

Institutionalized dualism: statistical significance testing as myth and ceremony

Marc Orlitzky

Published online: 13 August 2011
© Springer Verlag 2011

Abstract Several well-known statisticians regard significance testing as a deeply problematic procedure in statistical inference. Yet, in-depth discussion of null hypothesis significance testing (NHST) has largely been absent from the literature on organizations or, more specifically, management control systems. This article attempts to redress this oversight by drawing on neoinstitutional theory to frame, analyze, and explore the NHST problem. Regulative, normative, and cultural-cognitive forces partly explain the longevity of NHST in organization studies. The unintended negative consequences of NHST include a reinforcement of the academic-practitioner divide, an obstacle to the growth of knowledge, discouragement of study replications, and mechanization of researcher decision making. An appreciation of these institutional explanations for NHST as well as the harm caused by NHST may ultimately help researchers develop superior methodological alternatives to a controversial statistical technique.

Keywords Epistemology · Neoinstitutional theory · Null hypothesis significance testing · Quantitative methods · Sociology of science · Statistical significance test

“A science institutionalizes nonideological modes of observing, or ‘objectivity’.”

(Fuchs 2001, p. 34)

Statistical significance testing is essential to the growth of scientific knowledge in organization studies (OS). Over time, relying on significance tests, organizational researchers can decide whether particular theoretical propositions are true or false.

M. Orlitzky (✉)
The Pennsylvania State University Altoona, 214 Hawthorn, 3000 Ivyside Drive, Altoona,
PA 16601, USA
e-mail: moo3@psu.edu

If the results are not significant, researchers can conclude that the findings probably are only a chance occurrence and that the true result is a null effect. If the results are significant, then we know we have a reliable finding. The p value for a given significance test provides information on (1) the practical significance and (2) the replicability of a finding. Hence, significance testing is the *sine qua non* of objective, quantitative theory testing in OS. Without significance tests, organizational theories would be impossible to falsify and, thus, unscientific or pseudoscientific.

Arguments like those mentioned in the introductory paragraph (adapted from Schmidt 1996, p. 126) typically constitute the gist of the justification of null hypothesis significance testing (NHST), a dichotomous statistical inference technique for evaluating research hypotheses by assessing the probability of the observed data, given the null hypothesis is true (Cohen 1994). *Yet, every single sentence in the first paragraph of this paper is actually an overoptimistic misattribution to NHST* (Schmidt 1996; Schmidt and Hunter 1997). Thus, what on the surface appears to be a reasonable, objective data analysis tool can be shown to rest on myth and fantasy (Carver 1978; Malgady 2000). Yet, despite the growing methodological criticism of NHST over the last 70 years, which various researchers have documented (e.g., Anderson et al. 2000, p. 913; Harlow et al. 1997; Kirk 1996; Kline 2004; Schmidt 1996; Ziliak and McCloskey 2008), NHST has so far proved impervious to the controversy surrounding it. The persistence of NHST suggests this statistical technique has become institutionalized among quantitative researchers studying organizations and, more specifically, management control systems (see also Chenhall 2003). In fact, NHST has become such a taken-for-granted technique in the empirical validation of OS theory that quantitative researchers have generally embraced it uncritically and almost universally equated theory testing and NHST improperly (Hedges 1987; Schmidt and Hunter 1997). Because of the pervasiveness of NHST in quantitative theory testing in all behavioral and social sciences (e.g., Tracey 2000), some observers (e.g., Kaufman 1998; Kline 2004) predicted that the controversy over NHST could become the defining methodological issue of this generation of social and organizational scientists.

The primary goal of this article is to propose an institutional theory of the longevity of NHST, in which NHST is considered to be a statistical technique deeply embedded in the history of OS. This article presents a set of suppositions and conjectures explaining (1) the persistence of NHST in the social and organizational sciences and (2) its generally detrimental impact on theory development and the growth of knowledge. First, we must understand why the community of quantitative researchers continues to apply NHST in quantitative testing of theory, even though there are so few valid defenses of it. Methodological explanations must be insufficient because otherwise rational¹ scholars would long have abandoned this statistical method whose disadvantages outweigh its benefits (Cohen 1994; Schmidt and Hunter 2002). To my best knowledge, this is the first article in any social science discipline that analyzes NHST from a neoinstitutional perspective, which is

¹Rationality can be defined in a variety of ways (Scott 1995), but throughout the article, I follow the conventional definition from the OS literature: “instrumental behavior taken in order to reach desired ends” (Scott 1995, p. 138), which in this context is the empirical testing of theory.

able to surface the broader structural sources of problems that frequently go unnoticed (Pearce 2004).

Second, given the context of this inaugural issue of the *Journal of Management Control*, I will illustrate how NHST impedes theory development and knowledge growth by instilling a false sense of certainty about primary quantitative studies of organizations and management control systems (MCS). As the aim of research on MCS is to discover, describe, explain, and predict antecedents, contingencies, and outcomes of management accounting, control and governance, knowledge growth necessarily depends on the methods that we use to make those discoveries, descriptions, explanations, and predictions. Because NHST can be shown to impede theory testing and development in OS and MCS, it arguably fails OS and MCS on several counts. The specific negative consequences for OS will be pointed out toward the end of this article. In this context, OS includes the fields of organizational behavior, human resource management, strategy, organization theory, and many of the other business disciplines recognized by the Academy of Management. Because OS and studies of MCS have the same disciplinary roots in the social sciences (Birnberg et al. 2007; Bromwich 2007; Chenhall 2007; Hesford et al. 2007; Lambert 2007; Miller 2007), the explanations and propositions developed herein equally apply to quantitative research on MCS.

To pursue the two interrelated objectives, the article is structured as follows. First, it places the major problems of NHST (as identified by prominent statisticians) in an OS or MCS context in order to indicate the gravity of the situation. Second, it draws on a neoinstitutional perspective to suggest various social forces inside and outside of OS and MCS that may, at least partly, explain the continuing adherence to NHST. Third, the article illustrates the various unintended negative consequences with select concrete examples. Figure 1 summarizes the argument contained in the second and third sections, the core contributions of this article.

At this point, it ought to be noted that, in several aspects, the arguments in this article are similar to those of Meyer and Rowan's (1977) seminal article. Over 30 years ago, Meyer and Rowan revolutionized OS by unmasking its overemphasis on rational coordination and control. They showed how rationalized institutional rules might serve as myths giving meaning to organizational activities in a wider institutional arena. The puzzle I seek to solve in this paper is a much smaller one: the persistence of NHST despite all the rational methodological critiques of this theory-testing procedure. Inspired by Meyer and Rowan's article (including its title), I argue for a neoinstitutional view of NHST as "pseudoscientific ceremony" (McCloskey 1998, p. 115). The dualistic, or binary, coding of empirical evidence as *significant* and *non-significant*, so widely institutionalized in the social and organizational sciences, is behind the title of this paper. The broader epistemological sources of this dualism will be described in greater detail later, in the section on cultural-cognitive forces and the theoretical build-up to Proposition 4.

1 The problem of significance testing

This article rests on the assumption that, from a methodological perspective, NHST involves net costs rather than net benefits (Kline 2004; Schmidt and Hunter 2002;

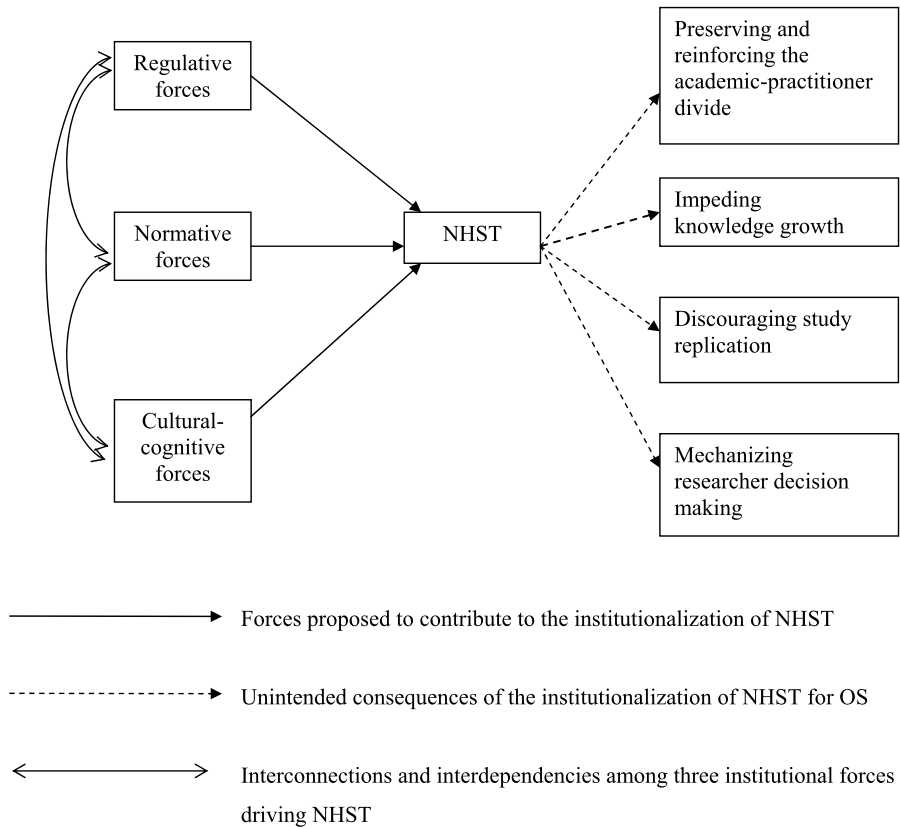


Fig. 1 The institutional embeddedness of Null Hypothesis Significance Testing (NHST)

