Editorial Essay: What Is Organizational Research For?

ARTICLE *in* ADMINISTRATIVE SCIENCE QUARTERLY · MAY 2015

Impact Factor: 4.21 · DOI: 10.1177/0001839215585725

1 AUTHOR:

Gerald F. Davis
University of Michigan

JOHNSON
Cornell University

Administrative Science Quarterly
2015, Vol. 60(2)179–188

Reprints and permissions:
sagepub.com/
journalsPermissions.nav
DOI: 10.1177/0001839215585725
asq.sagepub.com
SAGE

Editorial Essay: What Is Organizational Research For?

Gerald F. Davis[1]

Abstract

Organizational research is guided by standards of what journals will publish and what gets rewarded in scholarly careers. This system can promote novelty rather than truth and impact rather than coherence. The advent of big data, combined with our current system of scholarly career incentives, is likely to yield a high volume of novel papers with sophisticated econometrics and no obvious prospect of cumulative knowledge development. Moreover, changes in the world of organizations are not being met with changes in how and for whom organizational research is done. It is time for a dialogue on who and what organizational research is for and how that should shape our practice.

Keywords: philosophy of science, big data, theoretical progress, organizational research

When the *Administrative Science Quarterly* began publication in 1956, the post-war economy in the U.S. was populated by large and growing bureaucracies in the public and private sectors. The idea of an ''administrative science'' that would apply the insights of social science to the problem of managing bureaucracies was sensible, just as engineering applied the insights of natural science to design. *ASQ*'s aim was not to provide practical advice to managers but to build an interdisciplinary science of administration that both drew on and contributed to the broader enterprise of social science.

In the subsequent six decades, corporate bureaucracies grew, shrank, and radically changed forms. Organizational research largely migrated to business schools, and business administration grew to be by far the most popular undergraduate major. ''Administrative science'' became a global endeavor, and the number of journals and published articles in management expanded enormously. Judging by the number of scholars involved and their volume of research output, the field of organizational research has been an incredible success.

But are these the right measures? Academic careers reward publication, and our system of journals often privileges novelty over cumulative insights. Judgments of quality are difficult and time consuming, and so we rely on proxy

[1] Stephen M. Ross School of Business, University of Michigan

measures of ''impact'' that are easily gamed. Paradoxically, journals have evolved a sophisticated set of standards for evaluating research claims, but they reward being ''counterintuitive.'' As a result, while there is a lot of sound and fury, it is difficult to point to many areas of settled science when it comes to organizations.

Things are about to get worse unless we evolve new standards rooted in a clear sense of who and what organizational research is for. First, the coming tsunami of big data means that we have to reconsider how we evaluate research progress and impact. Thanks to the pervasiveness of electronic monitoring inside and outside organizations, we are on the verge of an all-you-can-eat buffet of organizational data suitable for regressionology. Coupled with our current standards of evaluation, this could result in a high volume of novel papers with sophisticated econometrics and no obvious prospect of cumulative knowledge development. Second, the changing nature of organizations means that we need to reconsider who our constituencies are, as ''managerial relevance'' is becoming an elusive goal. Enterprises today bear little resemblance to the postwar hierarchies that animated early research. Some of the best-known companies have few employees to manage, while some of the biggest rely on computer algorithms to schedule, monitor, and evaluate. This raises a fundamental question for our field: who should benefit from organizational research? Answering this question can help guide standards of evaluation for research.

The Architecture of Organizational Research

San Jose, California is home to one of the most peculiar structures ever built: the Winchester Mystery House, a 160-room Victorian mansion that includes 40 bedrooms, two ballrooms, 47 fireplaces, gold and silver chandeliers, parquet floors, and other high-end appointments. It features a number of architectural details that serve no purpose: doorways that open onto walls, labyrinthine hallways that lead nowhere, and stairways that rise only to a ceiling.

