

## Notes on the Fringes of the Replication Crisis in Management Science

Submitted: 01.08.19 | Accepted: 14.09.19

### Przemysław Hensel\*

The text aims to provide an introduction to the replicability and reproducibility issue in management science while at the same time highlighting possible problems associated with replication research. The study is based on the review of the current literature.

Three central issues are highlighted: a) the scarcity of replication attempts is caused by the incentive structure faced by management scholars, b) since the majority of published replications (interstudy and intrastudy) are authored by the researchers who conducted the original study, their results can be affected by the same incentives that affect the results of the original study, c) the popularity of research findings seems to be unaffected by failed attempts to reproduce them.

This introductory treatment of the issue suggests that further examination of the relationship between the authorship and replication results is warranted.

Increasing the number of replication studies requires a significant change in the incentive structure to which scholars are exposed.

**Keywords:** methods, replications, reproducibility crisis.

### Uwagi na marginesie kryzysu replikacji w naukach o zarządzaniu

Nadesłany: 01.08.19 | Zaakceptowany do druku: 14.09.19

Celem tekstu jest przedstawienie wprowadzenia do problemu replikowalności i reprodukowalności wyników w naukach o zarządzaniu, z jednoczesnym podkreśleniem wyzwań powiązanych z realizacją badań replikacyjnych. Tekst oparto na przeglądzie współczesnej literatury tematu.

Wskazano na trzy główne kwestie: a) rzadkość badań replikacyjnych wynika ze struktury zachęt, z którą mają do czynienia badacze zarządzania; b) ponieważ większość opublikowanych badań replikacyjnych została przeprowadzona przez autorów replikowanego badania, ich rezultaty mogą być pochodną tej samej struktury zachęt; c) publikacja badań, w których nie udało się ponownie uzyskać wyników z oryginalnego badania wydaje się nie wpływać na popularność opracowań, których nie udało się powtórzyć.

Tekst ma naturę wstępnej analizy zagadnienia, dlatego konieczne są dalsze badania poświęcone temu problemowi.

Zwiększenie liczby badań replikacyjnych będzie wymagać zmiany struktury zachęt, na którą wystawieni są naukowcy.

**Słowa kluczowe:** metody badań, badania replikacyjne, kryzys reprodukcyjności.

**JEL:** M10

---

\* **Przemysław Hensel** – prof. dr hab., Faculty of Management, University of Warsaw. <https://orcid.org/0000-0002-2352-2317>.

---

Correspondence address: Faculty of Management, University of Warsaw, 1/3 Szturmowa Street, 02-678 Warsaw; e-mail: [phensel@wz.uw.edu.pl](mailto:phensel@wz.uw.edu.pl).



Ministry of Science  
and Higher Education  
Republic of Poland

The creation of the English-language version of these publications is financed in the framework of contract No. 607/P-DUN/2018 by the Ministry of Science and Higher Education committed to activities aimed at the promotion of education.

## 1. Introduction

At the turn of the 1930s, a series of studies were conducted at the Western Electric plant that led to a breakthrough in management sciences. The impact of lighting (Snow, 1927) and other factors (Mayo, 1933; Roethlisberger & Dickson, 1941; Whitehead, 1938) on the productivity of line workers was studied for almost 8 years. The study lasted so long because employee productivity increased regardless of changes in the working conditions. The breakthrough did not come until the researchers understood that it was not the level of lighting but a change in interpersonal relationships to more inclusive that resulted in a productivity increase. Thereby, the so-called Hawthorne effect was discovered — the interest in employees, not just in their work, promotes productivity, and the awareness of participating in an experiment affects the behaviour of its participants (French, 1950). This important lesson is taught at business schools worldwide and is also described in both management and psychology textbooks (Adair, 1984; Jones, 1992). This beautiful story has only one deficiency – the described effect probably never occurred during the Hawthorne experiment. A re-analysis of the data collected at the Western Electric plant indicates that linking the interest in employees' affairs with an increase in their productivity is unjustified (Franke & Kaul, 1978; Izawa et al., 2011; Jones, 1992; Levitt & List, 2011; Parsons, 1974).

