses and employing the conventional $p < .05$ significance threshold would not have arrived at severely divergent conclusions. Both advertising expenses and R&D expenses are common proxies for intangible assets, and the choice of how to measure this IV is not especially theoretically rich. And yet, it does demonstrate that an encouraging convergence in results can emerge from a multitude of choices when one is highly specific about the empirical variables of interest (Auspurg & Brüderl, 2021).

Parsing the variance in estimates: Do analyst and research questions matter?

We observed no meaningful difference in the mean of effect sizes across the four research questions ($F = 0.81$, $p = 0.49$). Consequently, the research question failed to explain much variance in estimates for RQ1 (0.42%), RQ2 (0.07%), RQ3 (0.05%), or RQ4 (0.08%). In contrast, a substantial portion of the variance in effect sizes is attributable to analysts' specific approaches to the research question. Using R-squares, we calculate that 20.9% of the variance in effect sizes can be explained by analyst's unique preferences in modeling and variable choices. Notably, this is the mirror image of the outcome of a recent many designs initiative for behavioral experiments, which found that the conceptual hypothesis mattered far more than who created the study materials (Landy et al., 2020). More crowd initiatives are needed to examine when researchers matter most, which could differ across academic fields and methodologies.

Confirmation bias or rational belief updating?

Supplement 5 summarizes the distributions of pre-beliefs and post-beliefs. RQ2 was associated with the lowest variance in analyst pre- and post-beliefs among the four research questions (see Table S5-9 in Supplement 5).

We further ran linear regressions to examine the relationships between analysts' pre-beliefs and post-beliefs on the effect size estimates they obtained for the respective research question. The dependent variable is the estimated effect size. The independent variables are pre-beliefs and post-beliefs. When the analyst does not believe there is a relationship, it is coded as zero. If the analyst believes there is likely to be a positive/negative relationship, the IV is coded +1/−1. If the analyst believes there is very likely to be a positive/negative relationship, it is coded +2/−2.

Table 3 presents our analysis of pre- and post-beliefs for each research question as well as for all four research questions combined. For RQ1 and RQ4, neither pre-beliefs nor post-beliefs are correlated with effect size estimates. For RQ2, suggestive evidence emerged that both pre-beliefs and post-beliefs are related to effect size estimates. For RQ3, post-beliefs are related to effect size estimates, while pre-beliefs are not. In models 9 and 10, which include all four research questions, both pre-beliefs and post-beliefs

Table 3 Pre-beliefs and post-beliefs as correlates of effect size estimates

Research question	Model 1 RQ1	Model 2 RQ1	Model 3 RQ2	Model 4 RQ2	Model 5 RQ3	Model 6 RQ3	Model 7 RQ4	Model 8 RQ4	Model 9 All RQs	Model 10 All RQs
Independent variables										
Pre-beliefs	− 0.00 (0.01) [0.748]		24.55 (11.62) [0.039]		0.48 (0.76) [0.529]		4.86 (3.41) [0.160]		5.27 (2.37) [0.031]	
Post-beliefs		0.00 (0.04) [0.910]		18.46 (6.18) [0.004]		1.65 (0.60) [0.008]		5.64 (3.55) [0.119]		5.75 (1.86) [0.003]
Research question dummies	–	–	–	–	–	–	–	–	Included	Included
Constant	0.04 (0.02) [0.020]	0.05 (0.03) [0.182]	− 7.30 (10.62) [0.495]	4.50 (7.14) [0.532]	0.26 (1.07) [0.811]	0.71 (0.88) [0.423]	− 2.63 (3.78) [0.490]	− 2.03 (4.58) [0.659]	2.74 (1.42) [0.060]	3.62 (1.30) [0.007]
Number of observations	227	227	127	122	98	97	136	135	588	581
F value	0.10	0.01	4.47	8.91	0.40	7.52	2.03	2.52	2.15	4.00
R squared	0.0002	0.0001	0.0619	0.0578	0.0042	0.0812	0.0107	0.0123	0.0739	0.0823

The dependent variables in all regressions are the effect size estimates the focal analyst obtains. Standard errors are in parentheses. *P*-values are in brackets. Tests are two-tailed. Results are generated using Stata. Similar results are obtained using R (See Material 8 on the OSF).

Table 4 Regression results of analyst characteristics on estimates

Variables	Model 11	Model 12
DV	Effect size	*P*-value of the effect size
Regression Model	OLS	OLS
Related field	7.56 (3.22) [0.023]	0.11 (0.04) [0.007]
Research question dummies	Included	Included
Constant	− 6.55 (2.77) [0.022]	0.18 (0.04) [0.000]
Number of observations	588	580
F-value	2.43	23.05
R-square	0.0657	0.1035

Two-tailed tests. Standard errors are in parentheses. *P*-values are in brackets.

correlate positively with estimates overall. Based on this evidence, overall, we conclude that (1) pre-beliefs have a weak correlation with effect size estimates; and (2) post-beliefs have a stronger observed correlation with effect size estimates than pre-beliefs.

In addition, we used an ordered probit model to test the relationships between pre- and post-beliefs following Bürkner and Vuorre (2019) (see Tables S5-10 in Supplement 5). We find no reliable relationship between beliefs about the research hypotheses before analyzing the data and beliefs about the research hypotheses after analyzing the data. This suggests that analysts tended to change their beliefs over time, or at least that beliefs were unstable across these two time points. We further tested how estimates of effect sizes relate to post-beliefs (see Table S5-11 in Supplement 5) and found a positive correlation between the effect size estimates and post-beliefs for RQ2 and RQ3, where analysts appear to have been influenced by the results they drew from their analysis. Further details about the pre- and post-beliefs are provided in Supplement 6 and Materials 2 & 3 on OSF.

Does expertise matter?

Are expert analysts any more or less likely to obtain a given result? We run linear regressions to test which, if any, characteristics of analysts explain estimates related to effect sizes, using two distinct dependent variables. As shown in Table 4, the first dependent variable is effect size (Model 11), and the second is the *p* value of effect size (Model 12). The independent variable for both models is whether an analyst holds a doctoral degree in a related field (strategic management/international business) or not (another field). We also include research question dummies in the regressions.

The results show that analysts from a related field (strategic management and international business) are more likely to find larger effect sizes, and larger *p* values of effect size.

A related but distinct question is whether consistency in independently generated estimates is observed when selecting only experts. As analyzed in greater depth in Supplement 4, among our subsample of strategic management and international business scholars the range of estimates continues to cross zero and includes both positive and negative estimates for all four research questions. Further, topic experts show a high overall heterogeneity for RQ1–4 and conclude directional overall effects for RQ1 and RQ2 and yet not for RQ3 and RQ4. Analogous results emerge using the frequency of data analysis as our measure of expertise (Supplement 4). Thus, the results for experts are not qualitatively different from those for the crowd of analysts.

Next, expertise could plausibly moderate the relationship between pre-beliefs and/or post-beliefs and empirical estimates. For example, experts may exhibit less confirmation bias, or update their beliefs more carefully based on the evidence. However, as summarized in detail in Supplement 7, none of the interactions between the expertise variables and pre-/post-beliefs explained the variance in estimates.

Finally, regarding the use of different statistical software, an exploratory analysis revealed no differences in the mean of the reported effect sizes between the four analysts who used SPSS and the remaining 53 analysts who used Stata for their analyses (Material 16 on OSF).

Discussion

This crowd collaboration, the first of its kind in international business research, recruited 57 independent researchers from 41 institutions to attempt to answer four major, actively investigated research questions in IB. The results are informative about the focal research questions, as well as regarding the nature and processes of scientific inquiry more generally.

Dispersion of estimates

Our two clearest conclusions regarding researcher degrees of freedom are:

1. No two analysts employed the same specification for any research question.
2. The level of heterogeneity in empirical estimates was substantial.

For all four research questions, different analysts obtained estimates in opposite directions. This implies that a small science investigation relying on the most commonly used threshold for statistical significance ($p < 0.05$) could have, in principle, returned either a positive, negative, or non-effect and with varying effect sizes. Indeed, this observation aligns with the observed pattern of results for RQ1–4, across the many dozens of empirical studies reported in the IB and strategic management literatures, where there is no clear empirical consensus in results. Given this outcome, we must ask the question of whether empirical convergence for key research questions is likely across different small science investigations, where there is substantial, defensible latitude for researcher choices. We expand on this point later in this discussion.