According to legend, Sarah Winchester, the extremely wealthy widow of the founder of the Winchester rifle company, was told by a spiritual medium that she would be haunted by the ghosts of the people killed by her husband's rifles unless she built a house to appease the spirits. Construction began in 1884 and by some accounts continued around the clock until her death in 1922. There was no blueprint or master plan and no consideration of what it would mean to reach completion. The point was simply to keep building, hence the sprawling and incoherent result.

Is the Winchester Mystery House a good house? It's certainly beautiful in its own way. Any given room might be well proportioned and full of appealing features. A stairway might be made of fine wood, nicely joined and varnished, and covered in a colorful carpet. Yet it ends in a ceiling and serves no useful purpose other than keeping its builders busy. In assessing whether a house is good, we have to ask, ''Good for what? Good for whom?''—the questions we would ask about other kinds of constructions. An airport is designed for a specific function, is built according to a blueprint, and is straightforward to evaluate, although evaluations might vary widely depending on people's experience with the realized design. A cathedral has a plan that might take decades to realize, with adjustments along the way, guided by a shared vision for what its realization will be. But for the Winchester Mystery House, the act of building was an end in itself. It is a paradigmatic folly, according to one find on a Google search, ''a costly ornamental building with no practical purpose.''

The field of organization studies might be compared to a sprawling structure. There can be little doubt that a lot of activity goes into constructing the field: according to the Web of Knowledge, over 8,000 articles are published every year in the 170+ journals in the field of ''Management,'' adding more and more new rooms. The questions of good for what and good for whom are worth revisiting. There is reason to worry that the reward system in our field, particularly in the publication process, is misaligned with the goals of good science: we often reward novelty over truth. As a result, we may look more like a mystery house than a cathedral.

Novelty or Truth?

The unit of currency in academic research is the publication, typically the journal article. Articles are how we convey what we found and put it in context. They are the record on which individual scientists (and departments and journals) are judged. There is an emerging consensus in some quarters that the system of journals and academic career incentives often favors novelty over truth in publications, that individual academic researchers are often rewarded for being interesting rather than getting it right, leading to systematic biases in the published record. If the advancement of knowledge were the goal of science, then individual articles would be recognized as a means, not an end in themselves. In most cases, individual articles count only as part of a totality of evidence: they are one tile in a mosaic. In our world, however, career incentives turn publications into essential tools for individual advancement, and this is not always compatible with getting it right. ''To the extent that publishing itself is rewarded, then it is in scientists' personal interests to publish, regardless of whether the published findings are true'' (Nosek, Spies, and Motyl, 2012: 616). Aguinis and colleagues (2014) pointed out that scholars give different answers to the question ''What is good for the advancement of our knowledge?'' versus ''What is good for the advancement of a scholar's career?'' The misalignment between individual career incentives and the advancement of our science is the source of much mischief.

Nosek and colleagues described some of the many ways that the quest for publication can yield articles full of intriguing yet false or misleading results. This does not require fraud by researchers or excessive sloppiness on the part of reviewers and editors. Rewarding scholars for publication per se, abetted by standard processes of motivated reasoning, is sufficient. Novelty is prized in the literature, replication is devalued, and falsification is rare. Positive findings (which might reflect the idiosyncrasies of small samples, particular designs, and/or choices about how to analyze and present data) are published, while non-findings wind up in the proverbial file drawer. There are many rewards for innovation and being the first to make a novel claim, and few punishments for getting it wrong, so that questionable findings can persist in the literature indefinitely.

In a provocative and even alarming piece on publication bias in medical journals, Ioannidis (2005) concluded that ''most research findings are false for most research designs and most fields.'' His simulation models, based on entirely plausible assumptions, show how the bias varies by subfield within medicine. Epidemiology—perhaps the field most similar to non-experimental research on organizations—has an especially poor chance of getting it right, particularly if many statistical relationships are tested on archival data but only the significant ones get reported.