Ziliak and McCloskey 2008). Over the years, experts in quantitative methodology have identified over 15 problems inherent in NHST (Cohen 1994; Kline 2004; Schmidt 1996). For example, one major problem is that NHST trades off the control of more important statistical error rates—Type II error—for arguably less important ones—Type I error (Cohen 1994; Schmidt and Hunter 2002). Type I error is captured by the familiar alpha probability level, which researchers use to minimize the likelihood of false positives. That is, the smaller the alpha level, the lower the probability that observations show relationships that, in reality, do not exist. In contrast, a Type II error is the erroneous acceptance of a false null hypothesis, that is, the conclusion that the data do not show relationships that really do exist. NHST only controls Type I error rates (alpha levels), but not Type II error rates; that is, NHST does not maximize statistical power, which is defined as $1 - \beta$, where β refers to the Type II error rate (Cohen 1988, 1992). In fact, there is a well-known tradeoff between alpha and beta levels, so that, all else equal, the greater the concern with minimizing Type I errors, the higher the probability that the study may have low statistical power and, thus, fail to detect important relationships (Cohen 1994). For this reason, NHST may result in

the improper acceptance of the null hypothesis (Fisher 1935, p. 19; Schmidt 1996) because it ignores and, arguably, aggravates problems of low statistical power.

In addition, there are various other problems associated with NHST. For example, NHST typically is misapplied to nonprobability samples (Kline 2004; Pedhazur and Schmelkin 1991). Moreover, it leads to misinterpretations of the probability level p as an indicator of the likelihood of replication (Cohen 1994), practical significance (Campbell 1982), or the truth of the research hypothesis (Kline 2004; Oakes 1986). NHST also contains the deductive syllogism of *modus tollens*, which is invalid if applied to probabilistic statements (Hacking 1965; Hofmann 2002; Royall 1997; Sober 2005)² and, thus, lulls quantitative researchers in a false sense of epistemic certainty by leaving them with the “illusion of attaining improbability” (Cohen 1994, p. 998). These and other inherent problems with NHST have been reviewed by several experts in quantitative research methodology (e.g., Cohen 1990, 1994; Guttman 1985; Harlow et al. 1997; Kline 2004; Morrison and Henkel 1970; Nickerson 2000; Schmidt 1996). Therefore, yet another in-depth methodological review seems superfluous because the NHST-inherent methodological problems are the same in quantitative investigations of organizations and MCS as in psychology or economics, for example.

But aren't there pragmatic reasons explaining the continued use of NHST? This is exactly the argument put forth by the defenders of this statistical practice, who often claim that there really are no inherent problems with significance testing, only misinterpretations (e.g., Abelson 1997; Hagen 1997; Krueger 2001; Wainer 1999). These counter-arguments underestimate the extent and gravity of the problem. The aforementioned mistakes have been shown to be very common among social scientists and statisticians (Lecoutre et al. 2003; McCloskey 1998; Nunnally 1975, p. 195; Schmidt 1996). Oakes (1986) demonstrated that even well-trained and widely published researchers misinterpreted the output of NHST. Upon closer inspection, the NHST defenders' arguments can be shown to contain many of the myths that the critics of NHST have felt compelled to combat (see, e.g., Schmidt 1996, pp. 124–126; Schmidt and Hunter 1997).

The harm caused by NHST is significant for theory development and testing in OS. For one, NHST reinforces the academic-practitioner divide. Arguably even more important, NHST impedes the growth of knowledge in OS (Schmidt 1996), primarily because it leaves quantitative OS researchers with a false sense of certainty about their findings in primary studies as it suggests that there is such a thing as a critical study to decide the fate of OS theories (Hunter and Schmidt 2004). Furthermore,

²Valid *modus tollens*:

A1: If P , then Q .

A2: Not Q .

C: Hence, not P .

Invalid probabilistic *modus tollens*:

A1: If organizations use clan and informal controls, then their financial performance will increase.

A2: The performance of the organizations in our sample did not increase.

C: Hence, the organizations in our sample probably did not use clan and informal controls.

See also Cohen (1994) and Hofmann (2002) for further examples and explanations of this logical invalidity.

it discourages study replications and mechanizes researcher decision making in OS (Kline 2004). In short, from a sociology-of-science angle, NHST represents a “rhetorical scandal” (McCloskey 1998, p. 187) as it creates a mirage of objectivity (Labovitz 1972; Ziliak and McCloskey 2008). These undesirable effects of NHST will be explored and illustrated later—in the section on the unintended consequences of NHST.

Because NHST is as endemic in studies of organizations, MCS, and corporate governance as in the other social and behavioral sciences, the problems associated with it can no longer be ignored. NHST’s amazing run is partly due to the remarkable increase in quantitative studies, whose proportion has approximately doubled between the mid-1950s and now (Van Maanen 1998a). If we assume that quantitative articles constitute 83 percent of all empirical articles in premier management journals and if we further assume that 95 percent of quantitative organizational researchers rely on NHST to test their hypotheses (Kline 2004, p. 9), then 79 percent of all empirical articles published in academic management journals use the NHST in one form or another (e.g., in the context of *chi-square* tests, *t* tests, *F* tests, or goodness-of-fit tests).³ This estimate of 79 percent is conservative because, according to other estimates, NHST is used in 94 percent of all articles (Hubbard et al. 1997). My own review of the *Academy of Management Journal* found that in 2003, for example, about 6 percent of all articles were qualitative, and of the remaining 94 percent of quantitative articles, all relied on NHST. These estimates indicate that, consistent with reviews in psychology (Krueger 2001; Nickerson 2000), it might be difficult to find another theory validation procedure more widespread in OS journals today. It can safely be concluded that most quantitative normal-science activity in OS still relies on NHST as a firmly institutionalized methodological practice. Although, based on statisticians’ strident attacks on NHST, “null hypothesis testing should be dead” (Rindskopf 1997, p. 319), the premier journals do not seem to have implemented any reforms (Fidler et al. 2004; Finch et al. 2001). If, according to one researcher, NHST is “surely the most bone-headedly misguided procedure ever institutionalized in the rote training of science students” (Rozeboom 1997, p. 335), the central issue in this debate is not so much the institutionalization of NHST, but the failure of its deinstitutionalization.

In a way, the lack of deinstitutionalization is an outcome of the interdependence between theory and methods. Although scholars disagree about the characteristics of “good” theory (DiMaggio 1995; Sutton and Staw 1995; Weick 1995) or the value of a predominant paradigm in OS (Pfeffer 1993; Van Maanen 1995b), NHST has remained the one methodological constant in any quantitative testing of organizational theory. Granted, there are several criteria to judge the quality of a theory (Weick 1989). However, in practice, empirical validation seems to trump all other criteria. The likelihood that any OS theory will survive without empirical validation is small, which reviews of the most popular organizational theories can con-

³This percentage was estimated from Van Maanen’s (1998a) introduction to qualitative research in *Administrative Science Quarterly*. The 1986–1996 column shows 178 quantitative articles and the total number of empirical articles in that ten-year period was 36 (qualitative) + 178. A study of the *Journal of Applied Psychology* (Hubbard et al. 1997) and my own review of the *Academy of Management Journal* suggest that these estimates of the total number of quantitative articles (83%) and the number of NHST in those articles (95%) are most likely underestimates.

firm (Miner 2003). Given the perceived centrality of NHST for theory validation in quantitative research on organizations and MCS, especially in those areas derived from economics, sociology, and psychology (Birnberg et al. 2007; Bromwich 2007; Miller 2007), it is no surprise that researchers have not dropped the tool of NHST (see Weick 1996 on the importance of tool dropping).

