

## Methodology Corner

# Reordering Our Priorities by Putting Phenomena before Design: Escaping the Straitjacket of Null Hypothesis Significance Testing

Andy Lockett, Abigail McWilliams<sup>1</sup> and David D. Van Fleet<sup>2</sup>

Warwick Business School, Coventry CV4 7AL, UK, <sup>1</sup>College of Business Administration, University of Illinois at Chicago, 601 S. Morgan St (MC 243), Chicago, IL 60607-7123, USA, and <sup>2</sup>Morrison School of Agribusiness, W. P. Carey School of Business, Arizona State University, Mesa, Arizona, USA  
Emails: andy.lockett@wbs.ac.uk; abby@uic.edu; ddvf@asu.edu

**In this paper we reflect on two related questions. First, how have we arrived at a position where null hypothesis significance testing is the dominant criterion employed by quantitative researchers when deciding on whether or not a result is ‘significant’? Second, how might we change the practice of quantitative management research by promoting a greater plurality of methods, and in doing so better enable scholars to put phenomena before design? We conclude by arguing that quantitative management researchers need to focus on the epistemological issues surrounding the role of scholarly reasoning in justifying knowledge claims. By embracing a plurality of approaches to reasoning quantitative researchers will be better able to escape the straitjacket of null hypothesis significance testing and, in doing so, reorder their priorities by putting phenomena before design.**

*Judge a person by their questions, rather than their answers. (Voltaire)*

In this paper we explore the relationship between the phenomena and design of research, with a specific focus on the application of null hypothesis significance testing (NHST) in management research. We do so because the study of the practice of research matters as ‘research findings often say more about a researcher’s practices than the phenomena studied’ (Starbuck, 2004, p. 1). Management researchers face the dilemma that many of our existing practices mimic those in the

natural sciences but still fail to further our understanding of the world (Starbuck, 2004). This places us in the unenviable position of producing studies that may have a detrimental impact on management practice and on future management research if/when the mimicking of natural science methods is inappropriate (Ghoshal, 2005).

NHST is an interesting case in point of this possibility. Originally developed by Fisher to provide an ‘objective’ methodology for experiments in fundamental research, where little was known about potential effects, NHST was designed as a test for whether or not there is an effect; i.e. it is a pure test of the mere existence of an effect. Although NHST has a veneer of scientific rigour, commentators have detailed the problems of employing NHST across a range of different social science disciplines including

---

We would like to thank Saku Mantere for his invaluable advice on early versions of this manuscript, and three anonymous reviewers for their insights and improvements during the review process.

management<sup>1</sup> (Cortina, 2002; Schwab *et al.*, 2011), economics (McCloskey, 1983, 1985; McCloskey and Ziliak, 1996) and psychology (Cohen, 1994). In a recent paper in *Organization Science*, Schwab *et al.* (2011) provided exposition of the problems, which include (i) NHST portrays research findings as clear cut; (ii) NHST enables the validity of findings to be dependent on researchers' efforts to obtain enough data; (iii) NHST can disprove hypotheses that could not possibly be correct; (iv) NHST is difficult to understand and therefore often misinterpreted; (v) NHST highlights trivial findings; (vi) NHST obscures important findings; and (vii) NHST makes assumptions that are not satisfied in most research studies.

The problems of NHST are particularly acute because it is commonly employed as the sole criterion in deciding on whether or not an effect size is (statistically) significant, with no real consideration of whether or not the effect size is substantively significant (i.e. it really matters). The null hypothesis in regression is that the coefficient of a variable is equal to zero ( $\beta = 0$ ), but what does it really mean if the *t* value of a coefficient is significantly different, at some arbitrary level (say 5%), from zero? A correct interpretation of the *t* value of a coefficient may not tell us very much about the importance of a variable. Even if we reject the null this only means that we have some confidence about the direction of the effect; it does not necessarily say anything about the size of the effect. Presumably any empirical researcher is investigating something beyond the existence of an effect – they need to know whether or not the effect size is meaningful. As Tukey (1969) scolded psychologists to remember: 'amount as well as direction is vital'.<sup>2</sup>

<sup>1</sup>This list is not intended to be exhaustive in terms of authors who have written papers that are critical of the use of regression equations in their own discipline/field. It is merely intended to indicate that this problem is one that affects more than research in management. A related discussion surrounds statistical significance versus practical importance (see Dunnette, 1966, pp. 154–6). Numerous measures to help assess the practical significance of effects have been proposed (see Kirk, 1996).