Moreover, our results align with the dispersion pattern found in the majority of many-analysts and many-designs initiatives to date. The prominent exceptions of which we are aware are Silberzahn et al. (2018), which constrained variable operationalizations (e.g., only allowing red cards as the DV), and Hoogeveen et al. (2023), which relied on questionnaire responses on surveys rather than complex archival data. It seems reasonable to infer that less complex datasets and more specific hypotheses (e.g., “Are skin tone and red cards associated?” rather than “Is there any evidence of racial bias in referee decisions?”) are associated with a narrower spread of estimates (Auspurg & Brüderl, 2021; Breznau et al., 2022; Kummerfeld & Jones, 2023). Notably, however, quantitative research in international business typically relies on complex longitudinal archival data, and this research is oriented to identifying general principles of strategic decision making by organizations that are robust to how variables are measured.

Prior meta-scientific investigations reveal a surprising lack of evidence for the moderating effects of the expertise of contributors to crowd science. In citizen science initiatives which recruit members of the public to help with data collections, amateurs provide just as accurate observations (e.g., of butterfly migrations) as trained scientists (Kosmala et al., 2016; Thelen & Thiet, 2008). Highly cited and well-published researchers find the same results in direct replications of experiments as do early career researchers (Bench et al., 2017) and create novel study designs that return similar conclusions (Landy et al., 2020). The earliest many-analyst projects, such as Silberzahn et al. (2018), similarly obtained null effects of traditional indices of academic expertise and prestige.

In the present initiative, however, expertise did matter. Analysts with topic expertise in strategic management and international business were more likely to obtain large positive effects. This provides empirical support for the arguments of Auspurg and Brüderl (2021) that the variability in estimates in many-analyst projects can be due in part to analysts who have less expertise and utilize potentially suboptimal approaches. Menkveld et al. (2024) organized a large project that began to parse variance in estimates using individual differences at the analyst level and likewise observed a moderating effect of expertise. Yet, even when selecting only experts the variability in independent estimates remained very large, and the qualitative conclusions of both projects remained unchanged. This indicates that dispersion in estimates is not solely an artifact of the lesser expertise of some members of the crowd.

It is possible that an expertise effect was detected in the finance and international strategy many-analyst initiatives due to their comparatively larger samples than in Silberzahn et al. (2018) and Schweinsberg et al. (2021). Future investigations should recruit large samples of contributors across fields and ultimately meta-analyze to resolve whether expert scientists tend to obtain more (or less) extreme results, effects in a particular direction, and exhibit greater consensus with one another.

What these results mean for the four research questions

Two of the four initially posed research questions were associated with a directional conclusion based on aggregating across the different analyses employed by the crowd. Although many specifications were employed, these were carried out using only a single dataset, precluding strong inferences even in these two cases. For all four research questions, more analyses with further sets of independent observations are needed before drawing final conclusions. Below, we summarize what we believe the present dataset and crowd analyses have to say regarding each of the four specific research questions.

Research question 1: What is the relationship between entry mode and foreign subsidiary performance? Aggregating across the disparate analytical approaches, the overall effect indicates a directional relationship between entry mode and foreign subsidiary performance such that wholly owning a subsidiary is the best performing approach. This finding is consistent with the idea that ownership matters for foreign subsidiary performance (Chang et al., 2013; Lu & Beamish, 2001; Mata & Portugal, 2002). It is also in line with the predicted direction of the relationship from internationalization theory in terms of the ownership advantage of parent firms in running foreign subsidiaries (Delios & Beamish, 2001; Dunning, 2001). Yet, it should be emphasized that the aggregated estimate for Research Question 1 barely crosses the conventional $p < 0.05$ standard and would be considered only suggestive evidence by more conservative standards (Benjamin et al., 2018). This is yet another reason to carry out further investigations with new data sources before drawing strong inferences regarding RQ1.

Research question 2: What is the relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries? Based on the aggregation approach, greater levels of intangible assets do appear to lead parent firms to gain a higher percentage of equity ownership in their foreign subsidiaries. This result is anticipated in the literature on entry mode by multinational firms (Caves, 1971; Chang, 1995; Dunning, 1981; Wooster et al., 2016). In international joint ventures, multinational firms that contribute a high level of intangible assets tend to maintain advantageous ownership control positions (Ahuja et al., 2009). In the long run, they may even abandon cooperation and turn their joint ventures into wholly owned subsidiaries (Chang, 2019; Guillén, 2003).

Research question 3: What is the relationship between policy uncertainty and a firm's level of ownership in its foreign subsidiaries? The lack of an aggregate directional effect combined with a high heterogeneity in estimates leaves us without any answer regarding the relationship between policy uncertainty and parent firms' ownership in subsidiaries. This inconclusive result echoes the unresolved debate in the international business literature on this point. Organizations can use multiple methods to reduce external uncertainties, and one common way is to find partners to share the risks and create complementarity to enhance competitive advantages (Delios & Henisz, 2003; Kogut & Singh, 1988; Zhao et al., 2004). On the other hand, firms can accumulate knowledge to overcome external uncertainty (Kogut & Chang, 1996). Also, relaxing ownership control raises the issue of opportunism by partners (Luo, 2007; Madhok, 2006). Dhanaraj and Beamish (2004) provide evidence that equity ownership and the survival of foreign subsidiaries follow a nonlinear and declining curve. Firms need to design the ownership structure to manage the balance between help from partners and the risks from the environment.

Research question 4: How does the level of policy uncertainty moderate the relationship between intangible assets and a firm's level of ownership in its foreign subsidiaries? Similar to RQ3, there is no clear evidence to support the hypothesized moderating effect of policy uncertainty on the relationship between intangible assets and parent firms' ownership in foreign subsidiaries. Notably, since RQ4 is the only research question that involves a moderating effect, analysts have a particularly high degree of freedom to choose variables and model specifications ($I^2 = 100\%$). The interplay between uncertainty, intangible assets, and subsidiary ownership structure is complex. For example, firms may have to choose between focusing resources in one direction or spreading them to maintain flexibility (Wernerfelt & Karnani, 1987). Additionally, uncertainty can motivate firms to mimic the practices of other firms (Ang et al., 2015; Henisz & Delios, 2001). These possible contextual factors may contribute to the present inclusive results, which should be unpacked in future investigations, augmenting the mixed evidence on this topic to date (Delios & Henisz, 2000; Henisz, 2000b).

The dynamic interplay between beliefs and evidence

Beliefs about the research question measured at the end of the study were more strongly related to empirical estimates than beliefs measured before receiving the data. This is a similar pattern to that previously captured in Silberzahn et al. (2018), but with more units of observation, lending the results further credibility. It seems that scientists rationally updated their beliefs considering the evidence. We again find no apparent confirmation bias, such that researchers' preconceptions strongly predict their analytic choices and results. There may be little incentive to *p*-hack in many-analyst projects, since each individual estimate has a limited bearing on the publication potential of the paper or conclusions drawn. The transparency and observation by others may also curtail bias, or crowd collaboration may select researchers who are more open to different conclusions from the outset. Regardless, the dispersion in estimates seen here for all four research questions is more likely attributable to good-faith but subjective choices than it is attributable to a strategic exploitation of researcher degrees of freedom.

A meta-analysis will be needed to draw strong inferences about the relationship(s) between beliefs and evidence in crowd science collaborations. To this end, it should be standard to measure subjective beliefs about the research question before, after, and for multi-stage projects in the middle of the process as well. Another lingering question, best resolved through meta-analytic aggregation across projects, is the extent to which peer discussion leads to convergence in empirical estimates and subjective beliefs (Menkveld et al., 2024; Silberzahn et al., 2018).

Implications for international business research and other areas of inquiry

To date, the accumulated evidence from many-analysts and many-designs projects point to a challenge of subjectivity in science. Even with no incentive to produce a publishable finding, absent any apparent confirmation bias, and selecting only experts, we observe considerable variability in results when independent scientists and small teams attempt to address the same research question. Like many other disciplines, quantitative research in international business often relies on complex longitudinal datasets with numerous choice points; any such field is vulnerable to such subjectivity. Our project did not capture a preceding universe of choices in the building and cleaning of the dataset, and for this reason, may even underestimate the impact of researcher judgment calls.