2 Explaining the institution of statistical significance testing

A self-reflective understanding of the institutional impact of NHST requires an analytic approach that is broader than the statistical lens that predominated in previous critiques of NHST outside of OS. For one, NHST needs to be understood from an epistemological perspective, which is important work begun by Meehl (1997) and Rozeboom (1997). Furthermore, a broader institutional understanding may provide important insights. Because scientific activity encompasses more than the aggregation of individual researchers' work, a nonreductionist social systems view is necessary to capture institutional forces. In a nutshell, every science, including OS, is a social process, which may be affected by the same psychological, historical, and institutional dynamics (e.g., emotional motives, power plays, cognitive biases and heuristics) as other social processes (Barber 1971; Schofer 2004; Tsoukas and Knudsen 2003). This view of science as a social or institutional arena is consistent with one of the most influential accounts of scientific activity, namely Kuhn's (1970) *Structure of Scientific Revolutions*, as well as many other studies of scientists' work (e.g., Barnes et al. 1996; Bloor 1991; Brown and Malone 2004; Knorr-Cetina 1999; Latour and Woolgar 1986; Lynch 1993; MacKenzie 1981; Orlitzky 2011; Popper 1962, p. 217).⁴

The view of NHST as an institution within the broader social institution of OS may be instructive because this conceptualization can help solve the riddle of NHST's longevity despite the methodological controversy surrounding it. Institutions can be defined as "socially constructed, routine-reproduced [...] program or rule systems" (Jepperson 1991, p. 149). NHST is a rule system with which quantitative researchers have come to classify hypotheses as true or false. Of course, truth or falsity is no entity-inherent attribute (Rorty 1989, pp. 4–5). Rather, it is a convention built on the social construction of statistics and NHST (Hogben 1968; MacKenzie 1981; McCloskey 1998, pp. 120–124). In fact, NHST has become such a routine procedure that it has been likened to a ritual (e.g., by Gigerenzer 1998; Labovitz 1972).

⁴Like any other science, OS can be considered a social institution, which is a view that is neither novel nor new (Barnes et al. 1996; Kuhn 1970). Science not only institutionalizes objectivity (Fuchs 2001), but more generally fits the characteristics of institutions as discussed by Scott (2001, pp. 48–50). First, science is socially structured and resilient against outside forces (Brown and Malone 2004). Second, as an institution, science provides meaning to scientists' professional activities. Third, symbols and routines transmit the institution of science. Fourth, science is marked by long periods of stability, incremental change, and puzzle-solving (Kuhn's "normal science") interrupted by periods of discontinuous change, typically labeled "scientific revolutions" (Kuhn 1970, p. 92).

Institutions rest on three pillars or subsystems: *regulative*, *normative*, and *cultural-cognitive* (Scott 1995, 2001). Expressed slightly differently, institutions are composed of the following elements:

1. “meaning systems and related behavior patterns [i.e., the procedure of statistical significance testing itself and the beliefs expressed in the first paragraph of this article], which contain
2. symbolic elements, including representational, constitutive and normative components [*t* tests, *F* tests, χ^2 tests, etc.], that are
3. enforced by regulatory processes [review procedures at journals and book publishers]”

(Scott 1994, p. 56)

Analyzing developments within each of these (regulative, normative, and cultural-cognitive) subsystems, researchers can gain a critical understanding of NHST as an institution affected by broader forces in the sociology of OS—and its lack of deinstitutionalization. It is important to note that the following arguments, while introduced as propositions, are only conjectures. That is, instead of calling for future hypothesis testing, these theoretical propositions are rhetorical place markers and summary statements that will facilitate the reading of this particular institutional account of NHST. Therefore, the arguments and conjectures introduced in this article should only be considered conversation starters about the practice of NHST in research on organizations and MCS.

2.1 Regulative forces

On the surface level, the persistence of an institutionalized practice rests on regulatory forces, that is, coercive enforcement mechanisms. The proverbial “publish-or-perish” pressure guides career success in OS. In this process, journal editors and reviewers serve as regulatory gatekeepers (Beyer et al. 1995, pp. 1228–1230). In any social science, but especially in a science as pluralistic as OS (Burrell 1996; Donaldson 1995; Pfeffer 1993; Van Maanen 1995b), scientific value and “validity” generally are no naturally inherent qualities of the studies submitted to journals; instead, they are ultimately socially conveyed awards (Gans and Shepherd 1994; Peters and Ceci 1982). Intersubjective judgments of appropriate or “good” methods determine to a large extent publishing success and thus professional advancement (Kuhn 1970; Park and Gordon 1996). There are several procedures that make this inherently (inter)subjective process (Cohen 1989; Kincaid 1996) more objective, but NHST appears to be the most decisive and powerful, as argued below.

The examination of journal content and researcher experience confirms the socially constructed power of NHST. Generally, little journal space is devoted to statistically nonsignificant results, regardless of the magnitude of the effect size (Cortina and Folger 1998; Hill 2003). Many researchers presume that unpublished studies show smaller effect sizes and, thus, are less likely to be statistically significant than published studies (e.g., Begg 1994; Coursol and Wagner 1986; McNemar 1960). Mahoney (1976), who regarded NHST as highly problematic, affirmed the importance of the regulative force of editorial decisions by stating, in reference to NHST, that “until the rules of the science game are changed, one must

abide by at least some of the old rules or drop out of the game” (p. xiii). Thus, researchers who want to make a contribution to OS may partly be driven by force, fear and expedience in their use of NHST and evasion of professional opprobrium.

Undoubtedly, regulative insistence on NHST has some desirable institutional consequences for authors on the one hand and reviewers/editors on the other in an inherently conflict-laden process (Camic and Wilson 2003). Authors can (improperly) emphasize the conclusions from rejected null hypotheses as substantial and important evidence in support of their espoused theory. For example, McCloskey (1998, pp. 128–138) showed that approximately 70 percent of authors in one of the premier economics journals, *American Economic Review*, failed to distinguish between statistical significance and substantive significance. This suggests that, by substituting the former for the latter, most economists use this statistical *deus ex machina* to win the competition for journal space. If the uncertainty presented by sampling error were not easily circumvented by NHST, assessments of scientific contribution might have to invoke more elusive and subjective (but perhaps also more meaningful?) decision criteria (Gephart 1988). Hence, conformity to the NHST ritual can decrease the likelihood of researchers’ empirical efforts being dismissed.

Journal gatekeepers, on the other hand, can point to nonsignificant findings as evidence of the lack of scientific contribution (Gigerenzer 1993), without impugning the author’s theoretical or methodological expertise. In a quest to ensure that the journal’s publications are seen as scientific *and* important, editors and reviewers of the premier OS journals enforce and reinforce adherence to significance testing—and the term *significance* is not coincidental (McCloskey 1998, pp. 120–121). Conversely, the ability to reduce the quality of a study to the lack of significant results objectifies a type of conflict that is inherently (inter)subjective and embedded in scholarly discourse and evaluation (McCloskey 1998; Rorty 1991). Additionally, NHST also enhances review efficiency by introducing a front-end control of quantitative research submissions before these studies ever enter the publication process (although many of them, including statistically nonsignificant findings based on small sample sizes, might actually make quite valuable contributions, according to Schmidt 1996). In short, NHST makes success and failure more predictable in OS and, thus, may prevent open forms of conflict between researchers and gatekeepers (see, e.g., McSwite 1997, p. 255, for a contrary example when such control mechanisms are missing in academia).

In particular, high journal rejection rates increase the potential for conflict between authors and gatekeepers. For example, when journal space is increasing at a lower rate than the number of academics vying for it, NHST may become more salient because parties with conflicting interests (reviewers/editors versus authors) agree that it may be one of several useful techniques to objectify and diminish conflict. Interestingly, in many premier natural science journals where the acceptance rate of submitted papers is frequently over 75 percent (Adair 1982; Lazarus 1982), NHST is rare or completely absent (Schmidt 1996).⁵

⁵Articles in physics journals typically rely on point estimates of effect sizes with confidence intervals around them and emphasize the importance of replication because of the typically great empirical uncertainty of any primary study (Hedges 1987). Whether these style differences between journals in the social

Proposition 1 *Especially when journal space per potential contributor is decreasing, academic journal editors and reviewers, who act as regulators of “proper” science, rely on the outcomes of NHST as one of several criteria for the assessment of the value of a study.*

2.2 Normative forces

Scientific coercive regulation based on an instrumental logic would be a weak weapon against what many incumbents may consider the challengers of the status quo. What is needed in any ideological struggle is a normative framework that unifies and structures individual behaviors against the perceived “encroaching enemies of science.” Researchers, in other words, must feel compelled by a sense of social obligation underpinned by *norms*, which are the definitions of the legitimate means to pursue valued ends, and *values*, which are “conceptions of the preferred or the desirable, together with the construction of standards to which existing [...] behavior can be compared” (Scott 2001, pp. 54–55).