<sup>2</sup>The importance of the size of an effect is particularly important due to the problem of 'crud'. Meehl introduced the issue of the 'crud factor', which states that 'everything is related to everything else' (Meehl, 1990). The problem is that, as Meehl acknowledges, there is no

Given the well documented concerns researchers across different social sciences (paralleling similar concerns in medicine, natural sciences and engineering) have about the use of NSHT (Barnett, 2007; Cortina and Landis, 2011; Gliner, Leech and Morgan, 2002; Nord, 2012; Rogers, 2010; Zyphur and Oswald, 2013), we think it important that management scholars stop and reflect on the current dominance of NHST in management research. Our interest in doing so has been fuelled by the dawning realization that we have historically accepted and employed NHST without reflecting on the philosophical assumptions that accompany the practice. The process of reflection began with an observation that many management scholars use of statistical significance as the sole criterion of the importance of a finding, never even considering whether or not the effect size matters. In addition, over time we became increasingly aware that the use of NHST not only shapes the knowledge claims we make, through the interpretation of our results, it also shapes the nature of questions we ask. In writing this paper our intention is not to criticize quantitative researchers for adherence to an institutionalized practice but to promote reflection about two related questions. First, how have we arrived at a position where NSHT is the dominant criterion employed by quantitative researchers when deciding on whether or not a result is 'significant'? Second, how might we change the practice of quantitative management research by promoting a greater plurality of methods, and in doing so better enable scholars to put phenomena before design?

## Why is NHST so dominant in management research?

A number of commentators have suggested that NHST may have become the dominant practice in quantitative management research (see Barnett, 2007; Carlson and Hatfield, 2004; Mingers, 2006; Orlitzky, 2012; Schwab *et al.*, 2011; Zyphur and Oswald, 2013) for four primary reasons: (1) it is convenient; (2) it rewards large scale data collec-

accurate information about the size of the crud factor in a given research domain. He continues to argue that the assumption that the correlation between arbitrarily paired variables, although not zero, is of minimal importance is surely wrong in many cases (Meehl, 1990).

tion efforts; (3) novelty is valued over replication; and (4) it provides the illusion of scientific rigour.

### *NHST is convenient*

Management researchers, unlike scientists/engineers, seldom deal with constructs that involve a meaningful zero point. Consequently, discussing the substantive significance of a result in a meaningful manner is difficult. Statistical significance provides researchers with a convenient pseudo-scientific framework to justify their methodological decisions and results. Where judgements about substantive significance are complex, scholars are more likely to default to the use of statistical significance because it offers allegedly 'objective, mechanical and clear' decision criteria (Cohen, 1992).

The difficulty of discussing substantive significance is clear from considering the types of survey instruments regularly employed in management research. For example, the standard Likert-style questionnaire format presents standard answers to a question as a series of ratings from low to high on some dimension, with possible responses ranging over some categorical range, for example from 1 to 5 (or 7). The questions elicit an estimation of the perceived intensity of some condition or situation called for in the prompt. Difficulties with substantive significance arise as soon as one asks what it means for an individual to respond on a given variable with a '5' while responding with a '3' on a different but similarly structured variable. Such questions make it clear that the Likert scales commonly used express ordinal (numbers always in the same sequence) rating and not magnitude (where five is always five times the magnitude of one) rating. These ambiguities have prompted a wide variety of validation approaches for assessing different sorts of reliabilities in response patterns. Such validation approaches can help improve the power of a study's results but seldom, if ever, address issues of substantive significance or managerial relevance. The importance of survey results for specific organizations and managers is seldom addressed by academic researchers.

In addition to the problems of measures, the dominance of NHST in management research is promoted by difficulties created by the lack of specific standards for judging substantive significance. The closest we have are the rules of thumb

provided by Cohen (1988) for small, moderate and large effect sizes. However, even the guidelines provided by Cohen will need to vary according to the nature of the problem one is addressing. For example, a large effect size in one context may be viewed as trivial in another and vice versa. Even with available rules of thumb, researchers will need to consider issues of effect size and power in the design of their studies in order to undertake a meaningful analysis of substantive significance once study data have been collected (Sawyer and Ball, 1981). This increases the costs of appropriate study design and thus decreases the likelihood that the study will be undertaken.

Statistical significance seems to provide researchers with a ready-made solution for the problem of assessing the importance of research results which can be applied to all contexts. In effect, it obviates the need for researchers to make a value judgement about the importance of a size of effect (Cortina, 2002). In order to make judgements about effect size, however, scholars need to develop a body of evidence through replication of studies, because systematic replications provide a reliable basis for considering effect sizes in different research settings. This in turn would permit the development of more powerful research designs and support study hypotheses with fairly small changes identified. Many academic communities, however, do not reward replication (Kuhn, 1962), with little replication published in the leading, and even lower status, management journals (Hubbard, Vetter and Little, 1998). Furthermore, where replication has occurred, evidence from management (Hubbard, Vetter and Little, 1998) and economics (Dewald, Thursby and Anderson, 1986) suggests that the replication results have cast doubt on the reliability of the previously published research.

### *NHST rewards large scale data collection efforts*

Incentive structures in management research may work against any change in methodological practice through privileging the reporting of (statistically) significant over non-significant findings and discouraging replication studies (Hubbard, Vetter and Little, 1998). A fixation with large data sets (yes, it appears as though size does matter for many editors – see Rosenthal, 1979) merely leads to an increased probability that results will be statistically significant. This led Starbuck (in

Barnett, 2007, p. 118) to point out that ‘one can always disconfirm a point null hypothesis by gathering enough data. It is a mathematical certainty.’ McCloskey and Ziliak (1996) eloquently summarize this dilemma by arguing that authors need to pay attention to the trade-off between power and the size of the test, and the substantive significance of the power against alternatives. In doing so, researchers need to ensure that the importance of results in terms of statistical significance is not just driven by large samples and their associated power.