The above story very well illustrates the problem of a small number of replication studies (Evanschitzky et al., 2007; Hubbard & Armstrong, 1994; Hubbard & Vetter, 1996; Hubbard et al., 1998) in organisational and management research. The vast majority of results obtained in our discipline have never been reproduced and replicated, hence the level of their substantiation should be considered low, as evidenced by frequent failures to replicate research in other sub-disciplines of social sciences (Camerer et al., 2018) such as psychology (Open Science Collaboration, 2015) or economics (Camerer et al., 2016). Poorly substantiated results often gain great popularity, with new currents of practice and research emerging around them. This text seeks to define the concept of a replication study, to indicate what forms such studies can take and to show their usefulness for the development of research on organisations. The article relies on a current literature review based on the snowballing technique (Greenhalgh & Peacock, 2005) whereby the identified key sources lead to the discovery of seminal literature publications.

## 2. Reasons for Interest in Replication Studies

For many decades, replication studies remained on the fringes of interest of the scientific community trying to develop knowledge about organisations and management. Only in the last decade have they begun to attract

more attention from researchers active in this field. What are the reasons for this shift?

The growing awareness of widespread questionable research practices (Bedeian, 2007) should be considered the primary reason since such practices significantly affect the credibility of research results, hence also the legitimacy of our scientific discipline (Hensel, 2017). Questionable research practices can be defined as practices that allow obtaining the expected research result (i.e. statistical significance), while bending the rules of research methods. It is worth emphasising that this does not involve simple falsification of data (although such cases also occur, as evidenced by the growing number of retracted articles (Atwater et al., 2014; Karabag & Berggren, 2016)) but a modification of hypothesis testing, data manipulation and the reporting of analysis results.

Questionable research practices include, first and foremost, **multiple tests in search of “asterisks” indicating statistically significant results** (Bettis, 2012). A correctly conducted hypothesis testing procedure consists in putting forward a hypothesis first and then testing it on the basis of data collected. “Asterisk seekers” do the opposite: first, they search for statistically significant correlations in the data set and then put forward hypotheses that could be “tested” by means of the data. In other words, multiple tests to discover “statistical significance” enables so-called HARKing (N. L. Kerr, 1998), i.e. hypothesising after the results are known.

Unfortunately, this research practice undermines the reliability of the obtained statistical significance indicators. This problem can be presented by the following analogy: if there are 100 balls in the lottery box, including one black, then the probability of the black ball being randomly drawn by a blindfolded person is 1%. However, if the same person is looking for the black ball with open eyes, then the probability of it being pulled out is 100% because it is difficult to talk about a ‘random draw’ in such a situation. As early as in the 1960s, economists showed that the probability of finding statistically significant relationships in a large set of unrelated time series is high (Ames & Reiter, 1961). This problem has significantly intensified in recent decades due to two changes in researchers’ technical environment. Firstly, the widespread use of personal computers has made the search for “asterisks” extremely easy and cheap. All correlation coefficients among a hundred variables, even for a large research sample, can be calculated within a fraction of a second; for large datasets (consisting of millions of records), such an analysis can take several seconds. In the 1970s, a similar operation was practically impossible for large data sets. More importantly, it was often necessary to have a special computing centre perform the calculation, thus making it impossible to hunt for asterisks in the recess of one’s own office.

The second factor facilitating this practice is the popularity of all kinds of sensors (position and motion recorders, GPS, cameras, etc.) and automati-

cally created registers containing ‘big data’ (George et al., 2014; George et al., 2016). Consider this example: If we wish to see at what hours office employees are the most active, we can try correlating the time of day (hour) with the time the employee logs into the computer system. Should this relationship turn out to be statistically insignificant, we can look for a correlation with the number of mouse clicks. If this does not work, we can try to find a relationship with the number of open application windows. Or with the number of keys pressed. Or emails sent. Or emails received. Or with the number of files saved on the disk. Or with the length of the file names saved on the disk. The possibilities are practically infinite and a “statistically significant” relationship can be found sooner or later. And the identified relationship can always be convincingly justified (Watts, 2013). In other words, the results of an exploratory study are presented as the results of a study based on hypothesis testing, with negative results (e.g. having no statistical significance) being ignored in the final study report.