Regulative forces may be sufficient to stabilize social relations within scientific disciplines, but internalized scientific values and norms are necessary for demarcating the social system of science against intrusions from the social environment of science. Various norms have been proposed for the drawing of boundaries between science and non-science (Cohen 1989). The socially contingent character of these boundaries often is underemphasized when science is demarcated from non- or pseudo-sciences (Barnes et al. 1996; Collins and Pinch 1982). Instead, these criteria of demarcation are vested with an objective nature, an essentiality, which largely remains unsubstantiated in any objective sense (Fuchs 2001). How socially contingent these demarcation criteria really are has been shown by various sociological studies (e.g., Latour and Woolgar 1986; MacKenzie 1981; Shapin 1994; Shapin and Schaffer 1985). Social contingency implies that these demarcation criteria “could always be otherwise, and how precisely they have been applied could always have been otherwise” (Barnes et al. 1996, p. 154).

Although the epistemological science/non-science demarcation is tenuous and ultimately unsatisfying (Fuchs 2001; McSwite 1997), NHST can be considered the day-to-day signifier of the scientific nature of quantitative researchers’ work. Consistent with Popper (1969), theories become scientific due to their empirical testability. NHST, as the operationalization and implementation of Popper’s core scientific principle of falsifiability (Chow 1998, 2000), has become the central inferential step in ascertaining the truth or falsehood of theoretical propositions (Cohen 1989). In contrast, nonscientific statements fail to meet the falsifiability criterion. To avoid the supposed dead ends of, for example, astrology, Marxist theory, and Freud’s or Adler’s psychodynamic theories, which arguably suffer from a distinct lack of this criterion (Chalmers 1999), NHST is inculcated as a necessary norm sanctioning “good” and socially or even economically valuable science (McCloskey 1998).

sciences and journals in the natural sciences suggest any causal effect between acceptance rates and NHST is debatable, of course.

In line with this normative socialization of doctoral students, most methods textbooks do not even mention the NHST controversy (Gliner et al. 2002; McCloskey 1998). Instead, the myth is propagated that NHST is the ultimate arbiter in scientific inference and the only technique for testing hypotheses (see also Hedges 1987; Schmidt and Hunter 1997).

In researchers' quest for social status, insistence on hypothetico-deductive falsificationism and NHST becomes stronger as more external forces try to invade what scientists regard as their legitimate territory. Especially in an applied science like OS, these external pressures may be exerted by practitioners trying to redefine an area of inquiry. The Total Quality Management (TQM) movement is an example of the impact of science-external actors—consulting firms, management gurus, and the business mass media—as trendsetters in some OS area (e.g., organizational and small-group design) traditionally seen as the legitimate territory of organizational scholars (Abrahamson 1996).

Generally, practitioner encroachment into OS territory will set off two different institutional processes, particularly when resources (including managers' and policy makers' time and attention) are scarce and researchers regard these intrusions as illegitimate (for example, because they are “only” based on intuitions, induction, and individual cases). First, it increases the likelihood that academics will invoke objective selection criteria for these management fashions. Many quantitative researchers will want to differentiate themselves from the more subjective trendsetters. In order to accomplish this differentiation, they will emphasize the use of “objective” methodology to separate the wheat from the chaff (e.g., Hilmer and Donaldson 1996; Pfeffer and Sutton 2006). Second, although the trendsetters primarily employ rhetoric useful for achieving managers' acceptance of their fashions (typically well-narrated, idiographic, and current case examples), over time they, too, are bound by norms of rationality, which envelop the management fashion market (Abrahamson 1996). Hence, sooner or later, many trendsetters outside academia will also come to use “objective” methods such as NHST, though wrapping them in a more user-friendly package than academics. They will use this “objective” evidence to affirm the grounding of their intuitions (and interests) in data and material facts. The use of techniques such as NHST may legitimate quantitative research by emphasizing the fact- rather than value-based nature of their fashions because many research consumers—and especially those with limited statistical backgrounds—falsely conceive of NHST as a guarantee of objectivity (McCloskey 1998; Schmidt and Hunter 1997).

Proposition 2 *Perceived encroachments from nonscientific environmental forces increase adherence to NHST (a) first among organizational scientists as a normative sign of differentiation and implied superiority and (b) then and over time among practitioners as part of a normative rhetoric of “objectivity,” logocentrism, and rationality.*

The following three subsections on three challengers of positivism from within OS will illustrate how this normative call to (NHST) arms may come about in specific instances of a perceived threat to the “objectivity” engendered in positivist science. NHST may be one of several normative defense mechanisms against these epistemic challenges.

2.2.1 Qualitative research

Qualitative organizational research is typically seen to pursue goals different from quantitative research. Thus, its norms and values differ from those of quantitative research, too. Qualitative methodology is often portrayed as interactive and emergent (Ahrens and Chapman 2007; Van Maanen 1998a). Because the investigator takes a more central role in qualitative studies, the assessment of “good” qualitative research tends to be more subjective than that of “good” quantitative research. Also, qualitative researchers typically value interpretation, self-reflection, and often (but not always) induction as central research norms (Creswell 1998; Van Maanen 1998a). Hence, it is difficult to develop a uniform “cookbook” approach to qualitative methodology (Van Maanen 1988)—in contrast to, let’s say, Pedhazur and Schmelkin’s (1991) or Schwab’s (1999) methodical quantitative frameworks. However, overall instances of qualitative research tend to be more similar to each other than to quantitative research, particularly in terms of format, which often leads to different choices in publication outlets (Van Maanen 1988, 1998a).

Because of these differences in research aims, norms, values, and dissemination, quantitative researchers may regard qualitative research as either a threat or another important lens through which organizational phenomena can be analyzed. In turn, this view of qualitative research as either a scientific threat or valuable additional contribution may explain, to some extent, the continuing normative insistence on NHST. When quantitative researchers regard qualitative research as a threat, scientific activity needs to be sufficiently delimited so that it precludes the seemingly more subjective and supposedly nonscientific work of qualitative researchers. From this perspective, subjective insights are illegitimate and unscientific (Daston 1992). Instead, objectivity must be emphasized—in a context of material, quantitatively captured data that are typically presumed to reflect, or correspond to, reality. The goal of OS must be to distinguish true from false hypotheses. These dichotomous empirical tests must be replicable by anyone that chooses to apply them to the same data set. To this end, quantification can be used as a tool for overcoming distance and distrust (Daston 1992; Porter 1992). This propping up of the legitimacy of quantitative research by differentiating it from nonquantitative research can be detected in some prominent researchers’ failure to acknowledge nonquantitative research strategies in their OS methods books (e.g., Pedhazur and Schmelkin 1991; Schwab 1999).

NHST serves this purpose of differentiation from qualitative research and, thus, scientific legitimation quite well; for it erects an artificial barrier between “objective,” or “material,” data on the one hand and “subjective,” or “interpretive,” data on the other. Many quantitative researchers presume that whatever belongs to the nonquantifiable realm of researcher interpretation does not quite live up to the epistemological status of a *significant* finding and, therefore, is either ignored or relegated to the Discussion section in quantitative research reports. Few in the current generation of quantitative researchers seem to realize that the epistemological foundation of NHST does not ensure objectivity, either, as already pointed out by several authors (McCloskey 1998; Meehl 1997; Rozeboom 1997; Schmidt 1996; Schmidt and Hunter 1997). Yet, NHST is one of those stylistic practices “whose

words reinforce the conceit that phenomena are represented rather than created” (Weick 1996, p. 311). Hence, when qualitative perspectives are socially constructed as research strategies fundamentally different from quantitative research, increased adherence to NHST confers to quantitative researchers a gain in scientific stature as a signifier of their own “scientific” contributions

2.2.2 *Mixed research*

Because the divide between qualitative and quantitative research *is* a convention, mixed research, which gains valuable insights from both perspectives, may become a preferred and widely promoted way of conducting research (Creswell 2003; Lewis and Grimes 1999; Morris et al. 1999; Tashakkori and Teddlie 1998, 2002; Tsang and Kwan 1999). However, this mingling of research strategies will not predetermine the deinstitutionalization of NHST.

NHST may actually increase if positivist norms are imported into qualitative research activities. What Weick (1977, p. 139) called the “brute school” of quantitative research may superimpose its own values and assumptions on the “cute school” of qualitative research. Such totalizing tendencies are typical of functionalism, the paradigm of most quantitative research (Burrell and Morgan 1979). Thus, studies that appear to have a clearly formulated qualitative starting point may end up with NHST, for example in multimethod field studies that test hypotheses concerning MCS design in new product development (Davila 2000). Particularly in the context of content analysis (e.g., Crofts and Bisman 2010), invoking the desirability of computerized, depersonalized text analysis (Kabanoff 1996; Larsen and Monarchi 2004) may reinforce and institutionalize aperspectival objectivity, defined as the “elimination of individual (and occasionally group) idiosyncracies,” as well as mechanical objectivity, which is the “suppression of the universal human propensity to judge and aestheticize” (Daston 1992, p. 597).