Prospects for cumulation are further hobbled by a widely shared prescription for research to be ''interesting'': to find something that is generally believed to be the case and show that it is not (Davis, 1971). As a guide to delighting readers, the prescription is appealing. But if the aim of publishing articles is to advance our understanding through a cumulative process of building on prior findings, then it is hard to imagine a more nihilistic dictum than to be ''interesting.'' What is good for the career of the individual researcher might be very bad for the collective health of the scientific endeavor (cf. Pillutla and Thau, 2013).

Impact or Progress?

How do we evaluate organizational research? How do we know what a contribution is or how individual articles add up? In some sciences, progress can be measured by finding answers to questions, not merely reporting significant effects. How many moons does Jupiter have? How does the volume of a gas vary with pressure? What is the structure of DNA? Did Homo sapiens interbreed with Neanderthals? Does raising the minimum wage reduce employment? Some answers are more provisional than others, but the aim is clearly to answer questions. In many social sciences, however, including organization studies, progress is harder to judge, and the kinds of questions we ask may not yield firm answers (e.g., do nice guys finish last?). Instead we seek to measure the contribution of research by its impact.

There are many ways to assess the impact of scientific work, from book sales (where *Who Moved My Cheese?* is the best-selling management book in history) to prizes to Google hits and article downloads (Aguinis et al., 2014). By far the dominant measure of impact is citations: how often a piece is cited in subsequent works. An advantage of this measure is that it is easily accessible: Google Scholar and Web of Knowledge are just a click away. Citation metrics are widely used in faculty evaluations and routinely come up in tenure reviews. By this accounting, good science is widely cited science.

Yet there is now a vast literature on the inadequacy of these measures as indicators of research quality. Moreover, because so much is at stake, the incentives for gaming the system are irresistible to many editors and authors. According to the Web of Knowledge, seven of the ten most-cited articles published in the field of management in 2012 all appeared in the same journal. Not coincidentally, almost all of the citations came from the same small set of journals and authors, which resulted in the offending journal being suspended from the Web of Knowledge. There are many ways to engineer measures of impact that have little to do with the quality of ideas or the contribution to science, and not all of them result in suspensions.

At a more fundamental level, impact in this sense may not measure what we want. Consider what happens when police are evaluated according to their numbers of citations and arrests. We might imagine that as a society we want ''safety'' or ''justice,'' but if what we count when we evaluate police officers is ''number of arrests and citations issued,'' we get something rather different—

in the worst case, entire populations weighed down with criminal records for trivial offenses. Similarly, it is unclear why ''impact'' is an apt measure if the goal of research is to answer questions. If anything, raising questions that do not get answered, or being surprising and counterintuitive, may be better strategies for being widely cited than actually answering questions accurately (Davis, 2010). Being provocative may be more impactful than being right.

Progress in Standards of Evaluation

There are domains in which progress is unambiguous. Methodologically, we have become much more sophisticated in how we evaluate causal claims. We are really good at finding reasons to doubt, especially when we are acting as reviewers. As Cook and Campbell (1979: 38) said, ''Much of the researcher's task involves self-consciously thinking through and testing the plausibility of noncausal reasons why the two variables might be related and why 'change' might have been observed in the dependent variable even in the absence of any explicit treatment of theoretical or practical significance.'' Getting this right is particularly important in a field in which much of the research takes the form of analyzing archival data in which experimental control is absent. Thus we know enough to be wary of simple comparisons between treatment and control groups when the process that got them into these two groups might affect the outcome we care about. We know that before-and-after-treatment comparisons of a single group can be tainted by history, maturation, or regression to the mean. We know that ''dosages'' might not be equivalent in evaluating treatments, for example, how much ''total quality management'' was ingested (cf. Zbaracki, 1998).