Obviously, searching for unknown relationships in data sets is a valuable data analysis practice. However, in order for it to be considered appropriate, it is necessary to inform the reader that data mining was first carried out and only on that basis were interesting hypotheses put forward. This is the case of THARKing, i.e. transparent HARKing. Nonetheless, the real problem is SHARKing, meaning secret HARKing, which is hidden from readers, who are thereby misled as to the significance of the obtained statistical significance coefficients (Hollenbeck & Wright, 2017).

Questionable research practices may also involve **manipulation of data and data reporting to obtain “statistical significance”**. The number of questionable research practices that can lead to an “appropriate” analysis result is high. In this context, contemporary literature often refers to the term “p-hacking” proposed by psychologists (Simmons et al., 2011). In a situation where the original study design and pre-formulated hypotheses did not bring the expected results, some researchers follow practices such as: excluding a part of the data set from the analysis and reporting results only from other data, including or not including outliers, collecting additional data, adding or removing covariates, not reporting hypotheses that were statistically insignificant, and the above-mentioned post-hoc hypothesising. Simply put, there are many researcher degrees of freedom (Simmons et al., 2011) in choosing methods for analysing and reporting data. The elimination of these degrees of freedom is impossible because every research process requires a series of decisions without which raw data cannot be transformed into a scientific result. However, if these decisions are determined by the wish to obtain publishable results at all costs, the effect is research reports that falsely represent the phenomenon under study.

Opinion polls conducted in this decade suggest that the popularity of questionable research practices can be surprisingly huge. Surveys of academics employed in 104 AACSB-accredited management faculties indeed

revealed surprising and disturbing results (Bedeian et al., 2010). Of 384 people who responded to the survey, 79.2% said that they knew cases where the researcher had concealed methodological details or some study results. 77.6% knew about situations in which researchers had based their results only on data that allowed for obtaining the intended test results. 72.1% declared that they knew researchers who had used other people's ideas without permission and mention of the source. Almost 60% knew cases where other researchers had excluded data from the research sample as they felt that they might be inaccurate. Almost half of the respondents knew researchers who had concealed data that undermined their previous research. According to 26.8% of the respondents, other researchers went as far as to fabricate data. Finally, almost all respondents (91.9%) admitted that they were aware of cases where hypotheses were made after the results were already known (HARKing mentioned above). Similar results were obtained in surveys among scholars representing other fields of science where as many as 33.7% of respondents admit to using questionable research practices, while 72% suspect other scholars of such practices (Fanelli, 2009).

The interest in replication research is also a derivative of the results obtained in large projects aimed at reproducing previous studies. Employees of the Bayer pharmaceutical company who tried to reproduce the results of published oncology and cardiology studies failed in two-thirds of 67 replications (Prinz et al., 2011). Similar results were obtained in a project undertaken by the Amgen company (Begley & Ellis, 2012). In the field of psychology, which is closer to ours, the results of a project attempting to replicate 100 studies published in three major psychology journals attracted great attention (Open Science Collaboration, 2015). 97% of original studies contained statistically significant results, yet only 36% of the results of replicated studies were statistically significant. The authors of the project conclude that 39% of the original studies were successfully reproduced.

Finally, attention should be paid to an issue that is rarely raised in the literature on replication studies, but seems to be crucial. There are examples showing that the popularity of scientific results in our field is determined not so much by the quality of the research methods used and the reliability of results but by their compliance with the beliefs held by scientists. An excellent example of this is the research at the Western Electric plant in Hawthorne mentioned at the beginning of this text. As shown below, a re-analysis of the data collected in that research indicates that the conclusions drawn from that study are unjustified. However, the power of the Hawthorne legend is so great that even repeated unsuccessful reproductions of the study cannot undermine the belief in the truth of the results obtained by Elton Mayo. On this basis, it can be concluded that replication studies only slightly influence the beliefs of scientists (por. Sonnenfeld, 1985). Therefore, challenging erroneous results requires not just one but many replication projects. In other words, replications published

in journals concerning organisational and management research will bring the expected effect, i.e. will enhance the credibility of our knowledge about organisations, only if their number increases by leaps and bounds.