At the same time, it should be acknowledged that the superimposition of positivist research values in mixed research is possible, but not automatic or even likely. Researchers who genuinely value the aforementioned qualitative and interpretive norms and values in their mixed research may be less fixated on NHST as a symbol, or signifier, of aperspectival and mechanical objectivity. What may emerge instead is an interpretive attitude to quantitative data not unlike the one that could be observed in the early days of OS (Van Maanen 1998b). Even today, such a descriptive, interpretive stance can still be found, for example, in the study of pre-decision control mechanisms in strategic investment decisions (Alkaraan and Northcott 2007). As a further example, this one from OS (Hoffman 1997), important insights can be gained through a researcher’s personal interpretation of a multitude of longitudinal (including quantitative) data sources, rather than aperspectival objectivity, or a “view from nowhere” (Daston 1992; Nagel 1986). In short, a genuine appreciation of interpretive norms and values by OS researchers will lead to a decreased emphasis on NHST because the more subjectivist assumptions of qualitative research are allowed to be integrated into the research—rather than defended against as supposedly unscientific norms and values.

2.2.3 Postmodernism

An approach that undermines the conventional positivist norms, values, and assumptions of social science to an even greater extent than most qualitative research is postmodernism. Postmodernism challenges individual knowledge and reason, the construction of the world as objectively given, and language (including scientific knowledge claims) as truth-bearing. Instead, postmodernism posits communal rhetoric and discourse, the world as socially constructed, and language as pragmatic practice (Gergen 2001; Rorty 1997). Quantitative OS researchers, who are generally lovers of Truth,⁶ are unlikely to engage in debates about ontology and epistemological defenses of their worldviews—with exceptions, of course (e.g., Donaldson 1985, 1996). More likely, quantitative researchers tend to defend against the postmodern problematizing of normal science, Truth, representation, writing, and generalizability (Kilduff and Mehra 1997) via methodological means, since methodology is generally quantitative researchers' expertise, not philosophy of science or epistemology. NHST reaffirms the existence of real entities “out there” as few other quantitative procedures can—by presupposing it; for if these entities that were just observed, compared, contrasted, tested against the null hypothesis, or tackled in some other way by the “brute school” (Weick 1977, p. 139) did not objectively exist, how could we reasonably find and state something meaningful about their truth or falsehood?

Another epistemology that is seen as a threat to the deeply embedded values and norms of quantitative methodology may be confronted through an insistence on objectivist values and norms (e.g., Donaldson 1996; Locke and Becker 1998a, 1998b). Consistent with predictions derived from institutional theory (Oliver 1991), defiance may be the most likely stance toward postmodernism and other perceived “encroaching enemies of science.” Oliver (1991) pointed out how a multiplicity of institutional constituents on which the focal community is not dependent often leads to normative resistance. Such resistance may become even more likely and turn into defiance when the pressure is inconsistent with the focal community's norms and values (Oliver 1991). Although postmodernism does not dogmatically insist on (the superiority of) particular methodologies (Feyerabend 1975) and may contribute various new, valuable insights to OS (Alvesson and Deetz 1996; Boje et al. 1996; Calas and Smircich 1999; Mirchandani 2005), its norms and values are often perceived⁷ by modernists and positivists as threats to the foundations of reason, science, and knowledge (e.g., Brown 2001; Kincaid 1996; Nola 2003; Thompson 1993). In contrast to positivist assumptions of “real entities out there,” postmodern norms and values generally destabilize the old order (Gephart et al. 1996; Rosenau 1992), fracture modernist meta-narratives (Reed 1990), and, thus, undermine belief in scientific progress (Burrell 1996). Because of this postmodern debunking of modernists' “core possessions” and desecration of their “sacred symbols”

⁶In this instance and on other occasions throughout the article, this terminology follows Rorty. Capital letter *Truth* is defined as the accurate representation of a natural order (Rorty 1997, p. 23).

⁷The institutional impact of postmodernism is by way of modernists' *perceptions*, which may often be quite simplistic. For example, a self-confessed modernist once equated postmodernism to a “preference for pastiche” (Thompson 1993, p. 183). In addition, modernists may have a postmodernist out-group homogeneity bias when there really are a wide variety of different strands of postmodern theorizing (e.g., Alvesson and Deetz 1996; Boje et al. 1996; Calas and Smircich 1999; Rosenau 1992).

(Fuchs 2001, p. 23), a positivist call to arms often occurs (e.g., Brown 2001; Donaldson 1985, 1995, 1996; Pfeffer 1993). NHST is the core armament because its focus on replicability and intersubjective evaluation of knowledge claims is presumed to approximate and represent objective reality (Cohen 1989; Kincaid 1996). Thus, at the systems⁸ level positivists socially construct their epistemological response as a circling of the wagons,⁹ for example, by insisting on the incommensurability of other norms and values (Jackson and Carter 1993) or the superiority of objectivism (Locke and Becker 1998a, 1998b; Pfeffer 1993). Only in rare instances do the counter-attacks become more direct when positivists evoke war imagery, pointing out how “critics of the postmoderns score direct hits” (Brown 2001, p. 92).

The advance of postmodernism may often lead to a thickening of normative defenses and greater emphasis on the demarcation of science from (postmodern) non-science, a project for which NHST is generally seen as the essential ingredient (Chow 1998, 2000; Cohen 1989). NHST, as a “universal reflex” (McCloskey 1998, p. 123) that purports to test knowledge objectively, may become the subtle, polite, and politically correct journal equivalent of book-form refutations of postmodern anti-essentialism and subjectivism in the “many different battles that could be called ‘science wars’” (Brown 2001, p. xi). Clearly, NHST opposes the type of self-reflection that postmodernist (as well as many qualitative) researchers regard as an important scholarly activity. This implicit opposition emerges from the frequentist view of probability p that underlies sampling theory in NHST (as opposed to the subjectivist view that is briefly outlined later). The frequentist view assumes that probability is a property of nature that is independent of researcher beliefs (Kline 2004). The institutional practice of NHST superimposes this objectivist norm on researchers, who “in their daily lives probably take a subjective view of probabilities” (Kline 2004, p. 29). In sum, postmodernism, when perceived as a threat to knowledge generation, draws forth and thickens objectivist defense mechanisms and, thus, reinforces normative adherence to NHST.

2.3 Cultural-cognitive forces

In line with this regulative and normative self-assurance, NHST can be construed as one of the central procedures creating a scientific culture of confidence and good faith (Meyer and Rowan 1977). In the absence of technical validity, an activity’s legitimacy remains dubious or at least contested. Legitimacy refers to social acceptability, credibility (Scott 2001), or a “generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions” (Suchman 1995, p. 574). In the early days of science, the legitimacy of incontrovertible facts or scientific proofs was established by the direct correlation between scientist and gentleman (Barnes et al. 1996;

⁸This defense against postmodernism is not (primarily) an individual, deliberate response because few individual scientists would have the power to influence an institutional field anyway. Even more important, the normative pillar holds that researchers may be guided more by social rules, roles, and other obligations prescribing appropriate scientific behavior than by individual interests (Scott 2001).

⁹I would like to thank John Van Maanen for suggesting this metaphor.

Brown and Malone 2004). However, the assumption that scientists generally act in good faith and, therefore, can be trusted is as important today as it was then. In modern, less stratified societies, rational procedures tend to take the role of social class as symbols of legitimacy (Meyer and Scott 1983). In today's quest for scientific legitimacy, what partly absorbs epistemological uncertainty is the secular simulacrum of NHST.

Hardly any technique could artificially prop up good faith as much as one that subtly and cleverly replaces the norm of intersubjective verifiability with an ontological pretense of objectivity. In other words, the rhetoric of NHST may be seen as a "heavy tool" (Weick 1996, p. 301) that reinforces the belief that knowledge is derived from the facts that either do or do not live up to the demarcation criterion of $\alpha = .05$ or $.01$ (McCloskey 1998). Although even post-positivists regard such an ontological pretense as "rather silly" (Chalmers 1999, p. 12) or superfluous (Donaldson 2003), procedures that ritualistically invoke Truth and dualistic (true/false) certainty enhance epistemological confidence. For example, were conventional alpha levels not accepted as sacrosanct mirrors of reality, they would become open to subjective interpretation and negotiation. Consequently, organizational scholars' commitments to theories and frameworks would have to be much more tenuous. Many researchers untrained in the philosophy and sociology of science probably consider a normal (quantitative) science to be unimaginable without the legitimating backbone of NHST that accepts objective reality as given, found, or discovered—not as socially constructed and, thus, contingent on researcher presuppositions (Bloor 1991; Burrell and Morgan 1979; Kuhn 1970).

Proposition 3 *NHST is a cultural-cognitive heavy tool that artificially infuses confidence in the interplay between theory and evidence and good faith within the community of organizational scholars.*

The final and arguably most deep-seated and interesting force explaining the institutional adherence to NHST is the cultural-cognitive orthodoxy of science. Even in the absence of methodological, regulative, and normative justifications, NHST would probably still stand firm in the social system of science (Abelson 1997). This deeper cultural-cognitive explanation of the prevalence of NHST lies in the taken-for-granted cognitions embedded in positivist, dualistic conceptions of Truth and science.