We are skeptical that the current academic publication system, in which some reviewers may request further robustness tests from the authors, is sufficient to ensure the analytic robustness of research. Indeed, the review process is arguably also a contributor to the problem, since it provides authors an incentive to try many specifications and choose the subset (including robustness tests) that returns results that increase the chances of a favorable decision from editors and reviewers at prestigious journals (for empirical evidence, see Campbell et al., 2024). This usually means a $p < 0.05$ result in support of theories, the authors and/or the majority of scholars in the field find appealing. Notably, Delios et al. (2022) recently found that more than half of a set of published research results on international business could not be reproduced using the *same* analyses and observations, highlighting the shortcomings of the business-as-unusual approach to small science research with closed data and limited guardrails other than peer review. An ongoing project from our group is currently carrying out multiverse analyses for 28 published international business findings, with the preliminary results suggesting that this peer-reviewed work is typically not robust to alternative specifications.

At the same time, we must avoid what the Roman poet Virgil referred to as the *aegrescit medendo* phenomenon, meaning "it worsens with healing." Open science innovations that seem like improvements in theory may not be an improvement in practice if the innovation unduly burdens researchers or otherwise hinders the accumulation of knowledge in the field. In the modal situation in which organizing a crowd collaboration is excessively inefficient and impractical, international business scholars with access to sufficient computational resources could consider conducting a multiverse (Sala-i-Martin, 1997; Steegen et al., 2016) or specification curve (Simonsohn et al., 2020) in which a single researcher or small team carries out numerous preplanned analyses.

When computational resources are limited, scholars could specify an expanded suite of robustness checks and aggregate across them, with the caveat that, as with our project, their multiple alternative analyses were carried out using a single dataset, limiting generalizability. At the same time, we suggest that the expectation that each and every robustness test must return strong support for the key theoretical prediction should be relaxed (Sala-i-Martin, 1997): a result that is observed reliably aggregating across defensible specifications is sufficient to make a meaningful scientific claim, albeit not to draw final conclusions. Our results regarding expertise suggest there could be an elevated payoff from convening a select crowd of experts to carry out a handful of independent high-quality analyses (Mannes et al., 2014). This would be more efficient, and could lead to less noise, than open calls that assemble huge crowds of analysts with diverse profiles and backgrounds.

At a theoretical level, some quantitative investigations may benefit from clearly defined research questions that narrow the space of viable empirical approaches (Auspurg &

Brüderl, 2021; Breznau et al., 2022; Kummerfeld & Jones, 2023; Scheel, 2022; Schweinsberg et al., 2023). Such upstream theoretical choice points could be especially important contributors to effect size dispersion. A thorough understanding of the latitude of construal around a research question, as well as each independent investigator's interpretations and decisions within that space (Goldfarb & Yan, 2021; King et al., 2021), could be critical to cracking the parsing problem.

Importantly, at the same time, we consider theoretical choice points part of the phenomenon of the subjectivity in science that we seek to understand to enable better science, not only in international business research but in the social sciences more generally. Many scientific questions of great academic and public interest are by their nature broad and general (e.g., "Is there evidence of racial bias in top soccer leagues?" or "How do local partnerships affect the performance of global firms?"). In such cases, narrowing the theory to a specific independent variable and dependent variable, when many IVs and DVs are in fact relevant to testing the key claims, risks creating an illusion of consensus and certainty regarding the evidence. When many variable operationalizations and empirical approaches provide fair tests of a general theory (e.g., the value of strategic partnerships), they should all be carried out and the dispersion in estimates rendered transparent.

First try parsing, then consider aggregation

Aggregating across the disparate results reveals an overall effect for 2 of 4 research questions (RQ1 and RQ2), although the confidence interval almost includes zero for RQ1, suggesting caution in concluding a reliable directional effect. For the two remaining research questions (RQ3 and RQ4), aggregation suggests no directional conclusion.

With regards to parsing the variability by identifying meaningful moderators, we find that theoretically laden choices regarding how to operationalize variables explained meaningful variability in estimates (see also Schweinsberg et al., 2021). For 3 of 4 research questions, either the choice of the IV or the choice of the IV-DV pairing mattered a great deal. For RQ2, two of the three most common IV-DV pairs yielded consistent results. And yet, similar exploratory analyses examining the results within specific IV and DV operationalizations still failed to yield any directional answer to RQ3 and RQ4.

To date, meta-scientists have devoted innumerable person-hours to testing possible moderators of effect size dispersion, yielding hard-earned, yet modest rewards. Up until now, many-analysts and many-designs initiatives have explained only small slices of the variance despite sometimes multi-year efforts involving hosts of variables. The present comparative success with IV-DV pairs indicates that it is still too soon to give up on the perspectivist dream (McGuire, 1983) of uncovering intellectual spaces populated with theoretically rich moderators. We suggest that future investigations first attempt parsing, and if the variance explained remains modest, only then turn to aggregating across the diverse approaches and estimates in an effort to reach an initial conclusion.

Importantly, those who conduct and review and evaluate quantitative research should also consider that in some cases, the heterogeneity in approaches will be too high to aggregate. In such cases, parsing is the only viable route forward (Berchicci & King, 2022), for example, via epistemic maps of researcher theoretical assumptions, related analytic decisions, and their consequences for the reported results (Goldfarb & Yan, 2021; King et al., 2021).

One potential path forward may be to engage in formal qualitative analyses of quantitative statistical decisions (e.g., Koehler et al., 2020). In one step in this direction, Schweinsberg et al. (2021) used an online tool called Data Explained to track their crowd of analysts' paths taken and not taken as well as their justifications for their decisions. A team of coders then engaged in extensive qualitative coding of their workflow and reasoning. This inductive coding led to a model of the sensemaking process of data analysis in which scientists engage in iterative loops between their beliefs and knowledge, the data and empirical setting, their statistical choices and estimates, and new insights (see Figure 6 of Schweinsberg et al., 2021).

A recent study in international business reexamined the qualitative cases from previous studies to discover new findings (Rumyantseva & Welch, 2023). A similar qualitative approach might be applied to past quantitative decisions, coupling this with re-analyses exploring paths not taken in the original paper. In an integrated mixed-methods approach, qualitative interviews with experts could potentially be used to adjudicate between alternative viable approaches. For example, senior IB scholars and/or leaders of multinational companies could be consulted to help identify the ideal means of operationalizing key variables. This seemingly reasonable approach to parsing the variance in estimates may run into challenges if the experts disagree with each other on the "right" approach or qualitative coders fail to align regarding what the experts have to say, adding additional layers of subjectivity. Indeed, we expect that a "many coders" initiative would find that when expert qualitative researchers independently code the same text corpus they tend to draw diverging conclusions, similar to what is observed in many-analysts initiatives in quantitative research (Armstrong et al., 1997). Thus, the subjective aspects of the scientific process may hinder efforts to identify objectively superior means of answering a given research question.

Potential countermeasures

Addressing both strategic *p*-hacking and inherent subjectivity in data analysis is critical to ensuring the reliability and robustness of our science. Although both involve an uncertain garden of forking paths (Gelman & Loken, 2014) through a scientific inquiry, we believe the most effective countermeasures are different. In Table 5, we offer a list and typology of interventions and classify them as effective against either strategic or inherent elasticity in researcher decision making. Several interventions are in principle effective at curtailing *p*-hacking, in particular pre-registration of analyses (Wagenmakers et al., 2012), creating a test and holdout sample (Altman, 1968) and blind analyses (MacCoun & Perlmutter, 2015). In the complex archival data commonly used in international business studies, data must often be explored as part of the process of understanding, rendering pre-registration and blind analyses potentially over-restrictive (King et al., 2021). Thus, a test-holdout approach could be comparatively more useful in quantitative research using complex datasets, to allow exploration while at the same time preventing *p*-hacking (Goldfarb & Yan, 2021).