### 3. Types of Replication Studies

Replication can be briefly defined as a study consisting in the repetition of an earlier study in order to verify it (Frankfort-Nachmias & Nachmias, 1996). Projects that aim to accurately replicate the original study are called literal or exact replications while those in which changes are made to the original study design are usually called conceptual (Schmidt, 2009) or constructive (Lykken, 1968). Thus far, the most complete systematics of replication studies in organisational and management theories has been proposed by Tsang et al. (1999). In their view, six different types of such studies can be distinguished that differ in their compliance with the original.

**Analysis check.** This type, generally referred to as reproduction, uses the same data and analysis tools as the original study. The goal is to detect any errors made by the authors of the study. In management sciences, analysis checks are rarely performed since many studies are based on unique data sets that no one, except the authors of the original study, can access. The sharing of source data sets at the request of other researchers is ever more frequently suggested, which is to facilitate such reproductions (Atwater et al., 2014; Ethiraj et al., 2017). The potential of such studies is best illustrated by checks of analyses undertaken by economists. In an article from 2010, Carmen Reinhart and Kenneth Rogoff (Reinhart & Rogoff, 2010) showed that countries with high public (government) debt have a negative economic growth rate. That article gained tremendous popularity and became the justification for promoting the austerity policy. However, a re-analysis revealed (Herndon et al., 2014) that the authors failed to analyse several rows of data in the Excel file, which resulted in a negative, instead of a positive, correlation. Perhaps replication studies in management science are more difficult than in exact sciences but an analysis check is undoubtedly possible in almost every case when source data and the programme code used for calculations are made available.

**Data re-analysis.** A more advanced type of replication study is a data re-analysis that uses methods of analysis or theories unknown to the authors of the original study to verify the justification of the findings of that study. A great illustration of this replication method is provided by numerous studies on the aforementioned Hawthorne effect. For example, using the reinforcement theory, Parsons (1974) pointed out that an increase in productivity of employees participating in the mentioned experiment can be explained by two factors overlooked by Elton Mayo's team: feedback about the level of productivity and a change in the method of remuneration (Parsons, 1974). Western Electric employees were remunerated based on

the performance achieved by a team of 100 or more people, hence the impact of individual employees' performance on earnings was small. During the experiment, employees performed their work in teams of five, so the impact of each employee's productivity on remuneration was significantly greater. What is more, each day employees received information on the number of manufactured components and could monitor their productivity on an ongoing basis. In addition, a re-analysis of the experiment documentation revealed other factors that could affect productivity: for example, two workers partaking in the experiment were replaced by others who did their work faster. The use of multiple regression to analyse the time series collected at Hawthorne showed that over 90% of the productivity variance could be explained by the following factors: disciplinary actions undertaken by the management (employee replacement), the number of hours worked weekly, the outbreak of the Great Depression, the length of rest, and the change in the method of remuneration (Franke et al., 1978). The disciplinary action had the greatest impact on productivity, which is clearly in contradiction with Elton Mayo's findings that the improvement of interpersonal relationships played a key role. Neither did further analyses support the hypothesis that employees' productivity was influenced by factors other than the number of breaks, the number of hours worked or lighting<sup>1</sup> (Izawa et al., 2011; Jones, 1992; Levitt et al., 2011).

**Exact replication.** As the name implies, an exact replication involves a replication study that will be as close to the replicated study as possible. In particular, the same measures of constructs examined and methods of analysis are used and the sample is selected from the same population as in the original study.

**Conceptual extension.** Similar to exact replication, this type of repeated study is based on a sample from the population in which the original study was conducted. However, the methods of measuring variables and ways of analysis are changed. In other words, the goal is to investigate a phenomenon by means of a different research apparatus. As Tsang and Kwan (1999) rightly point out, such projects can provide interesting knowledge if previous results are confirmed. However, when a replication project fails, it is difficult to state whether the reason is an error in the original study or the difference between these two research projects.