Given the publication bias towards significant results, *post hoc* analyses will tend to show that the published studies will have had sufficient statistical power to detect significant effects (i.e. the effect size was sufficient for the size of the sample gathered) and effects not found to be significant were insufficiently large, given the study’s sample size. Any *a priori* estimates of power require the accumulation of reasonable estimates of the size of effect to be examined. Given the publication bias against replication studies, this is difficult to achieve. In the absence of a number of studies reporting the effect of interest, researchers are forced to fall back on making size estimates on the basis of the arbitrary small, medium or large effects outlined by Cohen (1988). The circularity of the whole venture is apparent, as Starbuck commented (2004, p. 18): ‘In a way statistical significance tests are statements about researchers’ willingness to invest effort (in compiling a large sample size) as much as they are about the phenomena studied.’ Without an appreciation of sample power, sample size becomes a decision criterion through which researchers can engineer publishable results.

#### *Novelty is valued over replication*

There are areas of management research whose characteristics are consistent with those of ‘normal science’ (Thomas and Wilson, 2011). In these research areas, organizations of interest typically consist of large numbers of people organized and managed on a regular basis to pursue established business with accepted, well diffused and well understood technologies. Common regulations, policies and procedures govern a large range of business activities. As a result, it is perfectly reasonable to expect that extensive areas of business research in accounting,

finance, human resource management, information systems, organizational behaviour and operations would be amenable to research approaches that focus on large samples, replicability and cumulative results, all with a view towards improving managerial practices and providing research results that are substantively important for practice.

Having said this, it is also the case that management researchers are often biased against viewing themselves as being engaged in ‘normal science’ and are often more comfortable defining their activities in terms of change, innovation, ‘breakthroughs’, ‘silver bullets’ and other discontinuities. Eccles, Nohria and Berkley (1992) note the strong prevalence of ‘hype’ in management research and show how even a casual perusal of major practitioner management journals suggests an atmosphere of almost continual crises in business. Popular business books mirror this bias towards the innovative and unusual or towards responding to the latest business crisis. Business schools are not immune from such biases, and compete with each other to show which is more innovative and leading edge in research (Khurana, 2010).

We suggest that a bias towards what is new, unusual and discontinuous tends to de-emphasize considerations of statistical power and effect size in management research for two main reasons. First, NHST provides researchers with an opportunity to overstate the novelty of their findings, employing the inference that a statistically significant result is important, irrespective of whether or not it is substantively significant. In contrast, the burden of proof required for substantive significance is much greater, making researchers’ claims about the importance of their work much more difficult. In essence, and related to the arguments above about NHST rewarding large scale data collection efforts, NHST increases the probability that researchers will be able to identify ‘novel’ existence effects.

Second, the desire for novelty will result in detailed research histories on topics failing to accumulate as replication is not valued. Furthermore, the leading management journals’ requirement for novelty has become more pronounced over time. For example, in looking at the *Academy of Management Journal* (AMJ), Colquitt and Zapata-Phelan (2007, p. 1291) point out that the information for contributors from



1973 to 1984 ‘referenced the testing of theoretical propositions, while noting that exploratory research and replications were still welcome’. The research note category (which tended to have less theory building) was introduced in 1973 but abandoned in 2005 (Colquitt and Zapata Phelean, 2007). Fast forward to 2014 and we find that while the Academy of Management states that the mission of the *Academy of Management Journal* is to ‘empirically examine theory-based knowledge’, the mission statement for the journal itself states that ‘to be published in *AMJ*, a manuscript must make strong empirical and theoretical contributions . . .’ – replications need not apply (*AMJ*, 2014, 57(1), page prior to page 1).<sup>3</sup>

### *The illusion of scientific rigour*

Some management research areas deal directly with unusual and discontinuous phenomena for which appropriate research designs may not be associated with ‘normal science’. Strategy and entrepreneurship are examples. A core argument in strategy focuses on how firms can behave differently from competitors to obtain an advantage that is often local and temporary (McWilliams and Smart, 1995). Sampling for strategy studies is often inconsistent with conventional norms of study design, since populations are assumed to be heterogeneous and competitors pay attention to each other and thus seldom act independently of each other.

The field of entrepreneurship deals with individuals and small firms that produce new products and practices that redefine entire areas of the economy. From its earliest conceptualization by Schumpeter, entrepreneurship has been concerned with ‘creative destruction’, i.e. transformative innovations, as well as how those innovations are brought to market and commercialized (Schumpeter, 1946). Entrepreneurship research has also developed around the entrepreneur and how such individuals differ from more conventional competitors and managers. As with strat-

egy, it is difficult to view entrepreneurship research in terms of ‘normal science’ and to consider how conventional norms of sampling, effect size and statistical power will apply, since the entire focus of the area is on what is by intention different from the norm, innovative and even disruptive.