Although inherent elasticity is arguably the greater challenge for science, it is not interpersonally inflammatory like allegations of *p*-hacking since the former does not call the ethical integrity of the scientist into question in any way. Demonstrating that subareas of the scholarly literature are characterized by publication bias, *p*-hacking by some research teams, and, in some cases, a lack of evidentiary value helps other scientists identify the most reliable bodies of knowledge upon which to build (Egger et al., 1997; Goldfarb & King, 2016; Simonsohn et al., 2014; Stanley, 2005). At the same time, accusations of strategic analyses against individual papers and research teams can be problematic and unfair—if we run publication bias or *p*-hacking tests on enough individual articles, inevitably some will be flagged for containing problematic data, even if none of them do (Simonsohn, 2013).

Moving forward, it is also worth considering our collective interest in maximizing the credibility of what appears in the published literature. We therefore suggest that with regard to individual future investigations, it is typically better to curtail strategic elasticity before the fact than attempt to expose it after the fact. Proposed new practices, such as pre-registration and test-holdout also have shortcomings, such as reduced efficiency or the requirement for very large samples and thus should be deployed on a case-by-case basis. The ultimate objective should be an improved research ecosystem, not the unattainable goal of scientific perfection.

In contrast, inherent elasticity, such as that revealed here, may be inevitable in scientific inquiry. Different scientists, acting entirely in good faith, will plan out different approaches in advance, analyze data blindly in different ways, and explore different paths through a test sample. The best available options are to render inherent elasticity transparent through multiverse and crowd analyses, expanded robustness tests, and the open posting of data for re-analysis by colleagues. Open data sharing should be encouraged by international business journals on publication, although this, of course, is not possible for datasets entailing confidentiality concerns and proprietary datasets obtained through agreements with partner firms and data providers. To the extent that data can be made open, the community can collectively probe the robustness of findings using alternative specifications (Goldfarb & Yan, 2021) and work together to identify issues post-publication.

We therefore present a vision of an open science of international business (Meyer et al., 2020) in which interventions are considered to prevent *p*-hacking before it happens when this is a significant concern, a single paper reports more analytic approaches and robustness tests than is currently the norm (Sala-i-Martin, 1997; Steegen et al., 2016), and the data and code are made publicly available to the community when possible. When faced with a wide dispersion in estimates across different specifications, international business scholars should seek to anticipate and detect theoretically informative moderators, and failing this, consider aggregation to reach a tentative conclusion. The principles of open science further encourage researchers to replicate analyses across different sets of observations, especially alternative time periods and countries. This helps develop a norm of examining the generalizability versus the context sensitivity of research findings (Delios et al., 2022). The approaches we advocate have advantages and disadvantages, as do traditional practices. However, open science leads to more transparent results and accurate inferences (Nosek et al., 2022) on behalf of the stakeholders we serve.

Testing hypotheses in many ways is considerably more work than social scientists are used to doing to make a single claim. Multiverse analyses, crowd analyses, aggregating across numerous robustness tests, and assessments of generalizability across time and geographies represent departures from standard scientific practices and incentive structures. However, a traditional small science paper reporting a single primary analysis of the results of one research design should not be solely relied on to make strong claims or determine policy, even in the absence of any researcher bias towards a desired conclusion. The research community will need to incorporate more reformed practices if we wish to draw strong inferences (Platt, 1964).

Conclusions

The major contributions of this first crowd science initiative in the field of international business are multifaceted and interrelated. We address four important unresolved

Table 5 Approaches to preventing and detecting strategic and inherent elasticity in analytic procedures

Approach	Description	Aimed at inherent or strategic elasticity	Prevents or detects
Test-holdout sample approach	The dataset is split into two pieces, with exploratory analyses conducted on the first piece and confirmatory analyses on the second piece (e.g., Altman, 1968; Orrù et al., 2020; Shrestha et al., 2021)	Strategic	Prevents
Pre-registration	Researcher specifies the analytic plan before collecting or obtaining the data (Wagenmakers et al., 2012)	Strategic	Prevents
Data-blind analysis	Original researcher relabels or recodes some of the key variables before conducting her analysis, such that she is unaware of the direction and nature of the observed relationships (MacCoun & Perlmutter, 2015)	Strategic	Prevents
Independent re-analysis	Independent researcher uses an alternative analytic approach to test the same theoretical idea (e.g., Simonsohn, 2011)	Both	Detects
Error analysis	Independent researcher identifies errors in data handling or analysis, and may show that these errors systematically favor the original hypothesis (Rosenthal, 1978)	Both	Detects
Tests for publication bias and questionable research practices	Meta-scientific analysis of a study set or individual study to test for omitted studies and data patterns consistent with questionable research practices (Ioannidis & Trikalinos, 2007; Simonsohn et al., 2014)	Strategic	Detects
"Chrysalis" analysis	Comparison of earlier vs. later versions of research reports (e.g., doctoral dissertations vs. published reports of the same work) (O'Boyle et al., 2017)	Strategic	Detects
Conducting an independent replication study	Independent research team collects new data using the same methodology and runs the same analyses as described in the original paper (e.g., Open Science Collaboration, 2015). *P*-hacked original findings are less likely to replicate, but failed replications can occur for many other reasons as well	Strategic	Detects
Red team approach	Independent research team attempts to debias and optimize an original investigation before it is submitted for publication (Lakens, 2020)	Strategic	Prevents
Retrospective re-assessment	Original researcher decides after-the-fact that the analysis was sub-optimal and a different approach would have been better (Rohrer et al., 2021)	Both	Detects
Comparing employed analyses	A single researcher surveys the various analytic approaches used in different studies on the same topic, identifies inconsistencies within and across research groups, and potentially applies all approaches to the same data (e.g., Elson et al., 2014; Harris et al., 2013)	Strategic	Detects
Strategic analysis	A crowd of analysts is asked to either produce statistically significant support for the original hypothesis (directional goal) or provided with no goal when analyzing the data (ongoing crowd initiative via the University of Pennsylvania's Adversarial Collaboration Project)	Strategic	Detects
Multiverse analysis or specification curve	One analyst employs numerous specifications (Simonsohn et al., 2020; Steegen et al., 2016)	Inherent	Detects
Many analysts	Many researchers analyze the same data to test the same hypothesis (Silberzahn et al., 2018)	Inherent	Detects

A longer version of this table is available in Supplement 9 questions in the field by leveraging the insights of 57 independent scholars. The insights collectively suggest a directional answer to two research questions, each the subject of numerous prior investigations across decades of study that only yielded mixed and conflicting results.

At the same time, we take stock of the small but growing literature on the many-analysts phenomenon (Silberzahn et al., 2018). Choice points matter in research using complex archival data, but we do not yet fully understand the reasons for and consequences of choice point effects. The present empirical results indicate that heterogeneity in naturally emerging approaches and resulting estimates poses a significant challenge for management scholarship (Ketokivi & Mantere, 2010; King et al., 2021; Mantere & Ketokivi, 2013). In terms of parsing this variance in estimates, we report the strongest evidence yet that variable operationalizations and researcher expertise matter. At the same time, we explore cross-question differences, finding that a high theoretical consensus regarding what "should" happen is no guarantee of empirical consistency in estimates. Encouragingly, however, where theory is strong, as for the direction of assets and ownership hypotheses, aggregation reveals an overall estimate consistent with theory, pending further evidence. Likewise, supporting the idea of a community building knowledge together, scholars updated their beliefs considering the empirical evidence rather than allowing their prior convictions to bias their analyses and conclusions.

To further strengthen this community, we advocate integrating the knowledge gains from meta-scientific crowd projects like this one into the everyday practices of traditional small teams of researchers. To facilitate open science practices in international business studies, we present a compendium of interventions against analytic elasticity (see Table 5) and a novel typology based on whether they target strategic or inherent elasticity and are prevention or detection oriented.

Although our present crowd initiative demonstrates the critical role of subjective researcher choices in shaping results in international business research, it also underscores that meaningful inferences remain possible in science. By sampling analyses and stimuli widely, and leveraging the insights derived from a diversity of approaches, we can make real progress on the key scientific questions of interest to a field.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1057/s41267-025-00808-9.