**Empirical generalisation.** This type of replication can be seen as the exact opposite of conceptual extensions: the replication project uses the same methods as the original study but the sample is selected from another population. The goal is to check whether previous results can be generalised to other populations. It can be suspected that many research projects conducted in our country use this type of research to check whether the results obtained, for example, in the United States are also true for Poland.

**Generalisation and extension.** In the last type of replications characterised by Tsang and Kwan (1999), everything changes, except for the tested



relationship: compared to the original study, the measurement methods, ways of analysis and the population from which the sample was drawn are different. Similarly to conceptual extension, the success of replication of original results informs about the possibility of their generalisation; however, a failure can be explained by both the errors in the original results and the differences between the original and replication projects (Bettis, Helfat, et al., 2016).

#### **4. Researchers' Attitudes Toward Replication Studies**

If replication studies are very rare in management sciences (Hubbard et al., 1994; Hubbard et al., 1996; Hubbard et al., 1998) and their number decreases rather than increases over time (Evanschitzky et al., 2007), this is due to, *inter alia*, the views of the scientific community on such projects. Replications are often seen as little creative (Hendrick, 1991; Hubbard et al., 1994; Mezas & Regnier, 2007), thus standing in contradiction to the desire to produce new (G.F. Davis, 2015; Geuens, 2011) and surprising results (M.S. Davis, 1971; Hensel, 2017) dominant in management sciences. Prejudice against replication studies has also been shown in surveys of editors and reviewers of social science journals (Easley et al., 2013; Hensel, 2019; S. Kerr et al., 1977; Madden et al., 1995; Neuliep & Crandall, 1991, 1993a, 1993b; Rowney & Zenisek, 1980; Yuksel, 2003). If editors seem reluctant to publish replication studies, it is probably not only because such studies are not particularly innovative but also because they attract fewer citations (Hubbard et al., 1994). Editors might be concerned that publishing too many replications may negatively affect the journal's impact factor and rankings.

So far, the views of organisation and management scholars on replications have been examined only once. In a survey of the Editorial Board of the *Academy of Management Journal*, which is a leading journal in our field, more or less the same number of respondents agreed and disagreed with the statement that the journal should be more open to replications (Ketchen Jr & Ireland, 2010). This indicates that even a small number of replications currently appearing in print is a problem for roughly half of the community.

#### **5. Conclusions**

It is assumed in the literature that replication studies allow for correcting erroneous results and make it possible to cope with the previously mentioned questionable research practices (Atwater et al., 2014; Haenlein, 2012; Honig et al., 2014; Kenworthy & Sparks, 2016; Leung, 2011). In addition, the results of replication projects can contribute to theory-building by indicating the limits of generalisation and the determinants of previous



results (Tsang et al., 1999; Uncles & Kwok, 2013). A series of replication studies can also help to more accurately estimate the magnitude of an effect (e.g., the coefficient of correlation between variables) (Hunter, 2001).

Nonetheless, a spoonful of tar spoils this barrel of honey. Supporters of replication studies seem to forget that the authors of replications are subject to the same incentives and pressures as the authors of original studies. Thus, they can stretch their results and seek statistical significance at all costs to the same extent as the latter. While in exact replications the design of the original study leaves the replicator with his or her hands strongly tied (although this is not a full tie – the research sample can still be selected in different ways, for example), conceptual replications leave considerable possibilities of result manipulation. Moreover, a recent analysis of 406 replication studies published in the *Academy of Management Journal*, *Journal of Applied Psychology* and *Journal of Management* indicates that the vast majority (78.6%) are dependent studies (Köhler & Cortina, forthcoming), meaning that a replication study was conducted by the same authors as the replicated study and usually published in the same paper. Therefore, we can suspect that the incentives that make authors stretch the results of original studies also have an impact on the results of replication studies – the authors have no interest in questioning the results of their own research.