One would expect to see strategy and entrepreneurship research characterized by smaller samples and more qualitative research designs, and this does occur. However, for these areas, there is often a countervailing bias to not adopt the study designs that are appropriate for the phenomena being studied and instead focus on research designs more consistent with large sample statistical studies. Arguably, the position is due to a desire for comparable status among business researchers to that enjoyed by more established areas that make claims to scientific rigour, such as business research based strongly in such disciplines as psychology and economics. This bias is often supported by journal editors seeking to boost the rigour of their areas and thus the prestige of their journals. For researchers in these areas the problem manifests itself in the need to publish in the ‘mainstream’ management journals, such as *AMJ*, in order to achieve tenure (de Rond and Miller, 2005).

In strategy research, one can see the results of this bias in cross-industry studies that assess highly generic strategies across widely differing industry contexts with the result that the research says very little about a broad range of phenomena (Schmalensee, 1985). In entrepreneurial studies, this bias is apparent in studies sampling from broad populations of small and medium sized enterprises, even though such populations will often contain large numbers of firms that are not entrepreneurial in their behaviours.

In the discussion above we have explored why NHST is so dominant in management research; however, it also highlights the error of this dominance. Studies should be designed to do the best job possible at explaining the phenomena of interest, i.e. phenomena should precede design. Where designs can make effective use of large samples and replication to accumulate useful results, they should do so. When the phenomena of interest are inconsistent with such statistical approaches, other research designs should be employed. Hence, biases that lead researchers to adopt study designs inconsistent with their phenomena are not

<sup>3</sup>We note, however, that the recent launch of the *Academy of Management Discoveries* is an interesting development in the field. The journal’s mission is ‘to promote the creation and dissemination of new empirical evidence that strengthens our understanding of substantively important yet poorly understood phenomena concerning management and organizations’ (<http://aom.org/amd/>, accessed 10 April 2014).

conducive to the development of management research as a field.

## Pluralism in quantitative research

To promote the reordering of our priorities, by putting phenomena before design, we now explore how management research could change if we began to de-institutionalize the practice of NHST in quantitative research. We suggest, however, that as long as quantitative researchers are wedded to NHST, and the underlying assumption of hypothetico-deductivism, any attempts to shift the status quo will need to move beyond addressing axiological concerns to consider more philosophical issues relating to the epistemology (Meehl, 1997; Orlitzky, 2012) and ontology of scholarly reasoning. In this paper we focus on the epistemological issues surrounding the role of scholarly reasoning in justifying knowledge claims, which resonate with our concern to promote the de-institutionalization of NHST in management research. Furthermore, Mantere and Ketokivi (2013, p. 70) argue that an important missing element from the extant literature is 'methodological – as opposed to rhetorical, psychological, or social – account of scientific reasoning. The missing piece is crucial, because the general understanding of how scientists reason and formulate explanations is surprisingly limited (Lipton, 2004), and yet prescriptive norms are essential in defining criteria for methodological rigor.'<sup>4</sup>

NHST is a hypothetico-deductive method that is predicated on deductive reasoning. Deduction is a logical process which involves moving from one or more premises (general statements) to reach a conclusion. For example, given the premises that (i) all dogs are mortal and (ii) Rover is a dog, then we can conclude that Rover is mortal. For deductive reasoning to be sound the premises

must be true, which then makes the conclusion true by definition. The problem arises when deductive reasoning is employed under NHST because NHST is probabilistic, which leads to 'a misapplication of deductive syllogistic reasoning' (Cohen, 1994, p. 998). For example, the null hypothesis is commonly assumed to be rejected based on the following logic. If the null hypothesis is correct then a statistically significant result cannot occur; therefore, finding a statistically significant result leads us to reject the null. The logic employed is formally correct, but it does not reflect the probabilistic nature of NHST, which should be re-stated as follows. If the null hypothesis is correct then we are unlikely to find a statistically significant result; therefore, finding a statistically significant result suggests that the null hypothesis is highly unlikely. The probabilistic reformulation of deductive reasoning allows for the possibility that the conclusion drawn is false even if the premises are true.

In addition, deductive reasoning does not sit comfortably with the real world (i.e. empirical data) as there is no way for observation and experimentation to test the validity of a premise (Mantere and Ketokivi, 2013). Premises, by definition, remain unproven and unprovable, and therefore must be accepted on face value, or by faith, or for the purpose of exploration. The finding of a statistically significant result may suggest some form of empirical regularity (although the extent to which this is true is questionable given the discussion above), but the explanation of the regularity may be explained using different theoretical lenses. Empirical data by itself cannot confirm any theory from alternative competing theories, which leads to the problem of 'indeterminacy of theory by data' (Quine, 1953). The indeterminacy of theory by data, in part, accounts for why quantitative scholars commonly spend large amounts of time re-framing the front ends of their papers, through the journal review process, as they try to convince editors and reviewers that they have the most compelling theory to explain their results.