The dualism of true/false has traditionally been considered the currency of all sciences (Luhmann 1990). In other words, any scientific endeavor is supposed to be about the assessment of the dichotomous truth content of scientific knowledge claims (Cohen 1989; Popper 1972, 1983). Unlike Weick (1989), who emphasizes the diversity of theory selection criteria, Luhmann considers the true/false dualism to be the core of scientific communication and, hence, selection and validation of theory, which hinges on the attachment of the linguistic symbols "true" and "false" to knowledge (Luhmann 1990, p. 578). NHST is the most common and institutionalized operationalization of this selection mechanism (Chow 1998, 2000). Even though equating hypothesis testing with NHST is actually specious (Schmidt and Hunter 1997), NHST, as a social convention, has been considered a natural fit for Popper's (1969) falsifiability criterion. In OS, hypothesis testing—invariably via NHST—has

certainly held a privileged position among other possible theory selection criteria (e.g., Donaldson 1985; see also Miner 2003).

The differentiation of objectively true hypotheses from false hypotheses may have psychological roots. NHST quite effectively satisfies the human drive to categorize and use cognitive shortcuts (Dawes 1988; Nicholson 1997; Tversky and Kahneman 1974). Because of the human predilection for categorical thinking or, less charitably, “splitting” (Berman 1981; Laing 1959; Lynne 1999), quantitative researchers may also naturally gravitate toward simplistic interpretations of the evidence, such as NHST, as it meets our need for facticity, that is, our drive toward believing that things are as they appear and that we face an obdurate reality (Turner 2002).

However, the true/false orthodoxy embodied in NHST is not unproblematic. First, insistence on Popper’s restrictive deductivism, which states that theories fall after the refutation of single hypotheses, is unreasonable (Goldstone 2004, pp. 47–48; Gorski 2004). Second, today’s more sophisticated *holistic* deductivism (Burawoy 1990; Lakatos 1970) can protect any theory against falsification—through what is generally called a protective belt of auxiliary hypotheses. Assessments of “progressive” and “degenerative” research programs, which are used to distinguish proper and improper protective belts, are subjective (Gorski 2004, p. 12). Consequently, if hypothesis testing depends on subjective decision frames, assessments of theories as *true* or *false* are ultimately subjective. For these two and a number of other reasons, empirical validation may in fact be an unrealistic aim (see also Lindblom 1987; Weick 1989). Even worse, though, many so-called organizational “facts” can take on a life of their own and become self-fulfilling (Ferraro et al. 2005; Ghoshal 2005).

Instead of showing signs of self-reflection and contextualization, much current OS theorizing reinforces the true/false dualism through an emphasis on the discovery of general, abstract laws in OS. When “transcendent forms of discourse” (Donaldson 1985, p. 76) are commended as the essence of scientific communication (see also Luhmann 1990) and general, causal laws are envisioned as the endgame of OS, rich context-bound descriptions of organizational contexts are often implicitly devalued as un- or pre-scientific. The implication is that narrative or subjectivist approaches are inferior to “objective” quantitative approaches that make the discovery of true general laws possible (Kincaid 1996), with measurement often taking center stage in what is considered legitimate OS (Donaldson 1985, 1996; Nunnally and Bernstein 1994; Pedhazur and Schmelkin 1991; Schwab 1999). Human meanings ascribed to organizational processes are factored out from such theory testing (Gephart 1986), so that the rich social complexity of organization theory often remains concealed in quantitative research. With general causal laws being seen as the Holy Grail of OS, inherently complex organizational processes are ultimately categorized as consistent or inconsistent with hypotheses derived from more or less abstract theory. Deductive, quantitative sensemaking seems to demand just this type of de-contextualization (Hempel 1965; Stinchcombe 1968). For example, Boxall and Purcell (2003) argued how previous hypothesis testing of the general economic benefits of so-called “high-performance work systems” often was too simplistic and not sufficiently attentive to national or industry contingencies. Even more so at the micro level, McGuire (1983) faults the over-generalizing intent of law-like, positivist deductivism for the inadequate epistemology in psychological research.

Proposition 4 *Emphasis on theoretical generalizability and lawfulness reinforces adherence to NHST as a result of cognitive shortcuts in the empirical validation of theory.*

It must be emphasized that lawfulness does not necessarily or automatically reinforce NHST under all circumstances. Even though general laws prevail in the physical sciences, use of NHST is much less common than in the social and organizational sciences (Hedges 1987; Schmidt 1996). Theoretical underspecification explains why dichotomous splitting occurs at the null point and, thus, why thinking in general laws may *at present* be detrimental to OS from a methodological perspective. Were OS theories more advanced, strong inference would be possible (Platt 1964). That is, more specific and concrete (i.e., stronger) theory would allow quantitative empirical work in OS to compare and contrast at least two competing hypotheses, which contain point estimates. “Not nothing” versus “nothing” comparisons are typically inadequate for hypotheses based on strong theory (Meehl 1990). At present, relatively strong theory that allows for strong inference and/or graphical data fitting is still rare in quantitative OS. Without better theory allowing for strong inference, quantitative researchers will most likely (have to) continue relying on this “kind of essential mindlessness in the conduct of research” (Bakan 1966, p. 436) and “corrupt form of the scientific method” (Carver 1978, p. 378).

More advanced theory is not the only aspect on which the (de)institutionalization of NHST depends; each of the three forces shows interdependencies and interconnections with the others, as depicted in Fig. 1 by the arched arrows. This implies that without a concerted effort driving its deinstitutionalization, NHST will persist in theory testing in the study of organizations or, more specifically, MCS. A wide variety of institutionwide reforms will be required to consign NHST to the ash heap of history. Solutions may range from surface-level changes (regulative reforms) to moderately deep reforms affecting the norms and values of OS to quite radical and/or slow changes (cultural-cognitive reforms). Due to space constraints, this article cannot develop these institutional implications for methodological reform. Instead, the epistemological reforms required for the deinstitutionalization of NHST either have already been proposed by others (e.g., Kline 2004; Rodgers 2010; Schmidt 1996; Schwab and Starbuck 2009) or will be developed in greater depth in a future paper.

3 Consequences of statistical significance testing for organization studies

The aforementioned institutional explanations may help us understand the longevity of NHST, but they do not justify it. After all, the “rhetorical damage” (McCloskey 1998, p. 121) caused by NHST may be severe; for it may “ruin empirical work” (McCloskey 1998, p. 112) in studies of organizations, management control systems, and corporate governance as much as it did in economics (Ziliak and McCloskey 2008). It is to this harmful impact of NHST (in terms of theory development and testing) that I now turn.

3.1 Preserving and reinforcing the practitioner-academic divide

NHST may signal boundary development and maintenance in science—within and across disciplines (Barnes et al. 1996). Because OS, like all other social sciences, does not have a natural monopoly on knowledge production, boundary work is important—perhaps even necessary—for the scientific community to hold onto its power base in society (Brown and Malone 2004). This boundary work can help the community of organizational scientists fend off unwelcome intrusions (see prior section on normative defensiveness) and minimize inspection and evaluation from outsiders (Meyer and Rowan 1977).

Arguably, prior conceptualizations of the academic-practitioner divide did not pay enough attention to the hypothesis-testing rituals and ceremonies that scientists use to preserve their autonomy. In contrast to the presumed objectivity and rationality of organizational scientists, practitioners are often portrayed as victims of irrational, emotional, or political influences (Abrahamson 1996). The out-group of practitioners can easily be swayed by rhetoric, it is assumed, whereas more objective criteria prevail in science (Brown and Malone 2004; Hilmer and Donaldson 1996). As the politics of persuasion is hidden behind a veil of objectivity, the rhetoric of science typically remains unacknowledged (Nelson et al. 1987; Van Maanen 1995a, 1995b). NHST is one such objectifying tool providing this veil, especially when it is combined with statistically complex multivariate analyses. The conventional assumption among many quantitative researchers is that NHST—and NHST alone—permits the scientific assessment of theories (Cohen 1989; Donaldson 1996). In this sense, NHST can be seen as yet another mechanism dividing scientists and practitioners.¹⁰

Yet, an organizational science divorced from organizational reality risks practical irrelevance. For example, although there are numerous findings (coming out of quantitative research) linking the personality trait of conscientiousness to job performance, the practically more important questions of managerial “good judgment” (Pearce 2004, p. 178) or “wisdom” (Weick 2001, pp. 362–363) in employee selection and advancement has proved to be more difficult to operationalize and subject to NHST. Though not typically identified as the culprit, NHST may contribute to this potential problem of irrelevance in two ways. First, NHST is typically concerned with differences at the group level (Kline 2004, p. 78, but also note Chap. 4), which may obscure, conceal, or gloss over the idiographic details typically needed in practitioner decision making (Pearce 2004). Second, because NHST is based on hypothetico-deductivism (Cohen 1989; Rozeboom 1997) rather than inference techniques grounded in the data (e.g., induction, abduction; see Rozeboom 1997), the problematic epistemological stance implied by NHST (Meehl 1997) does not really allow for a firm grounding in what Pearce (2004, p. 175) called *folk wisdom*. This in turn may contribute to the researcher assumption of the superiority

¹⁰Albeit, practitioner-oriented studies (e.g., by consulting firms) often prop up their legitimacy by relying on NHST as well. Although this practice arguably complicates the issues relating to the academic-practitioner divide and boundary control, such practitioner adherence to NHST actually strengthens the prior institutional logic of NHST (see Proposition 2).

of academic knowledge over practitioner knowledge, as observed to some extent by Rynes et al. (2001, in particular Table 1). It may lead to researchers' perception that a lot of folk wisdom represents dogma (see, e.g., Hilmer and Donaldson 1996; Pfeffer and Sutton 2006), whereas quantitative organizational science can somehow evade the cultural-cognitive traps of ideology and social ritual. Unfortunately, other research has documented the considerable impact of ideology and social ritual on OS (see, e.g., Orlitzky 2011).