It seems that all initiatives aimed at increasing the number of replication studies – and there are more and more of them (Bettis, Ethiraj, et al., 2016; Bettis, Helfat, et al., 2016; Van de Ven, 2015) – will not bring the expected result without first changing the structure of incentives that researchers are exposed to. A change of this structure is an extremely difficult task.

Finally, it should also be emphasised that the problem of replications is a nightmare only for representatives of the positivist (functionalist) paradigm (Burrell & Morgan, 1979). For researchers working under other paradigms, low replicability is not a serious problem because they either do not expect it (interpretivists) or believe that the primary purpose of science is a social change rather than discovery of ‘absolute’ truths. It can also be assumed that replicability is primarily dealt with by quantitative researchers employing experimental methods. Although attempts are being made to replicate qualitative studies (e.g. Boisot & Liang, 1992; Martinko & Gardner, 1990) and the level of replicability of such studies is being analysed (Aguinis & Solarino, 2019), it can be presumed that such efforts are mainly undertaken by qualitative researchers working within the positivist paradigm.

#### **Acknowledgment**

*This research was financed by the National Science Centre grant UMO-2018/31/B/H54/00390.*

*Translated from Polish language by Agata Ostrowska.*

## Endnotes

- <sup>1</sup> By the way, it is worth noting that all employees participating in the first Hawthorne experiment – on the impact of lighting on work productivity – were of Polish origin because it was easiest for the foreman, who was also Polish, to communicate with them (Izawa, French, & Hedge, 2011). This calls into question the representativeness of the sample used.

## References

- Adair, J.G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology*, 69(2), 334–345.
- Aguinis, H., & Solarino, A.M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*, 0(0).
- Ames, E., & Reiter, S. (1961). Distributions of Correlation Coefficients in Economic Time Series. *Journal of the American Statistical Association*, 56(295), 637–656.
- Atwater, L.E., Mumford, M.D., Schriesheim, C.A., & Yammarino, F.J. (2014). Editorial: Retraction of leadership articles: Causes and prevention. *The Leadership Quarterly*, 25(6), 1174–1180.
- Bedeian, A.G. (2007). Even if the Tower Is “Ivory,” It Isn’t “White:” Understanding the Consequences of Faculty Cynicism. *Academy of Management Learning & Education*, 6(1), 9–32.
- Bedeian, A.G., Taylor, S.G., & Miller, A.N. (2010). Management Science on the Credibility Bubble: Cardinal Sins and Various Misdemeanors. *Academy of Management Learning & Education*, 9(1), 715–725.
- Begley, C.G., & Ellis, L.M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), 531.
- Bettis, R.A. (2012). The search for asterisks: Compromised statistical tests and flawed theories. *Strategic Management Journal*, 33(1), 108–113.
- Bettis, R.A., Ethiraj, S., Gambardella, A., et al. (2016). Creating repeatable cumulative knowledge in strategic management. *Strategic Management Journal*, 37(2), 257–261.
- Bettis, R.A., Helfat, C.E., & Shaver, J.M. (2016). The necessity, logic, and forms of replication. *Strategic Management Journal*, 37(11), 2193–2203.
- Boisot, M., & Liang, X.G. (1992). The Nature of Managerial Work in the Chinese Enterprise Reforms. A Study of Six Directors. *Organization Studies*, 13(2), 161–184.
- Burrell, G., & Morgan, G. (1979). *Sociological Paradigms and Organizational Analysis: Elements of the Sociology of Corporate Life*. London: Heinemann.
- Camerer, C.F., Dreber, A., Forsell, E., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436.
- Camerer, C.F., Dreber, A., Holzmeister, F., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644.
- Davis, G.F. (2015). Editorial Essay: What Is Organizational Research For?. *Administrative Science Quarterly*, 60(2), 179–188. <https://doi.org/10.1177/0001839215585725>.
- Davis, M.S. (1971). That’s interesting! Towards a phenomenology of sociology and a sociology of phenomenology. *Philosophy of the Social Sciences*, 1(4), 309–344.
- Easley, R.W., Madden, C.S., & Gray, V. (2013). A tale of two cultures: Revisiting journal editors’ views of replication research. *Journal of Business Research*, 66, 1457–1459.
- Ethiraj, S.K., Gambardella, A., & Helfat, C.E. (2017). Improving Data Availability: A New SMJ Initiative. *Strategic Management Journal*, 38(11), 2145–2146.