The discussion above prompts the question: what new knowledge can be claimed on the basis of a statistically significant result? Under NHST, premises are deemed to be either true or false; however, such absolute language is problematic in management (and the social sciences more generally) because, rather than merely accepting or

<sup>4</sup>Issues of ontology are also important but lie outside the scope of this paper. For example, Laing (1967) highlights ontological concerns in that some social sciences are characterized by 'natural scientism', which is the error of turning people into things through a process of reification, and so 'results derived in this way have to be dequantified and dereified before they can be reassimilated into the realm of human discourse. The error fundamentally is the failure to realize that there is an ontological discontinuity between human beings and it-beings' (Laing, 1967, p. 53).

rejecting a hypothesis, ‘the primary aim of a scientific experiment is to make appropriate adjustments in the degree to which one accepts, or believes, the hypothesis or hypotheses being tested’ (Rozeboom, 1960, p. 420). Consequently: ‘The end product of a scientific investigation is a degree of confidence in some set of propositions, which then constitutes a basis for decisions’ (Rozeboom, 1960, p. 423). In essence, NHST may lead scholars to make misleading knowledge claims since ‘NHST was a creation of and for statisticians’ (Rogers, 2010, p. 341), and one we cannot hide behind. Rather, scholars need to accept that the extent to which they have confidence in a hypothesis constitutes a cognitive process, which ‘has an unpleasantly vague, subjective feel about it which makes it unpalatable for inclusion in a formalized theory of inference’ (Rozeboom, 1960, p. 421). In interpreting results researchers are required to make a subjective assessment about how they are able to interpret the results, in the context of other evidence, in arriving at a commentary about the analysis.

If we are to achieve Orlitzky’s (2012) aim of de-institutionalizing NHST, we suggest that a first step is to promote debate amongst quantitative management researchers around the different methods of scientific reasoning that may be employed when making knowledge claims from quantitative data. Opening up debate would promote a greater degree of methodological pluralism, which to date has been much more prevalent in qualitative research (see Johnson *et al.*, 2006 for a discussion). Furthermore, the dominance of NHST, and the limited engagement of quantitative scholars with different methods of scientific reasoning from data, has arguably exacerbated the separation of qualitative and quantitative research, with many qualitative management researchers viewing quantitative research as being illegitimately institutionalized (Symon *et al.*, 2008). We feel that the oppositional tone between qualitative and quantitative researchers, for which both are not without blame, is unhelpful. Further, we suggest that by focusing on the role of scientific reasoning in knowledge claims we may be able to narrow the divide between the two.

### *Deduction, induction and abduction*

In moving to a position of quantitative pluralism scholars may adopt a number of different forms

of scientific reasoning in making knowledge claims (Mantere and Ketokivi, 2013). The two dominant forms are deduction (as outlined above) and induction, but more recently scholars have taken an interest in abduction as a way of addressing the limitations of deduction and induction.<sup>5</sup>

Operating within the deductive camp, a number of scholars have sought to address the limitations of NHST through proposing more sophisticated forms of hypothetico-deductivism. Rather than adhering to the idea that a testing of an individual hypothesis can determine the fate of a whole theory, more holistic forms of hypothetico-deductivism (Burawoy, 1990; Lakatos, 1970) focus on a progressive research programme whereby ‘the new belts of theory expand the empirical content of the program, not only by absorbing anomalies but by making predictions, some of which are corroborated’ (Burawoy, 1990, p. 778). For the sophisticated hypothetico-deductivist ‘a theory is deemed “acceptable” or “scientific” only if it has corroborated excess empirical content over its predecessor or rival, that is, only if it leads to the discovery of more novel facts’ (Lakatos, 1970, p. 182). In moving away from the restrictive model of hypothetico-deductivism, scholars are required to engage in a shift in scientific reasoning because subjective assessments about what constitutes ‘excess empirical content’ to deem a theory ‘acceptable’ over a rival theory need to be stated. Consequently, more sophisticated forms of hypothetico-deductivism start to draw on additional forms of scientific reasoning rather than only deduction.

Inductive reasoning, in contrast to deductive reasoning where the truth of the premises guarantees the truth of the conclusion, is ampliative in nature, as the meaning and the conclusion go beyond what is (logically) contained in the premises (i.e. there are no necessary influences) (Hájek and Hall, 2002). Hence, and as first highlighted by David Hume (1888), the problem of inductive reasoning is the jump between empirical data and theoretical generalizations (the indeterminacy of data to theory) because any empirical regularity may be explained using multiple generalizations (see Ketokivi and Mantere, 2010 for a discussion in the context of organization research). Interestingly, the problem of inductive reasoning, in that

<sup>5</sup>Interestingly, abductive reasoning is a notable absentee from the vast majority of doctoral training programmes.

data cannot generate theory in itself, is the reverse of the problem of deduction in that data cannot prove theory. Both are rooted in the separation of theory and observation.

Another form of ampliative reasoning is abduction, whose modern usage is attributed to Charles Peirce (1839–1914), the ‘father of pragmatism’. In contrast to deductive and inductive reasoning, abductive reasoning is a more pragmatic form of reasoning that embraces the idea of the indeterminacy of data to theory. Peirce argued that whereas inductive reasoning requires a fairly full body of evidence from which to draw inference, abductive reasoning is characterized by lack of completeness, either in the evidence, or in the explanation, or both. The abductive process starts with an incomplete set of observations, but then the researcher works towards identifying the most likely explanation (i.e. which theory best explains the observations). Abduction is particularly well suited to exploring complex management issues for which there is incomplete data, and is considered to be the main model of reasoning in medical diagnosis. Doctors tend to diagnose according to the hypothesis that best explains the patient’s symptoms (Josephson and Josephson, 1994).