We also get better as a community in coming to appreciate non-obvious threats to causal inference. Denrell and Kovács (2008) discovered that studies of diffusion of innovation, which typically examine innovations that spread widely, may systematically underestimate the strength of contagion processes. If we study only successful diffusions, or populations that grew large enough to be observable, we end up introducing bias. Ioannidis's (2005) study and others like it also help us understand how the structure of the system of publications intersects with the way science is done to yield untrustworthy findings. The evolution of standards of evaluation within the community of scholars is one of the ways that journal review processes can make the field better. The oppositional process leads us to develop deeper understandings about when one can claim a causal relationship and what can go wrong. As McCloskey (1998) pointed out, an aim of research is to persuade a jury of sophisticated peers that you got it right. This is why assembling a thoughtful and developmental review process involving smart and critical scholars—a jury of sophisticated peers—is the distinctive feature of journals (Davis, 2014). Statistical significance and surprise alone cannot be the features we select for.

Big Data, Big Problems

The difficulties of evaluating impact or progress and assessing a contribution are going to get worse soon. Social life and the operation of organizations increasingly leave more or less permanent data traces, like the contrails left by jets. Sometimes these become available to researchers; even more often, they are available to corporations and government agencies such as the National Security Agency. A preview of this is the Enron e-mail archive, containing about one-half million e-mails from 150 senior managers of Enron, which was made public by the Federal Energy Regulatory Commission during its investigations and subsequently cleaned and shared by researchers (https://www.cs.cmu.edu/~./enron/). Researchers have mapped the networks created by From and To headers in messages, analyzed the prevalence of sentiments in the text, coded classification schemes, and more. The data provide a fascinating look into a corrupt organization in action. The 2014 hack of Sony Pictures Entertainment's computers, allegedly by North Korean operatives, yielded scores of embarrassing e-mails and memos, as well as sensitive data on thousands of employees, including ''a Microsoft Excel file that includes the name, location, employee ID, network username, base salary and date of birth for more than 6,800 individuals'' (http://time.com/3625326/sony-hack-files/). For some researchers, that spreadsheet could have been the start of an article and the Sony data an ethically questionable window into contemporary corporate practices.

Companies routinely collect far more intrusive data on their employees than what was revealed by the Sony hack. Most major companies have implemented enterprise resource planning (ERP) systems to keep track of their operations on a second-by-second basis. ID tags can tell where workers are and with whom, as well as how long they took for lunch. Retailers often have workforce-management systems that provide real-time data on the productivity of their workers, from the speed at which cashiers scan items to whether salespeople are successful at upselling customers—data that then feed into automated employee-scheduling algorithms that reward the productive with regular hours and punish the weak with irregular and insufficient hours. One online behemoth guides its warehouse workers step by step to fill customers' orders using a GPS device that beeps with increasing stridency as the time allotted for each item counts down. Financial institutions routinely monitor employees' e-mails and screen them for suspicious phrases and emotions, which can flag offenders for follow-up with human resources.

Data on the whereabouts, productivity, compensation, demographics, social networks, emotional expression, and perhaps medical records and Fitbit streams of employees can yield horrifyingly intrusive information within individual organizations, and modest versions of these are beginning to appear in the literature. A few companies such as Google and Oracle have some of these data for hundreds of client organizations, potentially allowing unprecedented comparative data on organizational structures and processes. It is inevitable that variants of such data will end up in the hands of researchers, one way or another. Moreover, A/B testing of experimental and control conditions is now standard practice in technology firms such as Google and Facebook, from the mundane (testing which headline yields the most click-throughs) to the less mundane (examining how exposure to positive or negative updates influences one's own expressed emotions or how knowledge of friends' voting influences the propensity to show up at the polls). Experiments are nearly costless in this environment, in which informed consent is evidently optional.

It is long past time for the field to have a serious conversation about the ethics of big data, and *ASQ* welcomes this conversation (see also George, Haas, and Pentland, 2014). My purpose here is a bit different: what will the advent of big data mean for organization science given the incentives in our publication process that I described above?