- Evanschitzky, H., Baumgarth, C., Hubbard, R., & Armstrong, J.S. (2007). Replication research's disturbing trend. *Journal of Business Research*, 60, 411–415.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *Plos One*, 4(5), 1–11.
- Franke, R.H., & Kaul, J.D. (1978). The Hawthorne Experiments: First Statistical Interpretation. *American Sociological Review*, (5), 623.
- Frankfort-Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences*. (5 ed.). New York: St. Martin's Press.
- French, J.R.P. (1950). Field experiments: Changing group productivity. In: J.G. Miller (Ed.), *Experiments in social process: A symposium on social psychology* (pp. 81–96). New York: McGraw-Hill.
- George, G., Haas, M.R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, 57(2), 321–326.
- George, G., Osinga, E.C., Lavie, D., & Scott, B.A. (2016). Big Data and Data Science Methods for Management Research. *Academy of Management Journal*, 59(5), 1493–1507.
- Geuens, M. (2011). Where does business research go from here? Food-for-thought on academic papers in business research. *Journal of Business Research*, 64(10), 1104–1107.
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ*, 331(7524), 1064–1065.
- Haenlein, M. (2012). From the incoming Editor: Happy Birthday EMJ! *European Management Journal*, 30(4), 293–294.
- Hendrick, C. (1991). Replications, strict replications, and conceptual replications: Are they important? In: J.W. Neuliep (Ed.), *Replication research in the social sciences* (pp. 41–49). Thousand Oaks: Sage.
- Hensel, P.G. (2017). *Legitymizacja badań organizacji*. Warszawa: PWN.
- Hensel, P.G. (2019). Supporting replication research in management journals: Qualitative analysis of editorials published between 1970 and 2015. *European Management Journal*, 37(1), 45–57.
- Herndon, T., Ash, M., & Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. *Cambridge Journal of Economics*, 38(2), 257–279.
- Hollenbeck, J.R., & Wright, P.M. (2017). Harking, Sharking, and Tharking: Making the Case for Post Hoc Analysis of Scientific Data. *Journal of Management*, 43(1), 5–18.
- Honig, B., Lampel, J., Siegel, D., & Drnevich, P. (2014). Ethics in the production and dissemination of management research: Institutional failure or individual fallibility? *Journal of Management Studies*, 51(1), 118–142.
- Hubbard, R., & Armstrong, J.S. (1994). Replications and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing*, 11(3), 233–248.
- Hubbard, R., & Vetter, D.E. (1996). An Empirical Comparison of Published Replication Research in Accounting, Economics, Finance, Management, and Marketing. *Journal of Business Research*, 35(2), 153–164.
- Hubbard, R., Vetter, D.E., & Little, E.L. (1998). Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal*, 19(3), 243–254.
- Hunter, J.E. (2001). The Desperate Need for Replications. *Journal of Consumer Research*, 28(1), 149–158.
- Izawa, M.R., French, M.D., & Hedge, A. (2011). Shining new light on the hawthorne illumination experiments. *Human Factors*, 53(5), 528–547.
- Jones, S.R.G. (1992). Was There a Hawthorne Effect? *American Journal of Sociology*, 98(3), 451–468.