To date, management research has become rather siloed, with research designs being aligned to specific types of scientific reasoning (Mantere and Ketokivi, 2013). For example, theory testing is based on a deductive style of research and tends to be based on quantitative data (Rumelt, Schendel and Teece, 1991), theory building based on qualitative data is inductive (Eisenhardt and Graebner, 2007), and interpretive scholarship is abductive (Hatch and Yanow, 2003). There is no reason, however, why specific research designs need to be aligned with specific forms of scientific reasoning. Mantere and Ketokivi (2013, p. 76) argue that ‘A closer look at research practice reveals that researchers across research traditions use all three forms of reasoning. It is hardly surprising to observe that we all make inferences to a case (use deduction), inferences to generalizations (use induction), and inferences to explanations (use abduction). Thus, using reasoning types as labels to describe entire research designs is misleading.’

We suggest that if quantitative management researchers question the basis of the knowledge claims they make, and open themselves up to employing different forms of scientific reasoning, they may be better able to reorder their priorities

and put phenomena before design. In many ways quantitative researchers already re-craft their ‘deductive’ models to fit their findings (and vice versa) in a manner that employs aspects of induction and abduction; however, the iterative nature of quantitative research is seldom acknowledged in journal articles. Being more aware of the nature (and thereby extent) of the knowledge claims we make may open up journals to publish papers that report interesting empirical ‘facts’, allowing subsequent researchers to then direct their efforts at the understanding of why and how those facts came to be (Helfat, 2007; Miller, 2007). For example, Hambrick (2007) suggests that Schmalensee’s (1985) paper in the *American Economic Review* is an interesting case in point, being ‘a straightforward, unvarnished exercise in fact-finding, but one that spawned an immensely important and influential stream of research in economics and strategy’ (Hambrick, 2007, p. 1348). Furthermore, embracing multiple forms of scientific reasoning will require authors to be much more explicit about the basis of their knowledge claims and the commonly unacknowledged subjective inference that is contained within them.

## Putting phenomena before design

Our intention in writing this paper has been to highlight that the methods we use not only influence the inferences we draw from our data, they also affect the nature of the questions we ask. If generating a large sample size is an important factor in determining the probability of a paper being published in a top-rated journal, then many interesting issues, where large scale data sets cannot be constructed, are less likely to be published. Furthermore, even if we assume that all papers that employ NHST address interesting research questions, the impact of the research may be significantly undermined because the results that are deemed important under the criterion of statistical significance may well be trivial in terms of substantive significance. In essence, the corollary of Hambrick’s (2007) argument about the dominance of NHST meaning that interesting facts cannot be published without theory is that it may also be promoting the publication of uninteresting and/or trivial facts. In either case, putting NHST design before phenomena reduces the impact of much scholarly activity.



The concern for ‘impact’ has important implications for the development of UK higher education, particularly with respect to promoting scholarship that is relevant (Hodgkinson and Starkey, 2011; Starkey and Madan, 2001). For example, the Research Excellence Framework (REF) 2014 saw the agenda evolve from a reliance on academics’ ‘Best 4’ publications to incorporate an assessment of the ‘impact’ of research, which is likely to become even more important in REF 2020. We feel that the shift of emphasis in the REF, towards a greater weighting of impact, may be an ideal time for the *British Journal of Management* to promote debate on two related issues.

First, axiological issues could be addressed through developing a code of best practice for quantitative analysis, which is something that has been implemented in a limited number of journals in other disciplines (e.g. *American Psychologist*, *Journal of Socio-Economics*, *Annals of Internal Medicine* and others). To date, the *BJM*’s guidelines to authors only provide guidance about the presentation of results pertaining to statistical significance, thereby indicating that statistical significance is the sole criterion for judging the significance of a result.<sup>6</sup> We suggest that a code of best practice should require quantitative researchers to determine and explain the importance of their results. Whilst we acknowledge that discussion of meaningful effect size is complex, we suggest that the act of engendering discussion of the meaningfulness of effect size will stimulate debate about the impact of quantitative research.

Second, and arguably of greater importance, would be to open up debate about the different forms of scientific reasoning that may be employed when making knowledge claims from quantitative (and qualitative) data. As Cohen warned, ‘don’t look for a magic alternative to NHST, some other objective mechanical ritual to replace it. It doesn’t exist’ (Cohen, 1994, p. 1001). We implore *BJM* to encourage scholars to use different forms of reasoning with quantitative data, and to be explicit about their approaches in their papers. If achieved, we suggest that it will help researchers escape the straitjacket of NHST and, in doing so, enable them to reorder their priorities and put phenomena before design.