The academic-practitioner divide that is frequently built on such epistemological defensiveness is particularly undesirable in an applied field such as OS (Pearce 2004; Rynes et al. 2001), in which NHST may both symbolize and contribute to this boundary control between science and non-science. But aren't there OS-internal benefits to NHST? Doesn't NHST sharpen and solidify scientific knowledge? Actually, closer examination of the track record of NHST also uncovers science-internal deficiencies and unintended negative consequences with respect to theory development and testing (see also Kline 2004; McCloskey 1998; Schmidt 1996). The following section will build on prior NHST critiques in the social sciences generally relevant to theory development and testing, illustrating these consequences for OS with a few concrete examples.

3.2 Impeding knowledge growth

Analogous to the dichotomization of continuous variables (Farewell et al. 2004; Hunter and Schmidt 2004; Labovitz 1972), NHST loses important information by dividing the empirical evidence in primary studies into real effects (when the null hypothesis is rejected) and null findings (when the null hypothesis cannot be rejected). An important factor determining this splitting of the evidence into so-called real and null findings is sample size—and, admittedly, NHST does address the study artifact of sampling error (Kline 2004, pp. 79–80). Yet, by dichotomizing the conclusions, NHST does not accurately reflect the empirical uncertainty caused by sampling error (or any other study artifact, such as measurement error) and, therefore, is likely to lead to the postulation of apparently “conflicting results in the literature” (Hunter and Schmidt 2004, p. 8). In contrast, proper acknowledgment of this uncertainty would lead to the correct conclusion that seemingly contradictory significant and nonsignificant findings may actually be reflective of the same underlying population parameter, a variability that is not adequately captured by NHST (Hunter and Schmidt 2004).

NHST may cause particular harm when OS researchers review a stream of research. They are often tempted to use the binary evidence from previous significance tests (i.e., statistically significant versus nonsignificant results) without considering the effects of any distorting study artifacts. Statisticians warned against this flawed vote counting of statistical significance tests several decades ago (Hedges and Olkin 1980; Light and Smith 1971). Although introductory methods textbooks in OS reiterate this caveat (Schwab 1999, p. 301), the vote counting technique (also known as the *box-score method*) still is widespread, even in research reviews compiled by quantitative researchers. Typically, literature reviews relying on the box-score method characterize research streams as suggestive of “variable findings,” “highly mixed results,” or the “indeterminate nature” (for examples, see Hunter and Schmidt 2004;

Schmidt 1992). These conclusions, or impressions, then need to be corrected by quantitative literature reviews, which do not rely on vote counting. For example, in research on corporate social and financial performance, two meta-analyses (Orlitzky and Benjamin 2001; Orlitzky et al. 2003) highlight that findings in this research stream are more consistent than is traditionally acknowledged by either narrative or vote-counting literature reviews.

In sum, dichotomous reporting in primary studies (“was our hypothesis supported or not?”) comes at the price of impeding knowledge growth from the perspective of research programs, which are often seen as the lifeblood of science (Lakatos 1970; Rozeboom 1997; Schmidt and Hunter 1997; Steiger 1990). NHST eschews cognitively complex sensemaking of effect sizes and sampling variations around them in favor of primary researchers’ psychological certainty. Paradoxically, this epistemic certainty at the level of primary studies may translate into uncertainty at the level of entire research programs (Hunter and Schmidt 2004). Any failure to recognize that the variability in NHST results may be merely due to differences in sample size (Hunter and Schmidt 2004) impedes the growth of OS knowledge because it may pull researchers in two different directions. First, an extended search for moderators, or contingency effects, may lead to costly and unnecessary investment in future empirical research (as cited in Kline 2004; Schmidt 1996), even though the “contradictory findings” often reflect only study artifacts and actually are to be expected—for example, because of low statistical power due to small sample size N (Cohen 1962; Schmidt et al. 1976; Sedlmeier and Gigerenzer 1989). Or, perhaps even more ominously, the analytic oversimplification inherent in NHST may result in the premature abandonment of seemingly contradictory research—and of theories that may be interesting, beautiful, or real (see Weick 1989, for an elaboration of these theory selection criteria on pp. 525–528).

3.3 Discouraging study replication

In a related vein, NHST is one of several practices that discourage research replications in OS and thus, indirectly, impede knowledge growth by extension (Kaplan 1964) because OS researchers do not explore the boundary conditions of their theories empirically (Tsang and Kwan 1999). In other words, without empirical replication, organizational researchers cannot know the range of phenomena to which their theories may apply. Yet, in OS and economics, replication has been shown to be extremely rare, far below 1% of all articles published (Kmetz 1998; see also Shaver and Norton 1980). NHST fails to encourage replication because, as mentioned before, it conveys a false sense of epistemological certainty (Schmidt and Hunter 1997). NHST perpetuates the myth that there are such things as crucial experiments (Cohen 1965) or the one “perfect study” (Hunter and Schmidt 2004, p. 17) that can corroborate a given theory. If organizational researchers were interested in approximating a specific parameter (as an effect size estimate) rather than rendering dichotomous true/false verdicts about population parameters, replication would be regarded as more valuable (as it is in the natural sciences).

This negative impact can be illustrated by highlighting the benefits of replication studies for theory development and testing (see also Tsang and Kwan 1999)—using an example from population ecology. For instance, without replication, OS researchers would only have recognized the liability of newness (Stinchcombe 1965), but not other liabilities associated with organizational adolescence and senescence (Carroll and Hannan 2000). More broadly, the entire research program in organizational ecology can be portrayed as a systematic inquiry into and modeling of population-level generalizations and extensions, made possible by replication studies in diverse industries (Baum 1996). More generally, Tsang and Kwan (1999) argued how replication might lead to more detailed and accurate theory. Certainly, OS would be poorer if such replications did not occur because such omissions might lead to an uncritical acceptance of sample-specific “truths” as the Truth.

3.4 Mechanizing researcher decision making

Standardized decision making is a core feature of quantitative research because procedures perceived as objective generally reduce distrust in research communities (Daston 1992; Porter 1992). However, NHST arguably carries this standardization too far because it uncritically accepts alphas of .01 or .05 as magic Type I error levels (Nelson et al. 1986; Rosnow and Rosenthal 1989) and neglects possible errors due to the lack of statistical power (Hunter 1997). In other words, researchers may accept null findings as the Truth of “no relationship,” even though those null findings may merely be due to Type II error, which, as mentioned before, NHST does not control. Often, small-*N* research (e.g., Orlitzky and Benjamin 2003) does not sufficiently acknowledge low statistical power, that is, the possibility of false null findings (i.e., Type II errors). Furthermore, even in large-*N* studies, NHST fails to control for errors of the third type (i.e., specification errors), which may make findings unreliable or spurious (Gujarati 1988; Leamer 1983).

Furthermore, by mechanizing decision making, NHST may encourage research fads that have little theoretical merit (Meehl 1990). NHST can lead to an unreflective acceptance of problem framing and distract researchers from a more nuanced and context-specific interpretation of their observed effect sizes, which was persuasively argued by Dunnette (1966). Thus, the meaning of quantitative theory testing often remains meager and trivial because of the NHST ritual (Nunnally 1960). NHST frequently leads to false conclusions when, in fact, the theoretical foundations should have been examined more carefully, before these theories are subjected to their ritualistic testing. For example, Priem and Butler (2001), Powell (2001), and Sastry (1997) pointed out the problematic, premature testing of logically and/or philosophically feeble ideas in three different OS literatures: the resource-based view of the firm, sustainable competitive advantage, and punctuated equilibrium, respectively. As many researchers find themselves in a race of fads and fashions (Abrahamson 1996; Dunnette 1966), OS researchers may disengage from the more subjective and open-ended task of uncovering weak conceptual foundations, which the quick-and-dirty tool of NHST often helps to hide behind a veil of so-called objective conclusions (Kline 2004).