Acknowledgements This project was supported by the Multi-Year Research Grant from the University of Macau (Reference No.: MYRG-GRG2024-00140-FBA), awarded to Tianyou Hu, and the National Natural Science Foundation of China (Project No.: 72572121, 72122016), awarded to Nan Zhou. Eric Uhlmann is grateful for funding from the INSEAD R&D committee. We acknowledge the following analysts as co-authors, with whom we have unfortunately lost contact: Liang Hu (Jinan University), Mahdi Forghani Bajestani (Old Dominion University), Aqi Liu (The Chinese University of Hong Kong), Fatemeh Askarzadeh (Old Dominion University), Benjamin Krebs (Paderborn University), and Jiao Li (China Europe International Business School). We are grateful to research assistance from Jimmy Chen, Yuehan Chu, Nihong Li, Yixiao Han, Yize Liu, Rafael Goldszmidt, Kerui Gao, Lixuan Xu, and Zhiyi Wu for their material organization, data preparation, and administrative assistance.

Data Availability Statement Our methods and analyses were preregistered at https://osf.io/4euaj. The supplementary document of this paper is accessible at https://doi.org/10.1057/s41267-025-00808-9. Further Results and Materials are posted on the Open Science Framework at https://osf.io/ew3vz/?view_only=52ab5518ada34e0ca1046aeab08a5122.

References

Aguinis, H., Archibold, E. E., & Rice, D. B. (2022). Let's fix our own problem: Quelling the irresponsible research perfect storm. *Journal of Management Studies, 59*(6), 1628–1642. https://doi.org/10.1111/joms.12814

Aguinis, H., Banks, G. C., Rogelberg, S. G., & Cascio, W. F. (2020). Actionable recommendations for narrowing the science-practice gap in open science. *Organizational Behavior and Human Decision Processes, 158*, 27–35. https://doi.org/10.1016/j.obhdp.2020.02.007

Aguinis, H., Bergh, D., & Molina-Azorin, J. F. (2023). Methodological challenges and insights for future international business research. *Journal of International Business Studies, 54*(2), 219–232. https://doi.org/10.1057/s41267-022-00578-8

Aguinis, H., Cascio, W. F., & Ramani, R. S. (2017). Science's reproducibility and replicability crisis: International business is not immune. *Journal of International Business Studies, 48*(6), 653–663. https://doi.org/10.1057/s41267-017-0081-0

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. https://doi.org/10.1002/smj.774

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609. https://doi.org/10.2307/2978933

Ang, S. H., Benischke, M. H., & Doh, J. P. (2015). The interactions of institutions on foreign market entry mode. *Strategic Management Journal, 36*(10), 1536–1553. https://doi.org/10.1002/smj.2295

Armstrong, D., Gosling, A., Weinman, J., & Marteau, T. (1997). The place of inter-rater reliability in qualitative research: An empirical study. *Sociology, 31*(3), 597–606. https://doi.org/10.1177/0038038597031003015

Auspurg, K., & Brüderl, J. (2021). Has the credibility of the social sciences been credibly destroyed? Reanalyzing the "Many Analysts, One Data Set" project. *Socius, 7*, 1–14. https://doi.org/10.1177/23780231211024421

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences of the United States of America, 115*(11), 2607–2612. https://doi.org/10.1073/pnas.1708285114

Bench, S. W., Rivera, G. N., Schlegel, R. J., Hicks, J. A., & Lench, H. C. (2017). Does expertise matter in replication? An examination of the Reproducibility Project: Psychology. *Journal of Experimental Social Psychology, 68*, 181–184. https://doi.org/10.1016/j.jesp.2016.07.003

Benjamin, D. J., & Berger, J. O. (2019). Three recommendations for improving the use of *p*-values. *The American Statistician, 73*(sup1), 186–191. https://doi.org/10.1080/00031305.2018.1543135

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E.-J., Berk, R., Bollen, K. A., Brembs, B., Brown, L., Camerer, C., et al. (2018). Redefine statistical significance. *Nature Human Behaviour, 2*(1), 6–10. https://doi.org/10.1038/s41562-017-0189-z

Berchicci, L., & King, A. A. (2022). Building knowledge by mapping model uncertainty in six studies of social and financial performance. *Strategic Management Journal, 43*(7), 1319–1346. https://doi.org/10.1002/smj.3374

Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (2017). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization, 15*(3), 423–436. https://doi.org/10.1177/1476127017701076

Berman, J. S., & Reich, C. M. (2010). Investigator allegiance and the evaluation of psychotherapy outcome research. *European Journal of Psychotherapy & Counselling, 12*(1), 11–21. https://doi.org/10.1080/13642531003637775

Bosco, F. A., Aguinis, H., Field, J. G., Pierce, C. A., & Dalton, D. R. (2016). Harking's threat to organizational research: Evidence from primary and meta-analytic sources. *Personnel Psychology, 69*(3), 709–750. https://doi.org/10.1111/peps.12111

Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature, 582*(7810), 84–88. https://doi.org/10.1038/s41586-020-2314-9

Breznau, N., Rinke, E. M., Wuttke, A., Nguyen, H. H. V., Adem, M., Adriaans, J., et al. (2022). Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences of the United States of America, 119*(44), e2203150119. https://doi.org/10.1073/pnas.2203150119

Brouthers, K. D. (2002). Institutional, cultural and transaction cost influences on entry mode choice and performance. *Journal of International Business Studies, 33*(2), 203–221. https://doi.org/10.1057/palgrave.jibs.8491013

Brouthers, K. D., Brouthers, L. E., & Werner, S. (2003). Transaction cost-enhanced entry mode choices and firm performance. *Strategic Management Journal, 24*(12), 1239–1248. https://doi.org/10.1002/smj.362

Buckley, P. J., & Casson, M. (1976). A long-run theory of the multinational enterprise. In P. J. Buckley & M. Casson (Eds.), *The future of the multinational enterprise* (pp. 32–65). London: Palgrave Macmillan. https://doi.org/10.1007/978-1-349-02899-3_2

Bürkner, P.-C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. *Advances in Methods and Practices in Psychological Science, 2*(1), 77–101. https://doi.org/10.1177/2515245918823199

Campbell, D., Brodeur, A., Dreber, A., Johannesson, M., Kopecky, J., Lusher, L., & Tsoy, N. (2024). *The robustness reproducibility of the American Economic Review*. Working Paper, I4R Discussion Paper Series, No. 124, Institute for Replication (I4R), s.l. Available at https://www.econstor.eu/handle/10419/295222.

Caves, R. E. (1971). International corporations: The industrial economics of foreign investment. *Economica, 38*(149), 1–27. https://doi.org/10.1016/j.cortex.2015.03.022

Chang, S.-J. (1995). International expansion strategy of Japanese firms: Capability building through sequential entry. *Academy of Management Journal, 38*(2), 383–407. https://doi.org/10.5465/256685

Chang, S.-J. (2019). When to go it alone: Examining post-conversion performance of international joint ventures. *Journal of International Business Studies, 50*(6), 998–1020. https://doi.org/10.1057/s41267-018-00211-7

Chang, S.-J., Chung, J., & Moon, J. J. (2013). When do wholly owned subsidiaries perform better than joint ventures? *Strategic Management Journal, 34*(3), 317–337. https://doi.org/10.1002/smj.2016

Chen, H., & Hu, M. Y. (2002). An analysis of determinants of entry mode and its impact on performance. *International Business Review, 11*(2), 193–210. https://doi.org/10.1016/S0969-5931(01)00055-5

Cyrus-Lai, W., Tierney, W., Schweinsberg, M., & Uhlmann, E. L. (2022). Exposing and overcoming the fixed-effect fallacy through crowd science. *Behavioral and Brain Science, 45*, 1–8. https://doi.org/10.1017/s0140525x21000297

Delios, A., & Beamish, P. W. (2001). Survival and profitability: The roles of experience and intangible assets in foreign subsidiary performance. *Academy of Management Journal, 44*(5), 1028–1038. https://doi.org/10.5465/3069446

Delios, A., Clemente, E. G., Wu, T., Tan, H., Wang, Y., Gordon, M., et al. (2022). Examining the generalizability of research findings from archival data. *Proceedings of the National Academy of Sciences of the United States of America, 119*(30), e2120377119. https://doi.org/10.1073/pnas.2120377119

Delios, A., & Henisz, W. J. (2000). Japanese firms' investment strategies in emerging economies. *Academy of Management Journal, 43*(3), 305–323. https://doi.org/10.5465/1556397

Delios, A., & Henisz, W. J. (2003). Policy uncertainty and the sequence of entry by Japanese firms, 1980–1998. *Journal of International Business Studies, 34*(3), 227–241. https://doi.org/10.1057/palgrave.jibs.8400031

Dhanaraj, C., & Beamish, P. W. (2004). Effect of equity ownership on the survival of international joint ventures. *Strategic Management Journal, 25*(3), 295–305. https://doi.org/10.1002/smj.372

Dunning, J. H. (1981). Explaining the international direct investment position of countries: Towards a dynamic or developmental approach. *Weltwirtschaftliches Archiv, 117*(1), 30–64. https://doi.org/10.1007/BF02696577

Dunning, J. H. (2001). The eclectic (OLI) paradigm of international production: Past, present and future. *International Journal of the Economics of Business, 8*(2), 173–190. https://doi.org/10.1080/13571510110051441

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the competitive reaction time task in aggression research. *Psychological Assessment, 26*(2), 419–432. https://doi.org/10.1037/a0035569

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE, 5*(4), e10068. https://doi.org/10.1371/journal.pone.0010068

Fey, C. F., Hu, T., & Delios, A. (2023). The measurement and communication of effect sizes in management research. *Management and Organization Review, 19*(1), 176–197.