## References

- Barnett, M. (2007). ‘(Un)learning and (mis)education through the eyes of Bill Starbuck: an interview with Pandora’s playmate’, *Academy of Management Learning and Education*, **6**, pp. 114–127.
- Burawoy, M. (1990). ‘Marxism as science’, *American Sociological Review*, **55**, pp. 775–793.
- Carlson, K. D. and D. E. Hatfield (2004). ‘Strategic management research and the cumulative knowledge perspective’, *Research Methodology in Strategy and Management*, **1**, pp. 273–301.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). ‘A power primer’, *Quantitative Methods in Psychology*, **112**, pp. 155–159.
- Cohen, J. (1994). ‘The earth is round ( $p < .05$ )’, *American Psychologist*, **49**, pp. 997–1003.
- Colquitt, J. A. and C. P. Zapata-Phelan (2007). ‘Trends in theory building and theory testing: a five-decade study of the Academy of Management Journal’, *Academy of Management Journal*, **50**, pp. 1281–1303.
- Cortina, J. M. (2002). ‘Big things have small beginnings: an assortment of “minor” methodological misunderstandings’, *Journal of Management*, **28**, pp. 339–362.
- Cortina, J. M. and R. S. Landis (2011). ‘The earth is not round ( $p = .00$ )’, *Organizational Research Methods*, **14**, pp. 332–349.
- Dewald, W. G., J. G. Thursby and R. G. Anderson (1986). ‘Replication in empirical economics: the journal of money credit and banking project’, *American Economic Review*, **76**, pp. 587–603.
- Dunnette, M. D. (1966). *Personnel Selection and Placement*. Belmont, CA: Wadsworth.
- Eccles, R. G., N. Nohria and J. D. Berkley (1992). *Beyond the Hype: Rediscovering the Essence of Management*. Boston, MA: Harvard Business School Press.
- Eisenhardt, K. M. and M. E. Graebner (2007). ‘Theory building from cases: opportunities and challenges’, *Academy of Management Journal*, **50**, pp. 25–32.
- Ghoshal, S. (2005). ‘Bad management theories are destroying good management practices’, *Academy of Management Learning and Education*, **4**, pp. 75–91.
- Gliner, J., N. Leech and G. Morgan (2002). ‘Problems with null hypothesis significance testing (NHST): what do the textbooks say?’, *Journal of Experimental Education*, **71**, pp. 83–92.
- Hájek, A. and N. Hall (2002). ‘Induction and probability’. In P. Machamer and M. Silberstein (eds), *The Blackwell Guide to the Philosophy of Science*, pp. 149–172. Malden, MA: Blackwell.
- Hambrick, D. C. (2007). ‘The field of management’s devotion to theory: too much of a good thing?’, *Academy of Management Journal*, **50**, pp. 1346–1352.
- Hatch, M. J. and D. Yanow (2003). ‘Organization theory as an interpretive science’. In H. Tsoukas and C. Knudsen (eds), *The Oxford Handbook of Organization Theory*, pp. 63–87. Oxford: Oxford University Press.
- Helfat, C. E. (2007). ‘Stylized facts, empirical research and theory development in management’, *Strategic Organization*, **5**, pp. 185–192.
- Hodgkinson, G. and K. Starkey (2011). ‘Not simply returning to the same answer over and over again: reframing relevance’, *British Journal of Management*, **22**, pp. 355–369.

<sup>6</sup>[http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1467-8551/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-8551/homepage/ForAuthors.html), accessed 28 March 2014.