Data used to be a constraining factor for organizational research. Gaining access to sufficient data to yield statistically significant results was often difficult, which encouraged researchers to use widely available sources such as Compustat or archives such as ICPSR. Some of the most influential papers in *ASQ* from the 1970s featured simple correlations and occasional regressions on cross-sectional data for modest samples, any of which might be desk-rejected today. Yet the availability of endless data may paradoxically make things worse.

Is an all-you-can-eat buffet likely to yield better research progress? Consider the norms for what counts as a contribution and what gets rewarded. If novelty, statistical significance, and citeability are our guides, then rather than seeing cumulative progress toward answering important questions, we may be headed for disaster. We are almost certain to see studies packed with highly significant and counterintuitive results crying out for citation—but not cumulation.

The Purpose of Organizational Research

In these new circumstances, it is appropriate to ask whose interests our research should serve. Who are the constituencies for organizational research? Answering these questions can guide our answers to what kind of research is worth doing and how we should be structured to do it.

Traditionally, the ultimate constituency for organizational research was managers. Scholars were encouraged to conduct research with ''managerial relevance'' or possibly ''policy relevance.'' From the start, *ASQ* has taken a longer view. The inaugural essay by founding editor James D. Thompson (1956: 102) was clear: ''The pressure for immediately applicable results must be reduced.''*ASQ*'s goal was not to provide tips on best practices for job interviews or how-to guides for creating incentive compensation systems. Research published in *ASQ* was intended to draw on and contribute to basic social science and ''must go beyond description and must be reflected against theory'' (Thompson, 1956: 102). This was not a call for managerial irrelevance; it was staking a claim for understanding our new organizational world.

In the decades after Thompson wrote, business corporations kept growing bigger, and the need for managers to staff their internal hierarchies spawned a massive expansion in management education. General Motors, the quintessential management-heavy corporation, expanded from 600,000 to over 800,000 employees within a few years. The demand for managerially relevant research was evident. Yet beginning in the 1980s, changes in the economy were reflected in the kinds of jobs taken by MBA students. Instead of seeking management jobs at GM or Eastman Kodak or Westinghouse, MBAs from elite schools went into finance and consulting, a shift that in turn empowered finance departments within business schools (see Khurana, 2007). Traditional corporations, particularly manufacturers, shrank or even disappeared through multiple rounds of outsourcing and downsizing, while the largest employers came to be in retail, where hierarchies within stores are relatively short. Meanwhile, information technologies increasingly turn the tasks of management (measuring and rewarding performance, scheduling) over to algorithms.

There are nearly 7 million Americans classified as ''managers,'' but the content of their tasks may not involve the actual supervision of other people.

More recently, alternative business models have arisen that dispense with ''employees'' and ''managers'' entirely. Uber reported that it had 162,000 ''driver-partners'' in the U.S. at the end of 2014. These are not employees of Uber— which itself employed perhaps 2,000 people—but independent contractors without need for management. Amazon expands and shrinks by tens of thousands of workers at a time through the use of temporary staffing companies for its warehouses—it added 80,000 temporary workers for the 2014 holiday season. The tasks are straightforward and largely supervised by computer. Retail, fast food, and the ''sharing economy'' are increasingly moving to a world in which algorithms and platforms replace human management. Meanwhile, GM's North American workforce has shrunk to under 120,000. Management of humans by other humans may be increasingly anachronistic. If managers are not our primary constituency, then who is? Perhaps it is each other. But this might lead us back into the Winchester Mystery House. If our standards of evaluation privilege what is interesting or novel to researchers over what is true, or what is valuable to the public that provides resources, then our sprawling enterprise is unlikely to continue forever.