Fitzgerald, J. (2024). *The need for equivalence testing in economics.* Working Paper, I4R Discussion Paper Series No. 125. Institute for Replication (I4R), s.l. Retrieved from https://hdl.handle.net/10419/296190.

Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis–a 'garden of forking paths'–explains why many statistically significant comparisons don't hold up. *American Scientist, 102*(6), 460–465.

Goldfarb, B., & King, A. A. (2016). Scientific apophenia in strategic management research: Significance tests & mistaken inference. *Strategic Management Journal, 37*(1), 167–176. https://doi.org/10.1002/smj.2459

Goldfarb, B., & Yan, L. (2021). Revisiting Zuckerman's (1999) categorical imperative: An application of epistemic maps for replication. *Strategic Management Journal, 42*(11), 1963–1992. https://doi.org/10.1002/smj.3290

Guillén, M. F. (2003). Experience, imitation, and the sequence of foreign entry: Wholly owned and joint-venture manufacturing by South Korean firms and business groups in China, 1987–1995. *Journal of International Business Studies, 34*(2), 185–198. https://doi.org/10.1057/palgrave.jibs.8400016

Harris, C. R., Chabot, A., & Mickes, L. (2013). Shifts in methodology and theory in menstrual cycle research on attraction. *Sex Roles, 69*(9), 525–535. https://doi.org/10.1007/s11199-013-0302-3

Henisz, W. J. (2000a). The institutional environment for economic growth. *Economics and Politics, 12*(1), 1–31. https://doi.org/10.1111/1468-0343.00066

Henisz, W. J. (2000b). The institutional environment for multinational investment. *Journal of Law, Economics & Organization, 16*(2), 334–364. https://doi.org/10.1093/jleo/16.2.334

Henisz, W. J. (2003). The power of the Buckley and Casson thesis: The ability to manage institutional idiosyncrasies. *Journal of International Business Studies, 34*(2), 173–184. https://doi.org/10.1057/palgrave.jibs.8400015

Henisz, W. J., & Delios, A. (2001). Uncertainty, imitation, and plant location: Japanese multinational corporations, 1990–1996. *Administrative Science Quarterly, 46*(3), 443–475. https://doi.org/10.2307/3094871

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A conceptual introduction to Bayesian model averaging. *Advances in Methods and Practices in Psychological Science, 3*(2), 200–215. https://doi.org/10.1177/2515245919898657

Hoogeveen, S., Sarafoglou, A., Aczel, B., Aditya, Y., Alayan, A. J., Allen, P. J., et al. (2023). A many-analysts approach to the relation between religiosity and well-being. *Religion, Brain & Behavior, 13*(3), 237–283. https://doi.org/10.1080/2153599X.2022.2070255

Huber, C., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Weitzel, U., et al. (2023). Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs. *Proceedings of the National Academy of Sciences of the United States of America, 120*(23), e2215572120. https://doi.org/10.1073/pnas.2215572120

Ioannidis, J. P. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), 696–701. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Journal of Clinical Trials, 4*(3), 245–253. https://doi.org/10.1177/1740774507079441

Isager, P. M., van Aert, R. C. M., Bahník, Š, Brandt, M. J., DeSoto, K. A., Giner-Sorolla, R., et al. (2021). Deciding what to replicate: A decision model for replication study selection under resource and knowledge constraints. *Psychological Methods, 28*(2), 438–451.

Ivey. (2015). *Japanese foreign direct investment publications* (Vol. 34, pp. 910–934). Retrieved March 5, 2015, from https://www.ivey.uwo.ca/media/1401423/jfdi-abstracts-march-2015.pdf.

Johanson, J., & Vahlne, J.-E. (1977). The internationalization process of the firm—a model of knowledge development and increasing foreign market commitments. *Journal of International Business Studies, 8*(1), 23–32. https://doi.org/10.1057/palgrave.jibs.8490676

Ketokivi, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. *Academy of Management Review, 35*(2), 315–333.

King, A., Goldfarb, B., & Simcoe, T. (2021). Learning from testimony on quantitative research in management. *Academy of Management Review, 46*(3), 465–488. https://doi.org/10.5465/amr.2018.0421

Koehler, T., Cassell, C., Langley, A., Rumyantseva, M., Smith, A., & Welch, C. (2020). Perspectives on replication in qualitative research. *Academy of Management Proceedings*. https://doi.org/10.5465/AMBPP.2020.18615symposium

Kogut, B., & Chang, S. J. (1996). Platform investments and volatile exchange rates: Direct investment in the U.S. by Japanese electronic companies. *The Review of Economics and Statistics, 78*(2), 221–231.

Kogut, B., & Singh, H. (1988). The effect of national culture on the choice of entry mode. *Journal of International Business Studies, 19*(3), 411–432. https://doi.org/10.1057/palgrave.jibs.8490394

Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment, 14*(10), 551–560. https://doi.org/10.1002/fee.1436

Kummerfeld, E., & Jones, G. L. (2023). One data set, many analysts: Implications for practicing scientists. *Frontiers in Psychology, 14*, 1094150. https://doi.org/10.3389/fpsyg.2023.1094150

Lakens, D. (2020). Pandemic researchers—Recruit your own best critics. *Nature, 581*(7807), 121–122. https://doi.org/10.1038/d41586-020-01392-8

Landy, J. F., Jia, M., Ding, I. L., Viganola, D., Tierney, W., Dreber, A., et al. (2020). Crowdsourcing hypothesis tests: Making transparent how design choices shape research results. *Psychological Bulletin, 146*(5), 451–479. https://doi.org/10.1037/bul0000220

Lee, K. H. (2018). Cross-border mergers and acquisitions amid political uncertainty: A bargaining perspective. *Strategic Management Journal, 39*(11), 2992–3005. https://doi.org/10.1002/smj.2944

Lu, J. W., & Beamish, P. W. (2001). The internationalization and performance of SMEs. *Strategic Management Journal, 22*(6–7), 565–586. https://doi.org/10.1002/smj.184

Luo, Y. (2007). Are joint venture partners more opportunistic in a more volatile environment? *Strategic Management Journal, 28*(1), 39–60. https://doi.org/10.1002/smj.564

MacCoun, R., & Perlmutter, S. (2015). Blind analysis: Hide results to seek the truth. *Nature, 526*(7572), 187–189. https://doi.org/10.1038/526187a

Madhok, A. (2006). Revisiting multinational firms' tolerance for joint ventures: A trust-based approach. *Journal of International Business Studies, 37*(1), 30–43.

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology, 107*(2), 276–299. https://doi.org/10.1037/a0036677

Mantere, S., & Ketokivi, M. (2013). Reasoning in organization science. *Academy of Management Review, 38*(1), 70–89.