- Karabag, S.F., & Berggren, C. (2016). Misconduct, Marginality and Editorial Practices in Management, Business and Economics Journals. *Plos One*, 11(7), 1–25.
- Kenworthy, T.P., & Sparks, J.R. (2016). A scientific realism perspective on scientific progress in marketing: An analysis of theory testing in marketing's major journals. *European Management Journal*, 34(5), 466–474.
- Kerr, N.L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality & Social Psychology Review*, 2(3), 196–217.
- Kerr, S., Tolliver, J., & Petree, D. (1977). Manuscript Characteristics Which Influence Acceptance for Management and Social Science Journals. *Academy of Management Journal*, 20(1), 132–141.
- Ketchen Jr, D.J., & Ireland, R.D. (2010). From the Editors: Upon Further Review: A Survey of the Academy of Management Journal's Editorial Board. *Academy of Management Journal*, 53(2), 208–217.
- Köhler, T., & Cortina, J.M. (forthcoming). Play It Again, Sam! An Analysis of Constructive Replication in the Organizational Sciences. *Journal of Management*. <https://doi.org/10.1177/0149206319843985>.
- Leung, K. (2011). Presenting Post Hoc Hypotheses as A Priori: Ethical and Theoretical Issues. *Management & Organization Review*, 7(3), 471–479.
- Levitt, S., D., & List, J., A. (2011). Was There Really a Hawthorne Effect at the Hawthorne Plant? An Analysis of the Original Illumination Experiments. *American Economic Journal: Applied Economics*, (1), 224.
- Lykken, D. T. (1968). Statistical Significance in Psychological Research. *Psychological Bulletin*, 70(3, Pt.1), 151–159.
- Madden, C.S., Easley, R.W., & Dunn, M.G. (1995). How Journal Editors View Replication Research. *Journal of Advertising*, 24(4), 77–87.
- Martinko, M.J., & Gardner, W.L. (1990). Structured Observation of Managerial Work: A Replication and Synthesis\*. *Journal of Management Studies*, 27(3), 329–357.
- Mayo, E. (1933). *The human problems of an industrial civilization*. New York: Macmillan.
- Mezias, S.J., & Regnier, M.O. (2007). Walking the walk as well as talking the talk: replication and the normal science paradigm in strategic management research. *Strategic Organization*, 5(3), 283–296.
- Neuliep, J.W., & Crandall, R. (1991). Editorial bias against replication research. *Journal of Social Behavior and Personality*, 5(4), 85–90.
- Neuliep, J.W., & Crandall, R. (1993a). Everyone was wrong: There are lots of replications out there. *Journal of Social Behavior and Personality*, 8(6), 1–8.
- Neuliep, J.W., & Crandall, R. (1993b). Reviewer bias against replication research. *Journal of Social Behavior and Personality*, 8(6), 21–29.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>.
- Parsons, H.M. (1974). What Happened at Hawthorne? *Science*, 183(4128), 922–932.
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews. Drug Discovery*, 10(9), 712–712.
- Reinhart, C.M., & Rogoff, K.S. (2010). Growth in a Time of Debt. *American Economic Review*, 100(2), 573–578.
- Roethlisberger, F.J., & Dickson, W.J. (1941). *Management and the Worker*. Cambridge: Cambridge University Press.
- Rowney, J.A., & Zenisek, T.J. (1980). Manuscript characteristics influencing reviewers' decisions. *Canadian Psychology/Psychologie canadienne*, 21(1), 17–21.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100.

- 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.
- Snow, C.E. (1927, November). Research on industrial illumination: A discussion of the relation of illumination intensity to productive efficiency. *Tech Engineering News*.
- Sonnenfeld, J.A. (1985). Shedding light on the Hawthorne studies. *Journal of Organizational Behavior*, 6(2), 111–130.
- Tsang, E.W.K., & Kwan, K.-M. (1999). Replication and Theory Development in Organizational Science: A Critical Realist Perspective. *Academy of Management Review*, 24(4), 759–780.
- Uncles, M.D., & Kwok, S. (2013). Designing research with in-built differentiated replication. *Journal of Business Research*, 66, 1398–1405.
- Van de Ven, A.H. (2015). Welcome to the Academy of Management Discoveries (AMD). *Academy of Management Discoveries*, 1(1), 1–4.
- Watts, D.J. (2013). *Everything is Obvious: Once You Know the Answer*. New York: Crown Business.
- Whitehead, T.N. (1938). *The industrial worker*. Cambridge: Harvard University Press.
- Yuksel, A. (2003). Writing publishable papers. *Tourism Management*, 24(4), 437–446.