A final possibility is that our obligation is to society more broadly. In a time of social transformation, when basic units of economy and society are undergoing upheaval with uncertain consequences, perhaps our best bet is to return to the mission laid out by Thompson (1956: 102), with an eye toward the new structures and new processes that are arising. Thompson wrote,

> The unique contribution of science lies in its combination of deductive and inductive methods for the development of reliable knowledge. The methodological problems of the basic sciences are shared by the applied fields. Administrative science will demand a focus on relationships, the use of abstract concepts, and the development of operational definitions. Applied sciences have the further need for criteria of measurement and evaluation. Present abstract concepts of administrative processes must be operationalized and new ones developed or borrowed from the basic social sciences. Available knowledge in scattered sources needs to be assembled and analyzed. Research must go beyond description and must be reflected against theory. It must study the obvious as well as the unknown. The pressure for immediately applicable results must be reduced.

In 1956, the object of administrative science was bureaucratic organizations and their administrators. Today, few companies identify as bureaucracies, and few individuals claim the mantle of ''administrator.'' What will an administrative science look like in a world administered by algorithms? Many aspects of Thompson's vision still hold: the combination of inductive and deductive methods, the use of the tools of basic social science, the benefits of an interdisciplinary orientation (which must surely include connections with information science), and the importance of theory. But it is time to update how we do administrative science and who we do it for.

As *ASQ* approaches its 60th year, we hope to reconsider who the core constituencies of organizational research are and how our field should be structured so that we are more like a cathedral and less like a mystery house. Businesses and governments are making decisions now that will shape the life chances of workers, consumers, and citizens for decades to come. If we want to shape those decisions for public benefit, on the basis of rigorous research, we need to declare ourselves.

Acknowledgments

I thank Joan Friedman, Linda Johanson, and Chris Marquis for their thoughtful suggestions on prior drafts, and apologize for not following more of the good ones.

REFERENCES

Aguinis, H., D. L. Shapiro, E. P. Antonacopoulou, and T. G. Cummings
2014 ''Scholarly impact: A pluralist conceptualization.'' Academy of Management Learning and Education, 13: 623–639.

Cook, T. D., and D. T. Campbell
1979 Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand McNally.

Davis, G. F.
2010 ''Do theories of organizations progress?'' Organizational Research Methods, 13: 690–709.

Davis, G. F.
2014 ''Why do we still have journals?'' Administrative Science Quarterly, 59: 193–201.

Davis, M.
1971 ''That's interesting! Towards a phenomenology of sociology and a sociology of phenomenology.'' Philosophy of the Social Sciences, 1: 309–344.

Denrell, J., and B. Kovács
2008 ''Selective sampling of empirical settings in organizational studies.'' Administrative Science Quarterly, 53: 109–144.

George, G., M. R. Haas, and A. Pentland
2014 ''Big data and management.'' Academy of Management Journal, 57: 321–326.

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

Khurana, R.
2007 From Higher Aims to Hired Hands: The Social Transformation of American Business Schools and the Unfulfilled Promise of Management as a Profession. Princeton, NJ: Princeton University Press.

McCloskey, D.
1998 The Rhetoric of Economics. Madison, WI: University of Wisconsin Press.

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

Pillutla, M. M., and S. Thau
2013 ''Organizational sciences' obsession with 'that's interesting!' Consequences and an alternative.'' Organizational Psychology Review, 3: 187–194.

Thompson, J. D.
1956 ''On building an administrative science.'' Administrative Science Quarterly, 1 (1): 102–111.

Zbaracki, M.
1998 ''The rhetoric and reality of total quality management.'' Administrative Science Quarterly, 43: 602–636.

Author's Biography

Gerald F. Davis, editor of the *Administrative Science Quarterly*, is the Wilbur K. Pierpont Collegiate Professor of Management at the Ross School of Business and a professor of sociology at the University of Michigan, 701 Tappan Street R6362, Ann Arbor, MI 48109-1234 (e-mail: gfdavis@umich.edu). His current research interests include finance and society, new forms of organization, social movements, networks, and corporate governance. He received a Ph.D. in organizational behavior from the Stanford Graduate School of Business.