Martin, X., & Salomon, R. (2003). Knowledge transfer capacity and its implications for the theory of the multinational corporation. *Journal of International Business Studies, 34*(4), 356–373. https://doi.org/10.1057/palgrave.jibs.8400037

Mata, J., & Portugal, P. (2002). The survival of new domestic and foreign-owned firms. *Strategic Management Journal, 23*(4), 323–343. https://doi.org/10.1002/smj.217

McGuire, W. J. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 16, pp. 1–47). Academic Press. https://doi.org/10.1016/S0065-2601(08)60393-7

Menkveld, A. J., Dreber, A., Holzmeister, F., Huber, J., Johannesson, M., Kirchler, M., et al. (2024). Nonstandard errors. *The Journal of Finance, 79*(3), 2339–2390. https://doi.org/10.1111/jofi.13337

Meyer, K. E., van Witteloostuijn, A., & Beugelsdijk, S. (2020). What's in a *p*? Reassessing best practices for conducting and reporting hypothesis-testing research. In L. Eden, B. B. Nielsen, & A. Verbeke (Eds.), *Research methods in international business.* JIBS Special Collections, Palgrave Macmillan. https://doi.org/10.1007/978-3-030-22113-3_4

Morck, R., & Yeung, B. (1992). Internalization: An event study test. *Journal of International Economics, 33*(1–2), 41–56. https://doi.org/10.1016/0022-1996(92)90049-P

Murphy, K. R., & Aguinis, H. (2019). HARKing: How badly can cherry-picking and question trolling produce bias in published results? *Journal of Business and Psychology, 34*(1), 1–17. https://doi.org/10.1007/s10869-017-9524-7

Musani, S. K., Shriner, D., Liu, N., Feng, R., Coffey, C. S., Yi, N., Tiwari, H. K., & Allison, D. B. (2007). Detection of gene × gene interactions in genome-wide association studies of human population data. *Human Heredity, 63*(2), 67–84. https://doi.org/10.1159/000099179

Nandialath, A. M., & Rogmans, T. (2019). Assessing the determinants of FDI in emerging markets: Do natural resources and institutions matter? *The IUP Journal of Applied Economics, 18*(3), 80–99.

Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology, 69*(1), 511–534. https://doi.org/10.1146/annurev-psych-122216-011836

North, D. C. (1990). A transaction cost theory of politics. *Journal of Theoretical Politics, 2*(4), 355–367. https://doi.org/10.1177/0951692890002004001

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

O'Boyle, E. H., Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphosize into beautiful articles. *Journal of Management, 43*(2), 376–399. https://doi.org/10.1177/0149206314527133

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

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Frontiers in Psychology, 10*, 1–10. https://doi.org/10.3389/fpsyg.2019.02970

Platt, J. R. (1964). Strong inference. *Science, 146*(3642), 347–353. https://doi.org/10.1126/science.146.3642.347

Rohrer, J. M., Tierney, W., Uhlmann, E. L., DeBruine, L. M., Heyman, T., Jones, B., et al. (2021). Putting the self in self-correction: Findings from the loss-of-confidence project. *Perspectives on Psychological Science, 16*(6), 1255–1269. https://doi.org/10.1177/1745691620964106

Rosenthal, R. (1978). How often are our numbers wrong? *American Psychologist, 33*(11), 1005–1008. https://doi.org/10.1037/0003-066X.33.11.1005

RRBM (Responsible Research for Business and Management) Network. (2024). *A vision of responsible research in business and management*. Retrieved October 13, 2024, from https://www.rrbm.network/.

Rumyantseva, M., & Welch, C. (2023). The born global and international new venture revisited: An alternative explanation for early and rapid internationalization. *Journal of International Business Studies, 54*(7), 1193–1221. https://doi.org/10.1057/s41267-023-00613-2

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

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development, 31*(1), e2295. https://doi.org/10.1002/icd.2295

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods, 17*(4), 551–566. https://doi.org/10.1037/a0029487

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C. M., van Assen, M. A. L. M., et al. (2021). Same data, different conclusions: Radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organizational Behavior and Human Decision Processes, 165*, 228–249. https://doi.org/10.1016/j.obhdp.2021.02.003

Schweinsberg, M., Thau, S., & Pillutla, M. (2023). Research-problem validity in primary research: Precision and transparency in characterizing past knowledge. *Perspectives on Psychological Science, 18*(5), 1230–1243. https://doi.org/10.1177/17456916221144990

Scott, W. R. (2008). Approaching adulthood: The maturing of institutional theory. *Theory and Society, 37*(5), 427–442. https://doi.org/10.1007/s11186-008-9067-z

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: Does entry mode choice affect FDI survival? *Management Science, 44*(4), 571–585. https://doi.org/10.1287/mnsc.44.4.571

Shen, Z., Puig, F., & Paul, J. (2017). Foreign market entry mode research: A review and research agenda. *International Trade Journal, 31*(5), 429–456. https://doi.org/10.1080/08853908.2017.1361368

Shrestha, Y. R., He, V. F., Puranam, P., & Krogh, G. (2021). Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science, 32*(3), 856–880. https://doi.org/10.1287/orsc.2020.1382

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science, 1*(3), 337–356. https://doi.org/10.1177/2515245917747646

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U. (2011). Spurious? Name similarity effects (implicit egotism) in marriage, job, and moving decisions. *Journal of Personality and Social Psychology, 101*(1), 1–24. https://doi.org/10.1037/a0021990

Simonsohn, U. (2013). It really just does not follow, comments on Francis (2013). *Journal of Mathematical Psychology, 57*(5), 174–176. https://doi.org/10.1016/j.jmp.2013.03.006

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*, 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour, 4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Stanley, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys, 19*(3), 309–345. https://doi.org/10.1111/j.0950-0804.2005.00250.x

Stanley, T. D., & Doucouliagos, H. (2012). Meta-regression analysis in economics and business. *Meta-Regression Analysis in Economics and Business, 5*, 1–190. https://doi.org/10.4324/9780203111710

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science, 11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Sun, P., Hu, H. W., & Hillman, A. J. (2016). The dark side of board political capital: Enabling blockholder rent appropriation. *Academy of Management Journal, 59*(5), 1801–1822. https://doi.org/10.5465/amj.2014.0425

Thelen, B. A., & Thiet, R. K. (2008). Cultivating connection: Incorporating meaningful citizen science into Cape Cod National Seashore's estuarine research and monitoring programs. *Park Science, 25*(1), 74–80.

Tierney, W., Cyrus-Lai, W., Hoogeveen, S., Haaf, J., Landy, J., Hardy, J., et al. (2025). *Who respects an angry woman? A crowdsourced systematic reexamination of the relationships between gender, emotion expression, and status conferral*. Working paper.

Tihanyi, L., Griffith, D. A., & Russell, C. J. (2005). The effect of cultural distance on entry mode choice, international diversification, and MNE performance: A meta-analysis. *Journal of International Business Studies, 36*(3), 270–283. https://doi.org/10.1057/palgrave.jibs.8400136

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*(6), 632–638. https://doi.org/10.1177/1745691612463078

Weingarten, E., Chen, Q., McAdams, M., Yi, J., Hepler, J., & Albarracin, D. (2016). On priming action: Conclusions from a meta-analysis of the behavioral effects of incidentally-presented words. *Current Opinion in Psychology, 12*, 53–57. https://doi.org/10.1016/j.copsyc.2016.04.015

Wernerfelt, B., & Karnani, A. (1987). Competitive strategy under uncertainty. *Strategic Management Journal, 8*(2), 187–194. https://doi.org/10.1002/smj.4250080209

Wiedersheim-Paul, F., & Johanson, J. (1975). The internationalization of the firm: Four Swedish case. *Journal of Management Studies, 22*(4), 305–322. https://doi.org/10.1111/j.1467-6486.1975.tb00514.x

Wooster, R. B., Blanco, L., & Sawyer, W. C. (2016). Equity commitment under uncertainty: A hierarchical model of real option entry mode choices. *International Business Review, 25*(1, Part B), 382–394. https://doi.org/10.1016/j.ibusrev.2015.07.006

Wu, J., Zhou, N., Park, S. H., Khan, Z., & Meyer, M. (2022). The role of FDI motives in the link between institutional distance and subsidiary ownership choice by emerging market multinational enterprises. *British Journal of Management, 33*(3), 1371–1394. https://doi.org/10.1111/1467-8551.12508

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099

Young, C., & Holsteen, K. (2017). Model uncertainty and robustness: A computational framework for multimodel analysis. *Sociological Methods & Research, 46*(1), 3–40. https://doi.org/10.1177/0049124115610347

Zeng, J., Khan, Z., & De Silva, M. (2019). The emergence of multi-sided platform MNEs: Internalization theory and networks. *International Business Review, 28*(6), 101598. https://doi.org/10.1016/j.ibusrev.2019.101598

Zhao, H., Luo, Y., & Suh, T. (2004). Transaction cost determinants and ownership-based entry mode choice: A meta-analytical review. *Journal of International Business Studies, 35*(6), 524–544. https://doi.org/10.1057/palgrave.jibs.8400106

Andrew Delios (PhD, Ivey Business School, Western University of Ontario) is Professor and Vice Dean (MSc Programs), NUS Business School, National University of Singapore. He is a Fellow of the Academy of International Business and the Asia Academy of Management. His research interests are strategy and the internationalization process, with a focus on firms operating in the Asia-Pacific region.