- Hubbard, R., D. E. Vetter and E. L. Little (1998). 'Replication in strategic management: scientific testing for validity, generalizability, and usefulness', *Strategic Management Journal*, **19**, pp. 243–254.
- Hume, D. (1888). *Hume's Treatise of Human Nature*, ed. by L. A. Selby Bigge. Oxford: Clarendon Press. Originally published in 1739–40.
- Johnson, P., A. Buehring, C. Cassell and G. Symon (2006). 'Evaluating qualitative management research: towards a contingent criteriology', *International Journal of Management Reviews*, **8**, pp. 131–156.
- Josephson, J. R. and S. G. Josephson (eds) (1994). *Abductive Inference: Computation, Philosophy, Technology*. Cambridge: Cambridge University Press.
- Ketokivi, M. and S. Mantere (2010). 'Two strategies for inductive reasoning in organizational research', *Academy of Management Review*, **35**, pp. 315–333.
- Khurana, R. (2010). *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.
- Kirk, R. E. (1996). 'Practical significance: a concept whose time has come', *Educational and Psychological Measurement*, **56**, pp. 746–759.
- Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Laing, R. D. (1967). *The Politics of Experience and the Bird of Paradise*. London: Penguin.
- Lakatos, I. (1970). 'Falsification and the methodology of scientific research programmes'. In I. Lakatos and A. Musgrave (eds), *Criticism and the Growth of Knowledge*, pp. 91–196. Cambridge: Cambridge University Press.
- Lipton, P. (2004). *Inference to the Best Explanation*. London: Psychology Press.
- Mantere, S. and M. Ketokivi (2013). 'Reasoning in organization science', *Academy of Management Review*, **38**, pp. 70–89.
- McCloskey, D. N. (1983). 'The rhetoric of economics', *Journal of Economic Literature*, **21**, pp. 481–517.
- McCloskey, D. N. (1985). 'The loss function has been mislaid: the rhetoric of significance tests', *American Economic Review*, **75**, pp. 201–205.
- McCloskey, D. N. and S. T. Ziliak (1996). 'The standard error of regressions', *Journal of Economic Literature*, **34**, pp. 97–114.
- McWilliams, A. and D. L. Smart (1995). 'The resource-based view of the firm: does it go far enough in shedding the assumptions of the SCP paradigm?', *Journal of Management Inquiry*, **4**, pp. 309–316.
- Meehl, P. E. (1990). 'Why summaries of research on psychological theories are often uninterpretable', *Psychological Reports*, **66**, pp. 195–244.
- Meehl, P. E. (1997). 'The problem is epistemology, not statistics: replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions'. In L. L. Harlow, S. A. Mulaik and J. H. Steiger (eds), *What If There Were No Significance Tests?*, pp. 393–425. Mahwah, NJ: Erlbaum.
- Miller, D. (2007). 'Paradigm prison, or in praise of atheoretic research', *Strategic Organization*, **5**, pp. 177–184.
- Mingers, J. (2006). 'A critique of statistical modelling in management science from a critical realist perspective: its role within multimethodology', *Journal of the Operational Research Society*, **57**, pp. 202–219.
- Nord, W. (2012). 'On doing the wrong things for the wrong reasons: two misguided organization studies practices', *Journal of Management Inquiry*, **21**, pp. 443–447.
- Orlitzky, M. (2012). 'How can significance tests be deinstitutionalized?', *Organizational Research Methods*, **15**, pp. 199–228.
- Quine, W. (1953). 'Two dogmas of empiricism'. In W. Quine (ed.), *From a Logical Point of View*, pp. 20–46. Cambridge, MA: Harvard University Press.
- Rogers, J. (2010). 'Statistical and mathematical modeling versus NHST? There is no competition', *Journal of Modern Applied Statistical Methods*, **9**, pp. 340–347.
- de Rond, M. and A. Miller (2005). 'Publish or perish: bane or boon of academic life?', *Journal of Management Inquiry*, **14**, pp. 321–329.
- Rosenthal, R. (1979). 'The "file draw problem" and tolerance of null results', *Psychological Bulletin*, **86**, pp. 638–641.
- Rozeboom, W. W. (1960). 'The fallacy of the null-hypothesis significance test', *Psychological Bulletin*, **57**, pp. 416–428.
- Rumelt, R. P., D. Schendel and D. J. Teece (1991). 'Strategic management and economics', *Strategic Management Journal*, **12**, pp. 5–29.
- Sawyer, A. G. and A. D. Ball (1981). 'Statistical power and effect size in marketing research', *Journal of Marketing Research*, **18**, pp. 275–290.
- Schmalensee, R. (1985). 'Do markets differ much?', *American Economic Review*, **75**, pp. 341–351.
- Schumpeter, J. A. (1946). 'Comments on a plan for the study of entrepreneurship'. Reprinted in R. Swedberg (ed.), *Joseph A. Schumpeter: The Economics and Sociology of Capitalism*, pp. 406–428. Princeton, NJ: Princeton University Press.
- Schwab, A., E. Abrahamson, W. H. Starbuck and F. Fidler (2011). 'Perspective – Researchers should make thoughtful assessments instead of null-hypothesis significance tests', *Organization Science*, **22**, pp. 1105–1120.
- Starbuck, W. H. (2004). *The Production of Knowledge: The Challenge of Social Science Research*. Oxford: Oxford University Press.
- Starkey, K. and P. Madan (2001). 'Bridging the relevance gap: aligning stakeholders in the future of management research', *British Journal of Management*, **12**, pp. S3–S26.
- Symon, G., A. Buehring, P. Johnson and C. Cassell (2008). 'Positioning qualitative research as resistance to the institutionalization of the academic labour process', *Organization Studies*, **29**, pp. 1315–1336.
- Thomas, H. and A. D. Wilson (2011). '"Physics envy", cognitive legitimacy or practical relevance: dilemmas in the evolution of management research in the UK', *British Journal of Management*, **22**, pp. 443–456.
- Tukey, J. W. (1969). 'Analyzing data: sanctification or detective work?', *American Psychologist*, **24**, pp. 83–91.
- Zyphur, M. and F. Oswald (2013). 'Bayesian estimation and inference: a user's guide', *Journal of Management*, doi: 10.1177/0149206313501200.

Andy Lockett is Professor of Strategy and Entrepreneurship, and Pro Dean, at Warwick Business School. His research interests span the interface of strategy and entrepreneurship focusing on new and established ventures in both the private and public sectors.

Abigail McWilliams is Professor of Strategic Management in the University of Illinois – Chicago Business School. Her primary research interest is in the interaction of social responsibility, sustainability and enduring competitive advantage.

David Van Fleet is Professor of Management in the Morrison School of Agribusiness, W.P. Carey School of Business at Arizona State University. His research interests include the history of management thought, leadership, and workplace violence.

Copyright of British Journal of Management is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.