Tianyou Hu is Assistant Professor in the Faculty of Business Administration at the University of Macau. He received his undergraduate degree from Peking University and his PhD from National University of Singapore. His current research in strategy and international business focuses on firm strategy within alliances and networks, and how multinational firms navigate institutional environments in foreign direct investment.

Shu Yu (PhD, National University of Singapore) is Senior Research Fellow at the Suzhou Industrial Park Monash Research Institute of Science and Technology and Senior Lecturer at the Monash University Suzhou Campus. Her research interests lie in corporate and international strategy, with a focus on emerging economies.

Nan Zhou is Professor at the School of Economics and Management, Tongji University in China. She received a PhD from the Wharton School, University of Pennsylvania. Her research addresses questions that intersect the fields of corporate strategy and international business, focusing primarily on understanding how firm growth is influenced by firm resources and institutional environments in the context of emerging markets.

Eric Uhlmann is Professor of Organizational Behavior at INSEAD. His research centers on stereotyping and discrimination, moral judgments and behaviors, cross-cultural similarities and differences, and crowdsourcing science. He received a PhD in Social Psychology from Yale University in 2006 and was a postdoctoral scholar at the Kellogg School of Management.

Authors and Affiliations

Andrew Delios[1] · Tianyou Hu[2] · Shu Yu[3] · Nan Zhou[4] · Faisal M. Ahsan[5] · Mona Bahl[6] · Tao Bai[7] · Madhurima Basu[8] · Hanoku Bathula[9] · Georgios Batsakis[10] · Jorge Carneiro[11] · Dwarka Chakravarty[12] · Danyang Chen[13] · Weihong Chen[14] · Ying-Yu Chen[15] · Luis Alfonso Dau[16] · Shu Deng[17] · Desislava Dikova[18] · Xiaomin Fan[19] · Viswa Prasad Gada[20] · Dongdong Huang[21] · Hyun Gon Kim[22] · Kyungjoong Kim[23] · Aleš Kubíček[24] · Chengguang Li[25] · Wen Helena Li[26] · Yi Li[27] · Yuanyuan Li[28] · Yung-Chih Lien[29] · Huanchen Liu[30] · Wei Liu[31] · Grigorij Ljubownikow[32] · Racheal Louis Vincent[33] · Ondřej Machek[34] · Parul Manocha[35] · Yoichi Matsumoto[36] · Atthaphon Mumi[37] · Chao Niu[38] · N. Nuruzzaman[39] · Vidya Sukumara Panicker[40] · K. Praveen Parboteeah[41] · Wei Qiao[42] · Xiaole Qiao[43] · Shyamala Sethuram[44] · Ao Shen[45] · Lei Shi[46] · Evis Sinani[47] · Sandeep Sivakumar[48] · Pei-Shan Soon[49] · Maximilian Stallkamp[50] · Priit Tinits[51] · Daniel Tolstoy[52] · Andranik Tumasjan[53] · Mayank Varshney[54] · Andres Velez-Calle[55] · Chris Wagner[56] · Peng Wang[57] · Xingang Wang[58] · Yong Wang[59] · Liang Wen[60] · Tao Wu[61] · Sandeep Yadav[62] · Jiaju Yan[63] · Jing Yu Yang[64] · Megan Zhang[65] · Weihao Zhang[66] · Yameng Zhang[67] · Yang Zhao[68] · Eric Uhlmann[69]

✉ Tianyou Hu
tianyouhu@um.edu.mo

✉ Nan Zhou
zhounan38@hotmail.com

1 National University of Singapore, Singapore, Singapore

2 Faculty of Business Administration, University of Macau, E22, Avenida da Universidade, Taipa, Macau

3 Monash University & Suzhou Industrial Park Monash Research Institute of Science and Technology, Suzhou, China

4 School of Economics and Management & Advanced Institute of Business, Tongji University, 1500 Siping Road, Shanghai 200092, China

5 Xavier School of Management, Jamshedpur, India

6 Illinois State University, Normal, USA

7 The University of Queensland, Brisbane, Australia

8 Symbiosis Institute of International Business & Symbiosis International (Deemed University), Pune, India

9 The University of Auckland, Auckland, New Zealand

10 The American College of Greece & Brunel University of London, Uxbridge, UK

11 FGV EAESP Sao Paulo, School of Business Administration, São Paulo, Brazil

12 San Diego State University, San Diego, USA

13 Shanghai University of Finance and Economics, Shanghai, China

14 Guangxi University, Nanning, China

15 National Dong Hwa University, Hualien, Taiwan

16 Northeastern University, Boston, USA

17 The University of Mississippi, Oxford, USA

18 Vienna University of Economics and Business, Vienna, Austria

19 Nanjing University of Science and Technology, Nanjing, China

20 Glasgow Caledonian University, Glasgow, UK

21 Nankai University, Tianjin, China

22 The State University of New Jersey, New Brunswick, USA

23 Northwest Missouri State University, Maryville, USA

24 Prague University of Economics and Business, Prague, Czechia

25 Technical University of Munich, Munich, Germany

26 University of Technology Sydney, Ultimo, Australia

27 The University of Sydney, Camperdown, Australia

28 California State University, Los Angeles, Los Angeles, USA

29 National Taiwan University, Taipei, Taiwan

30 Nanjing University of Aeronautics and Astronautics, Nanjing, China

31 Qingdao University, Qingdao, China

32 The University of Auckland, Auckland, New Zealand

33 Monash University Malaysia, Subang Jaya, Malaysia

34 Prague University of Economics and Business, Prague, Czechia

35 University of Alabama at Birmingham, Birmingham, USA

36 Keio University, Minato, Japan

37 Mahasarakham University and National Institute of Development Administration, Kham Riang, Thailand

38 City University of Hong Kong, Kowloon, Hong Kong

39 University of Manchester, Manchester, UK

40 Loughborough Business School, Loughborough, UK

41 University of Wisconsin – Whitewater, Whitewater, USA

42 Xiamen University, Xiamen, China

43 Chang'an University, Xi'an, China

44 Baruch College, New York, USA

45 Xidian University, Xi'an, China

46 University of International Business and Economics, Beijing, China

47 Copenhagen Business School, Frederiksberg, Denmark

48 Indian Institute of Management Raipur, Raipur, India

49 Sunway University, Subang Jaya, Malaysia

[50] East Carolina University, Greenville, USA

[51] Permanent Representation of Estonia to the OECD, Paris, France

[52] Stockholm School of Economics, Stockholm, Sweden

[53] Johannes Gutenberg University Mainz, Mainz, Germany

[54] Indian Institute of Management Ahmedabad, Ahmedabad, India

[55] Universidad EAFIT, Medellín, Colombia

[56] ananki.ai GmbH, Menlo Park, USA

[57] Beijing Normal-Hong Kong Baptist University, Zhuhai, China

[58] The University of Auckland, Auckland, New Zealand

[59] Southwest Jiaotong University, Chengdu, China

[60] Xi'an Jiaotong-Liverpool University, Suzhou, China

[61] The Chinese University of Hong Kong, Shenzhen, China

[62] Indian Institute of Management Bangalore, Bengaluru, India

[63] Baylor University, Texas, USA

[64] The University of Sydney, Sydney, Australia

[65] Independent Researcher, New York, USA

[66] Guangxi University, Nanning, China

[67] Xi'an Jiaotong-Liverpool University, Suzhou, China

[68] Northwest Normal University, Lanzhou, China

[69] INSEAD, Singapore, Singapore