4 Conclusion

Many OS researchers still widely regard NHST as a best practice rather than an institutional one (see, e.g., Gigerenzer 2004; Oakes 1986; Schmidt 1996; Seth et al. 2009; Ziliak and McCloskey 2008). To correct such a misperception, this article highlighted the gravity and extent of the problem of statistical significance testing and explained how various regulative, normative, and cultural-cognitive forces may have contributed to the institutionalization of NHST. Then, as shown in Fig. 1, the article illustrated four specific unintended negative consequences of NHST's structural embeddedness on theory testing and development in OS. My hope is that this article may stimulate the kind of meta-theoretical debate so important for a vibrant social science (Tsoukas and Knudsen 2003; Van Maanen 1995a). In contrast to previous reviews of research methodology in OS ignoring the NHST controversy (e.g., Scandura and Williams 2000), the arguments presented in this paper highlighted why OS can no longer afford to gloss over the dysfunctions associated with the institution of NHST.

So, by way of conclusion and analogous to Brooks' (2004) arguments, I wonder: Do we, as organizational researchers, want to exhibit a "tendency to impose a false order on reality"? Will we want to maintain an "aura of objectivity" at all cost, reinforced by a technique that has a "record studded with error," as pointed out by several NHST critics (e.g., Cohen 1994; Kline 2004; Schmidt 1996; Ziliak and McCloskey 2008)? Can quantitative OS really improve its methodology, if epistemology and the institutional practice of social science are, in fact, at the root of the problem? When seemingly "nonideological modes of observing" (Fuchs 2001, p. 34) have been embraced as scientific doctrine, the reluctance to drop such "heavy tools" (Weick 1996, p. 301) as NHST may turn the drama of OS into a tragedy. Now is the time to develop a healthy skepticism toward the error-prone ways of "false scientism" (Brooks 2004) and a culture of objectivity and quantifiable certainty (Feldman 2004). Now is the time to pause and reflect on the ritual and rationality of statistical significance testing.

References

- Abelson, R. P. (1997). A retrospective on the significance test ban of 1999 (If there were no significance tests, they would be invented). In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* Mahwah: Erlbaum.
- Abrahamson, E. (1996). Management fashion. *The Academy of Management Review*, 21(1), 254–285.
- Adair, R. K. (1982). A physics editor comments on Peters and Ceci's peer-review study. *Behavioral and Brain Sciences*, 5(2), 196.
- Ahrens, T., & Chapman, C. S. (2007). Doing qualitative field research in management accounting: Positioning data to contribute to theory. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 299–318). Amsterdam: Elsevier.
- Alkaraan, F., & Northcott, D. (2007). Strategic investment decision making: The influence of pre-decision control mechanisms. *Qualitative Research in Accounting & Management*, 4(2), 133–150.
- Alvesson, M., & Deetz, S. (1996). Critical theory and postmodernism approaches to organization studies. In S. R. Clegg, C. Hardy, & W. R. Nord (Eds.), *Handbook of organization studies* (pp. 191–217). Thousand Oaks: Sage.
- Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *The Journal of Wildlife Management*, 64, 912–923.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 423–437.

- Barber, B. (1971). Resistance by scientists to scientific discovery. *Science*, 134(3479), 596–602.
- Barnes, B., Bloor, D., & Henry, J. (1996). *Scientific knowledge: A sociological analysis*. Chicago: University of Chicago Press.
- Baum, J. A. C. (1996). Organizational ecology. In S. R. Clegg, In C. Hardy, & W. R. Nord (Eds.), *Handbook of organization studies* (pp. 77–114). London: Sage.
- Begg, C. B. (1994). Publication bias. In H. M. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York: Sage.
- Berman, M. (1981). *The reenchantment of the world*. London: Cornell University Press.
- Beyer, J. M., Chanove, R. G., & Fox, W. B. (1995). The review process and the fates of manuscripts submitted to AMJ. *Academy of Management Journal*, 38(5), 1219–1260.
- Birnberg, J. G., Luft, J., & Shields, M. D. (2007). Psychology theory in management accounting research. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 113–136). Amsterdam: Elsevier
- Bloor, D. (1991). *Knowledge and social imagery* (2nd edn.). Chicago: University of Chicago Press.
- Boje, D. M., Gephart, R. P. Jr., & Thatchenkery, T. J. (1996). *Postmodern management and organization theory*. Thousand Oaks: Sage.
- Boxall, P., & Purcell, J. (2003). *Strategy and human resource management*. New York: Palgrave Macmillan.
- Bromwich, M. (2007). Economics in management accounting. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 137–162). Amsterdam: Elsevier.
- Brooks, D. (2004). The CIA: Method and madness. *New York Times online*, February 2.
- Brown, J. R. (2001). *Who rules in science? An opinionated guide to the wars*. Cambridge: Harvard University Press.
- Brown, R. H., & Malone, E. L. (2004). Reason, politics, and the politics of truth: How science is both autonomous and dependent. *Sociological Theory*, 22(1), 106–122.
- Burawoy, M. (1990). Marxism as science. *American Sociological Review*, 55, 775–793.
- Burrell, G. (1996). Normal science, paradigms, metaphors, discourses and genealogies of analysis. In S. R. Clegg, C. Hardy & W. R. Nord (Eds.), *Handbook of organization studies* (pp. 642–658). London: Sage.
- Burrell, G., & Morgan, G. (1979). *Sociological paradigms and organisational analysis*. London: Heinemann.
- Calas, M. B., & Smircich, L. (1999). Past postmodernism? Reflections and tentative directions. *The Academy of Management Review*, 24(4), 649–671.
- Camic, C., & Wilson, F. D. (2003). Authors and reviewers: In appreciation. *American Sociological Review*, 68(6), iii–v.
- Campbell, J. P. (1982). Editorial: Some remarks from the outgoing editor. *Journal of Applied Psychology*, 67, 691–700.
- Carroll, G. R., & Hannan, M. T. (2000). *The demography of corporations and industries*. Princeton: Princeton University Press.
- Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48, 378–399.
- Chalmers, A. F. (1999). *What is this thing called science?* (3rd edn.). Indianapolis: Hackett.
- Chenhall, R. H. (2003). Management control systems design within its organizational context: Findings from contingency-based research and directions for the future. *Accounting, Organizations and Society*, 28, 127–168.
- Chenhall, R. H. (2007). Theorising contingencies in management control systems research. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 163–206). Amsterdam: Elsevier.
- Chow, S. L. (1998). Précis of *statistical significance: rationale, validity, and utility*. *Behavioral and Brain Sciences*, 21(2), 169–239.
- Chow, S. L. (2000). The Popperian framework, statistical significance, and rejection of chance. *Behavioral and Brain Sciences*, 23(2), 294–298.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145–153.
- Cohen, J. (1965). Some statistical issues in psychological research. In D. D. Wolman (Ed.), *Handbook of clinical psychology*. New York: McGraw-Hill.
- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd edn.). Hillsdale: Erlbaum.

- Cohen, B. P. (1989). *Developing sociological knowledge: Theory and method* (2nd edn.). Chicago: Nelson-Hall.
- Cohen, J. (1990). Things I have learned (so far). *The American Psychologist*, *45*, 1304–1312.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.
- Cohen, J. (1994). The Earth is round ($p < 0.05$). *The American Psychologist*, *49*, 997–1003.
- Collins, H. M., & Pinch, T. J. (1982). *Frames of meaning: The social construction of extraordinary science*. London: Routledge & Kegan Paul.
- Cortina, J. M., & Folger, R. G. (1998). When is it acceptable to accept a null hypothesis: No way, Jose? *Organizational Research Methods*, *1*(3), 334–350.
- Coursol, A., & Wagner, E. E. (1986). Effect of positive findings on submission and acceptance rates: A note on meta-analysis bias. *Professional Psychology*, *17*, 136–137.
- Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks: Sage.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks: Sage.
- Crofts, K., & Bisman, J. (2010). Interrogating accountability: An illustration of the use of Leximancer software for qualitative data analysis. *Qualitative Research in Accounting & Management*, *7*(2), 180–207.
- Daston, L. (1992). Objectivity and the escape from perspective. *Social Studies of Science*, *22*, 597–618.
- Davila, T. (2000). An empirical study on the drivers of management control systems' design in new product development. *Accounting, Organizations and Society*, *25*(4/5), 383–409.
- Dawes, R. M. (1988). *Rational choice in an uncertain world*. Fort Worth: Harcourt Brace.
- DiMaggio, P. J. (1995). Comments on "What theory is not". *Administrative Science Quarterly*, *40*(3), 391–397.
- Donaldson, L. (1985). *In defence of organization theory*. Cambridge: Cambridge University Press.
- Donaldson, L. (1995). *American anti-management theories of organization: A critique of paradigm proliferation*. Cambridge: Cambridge University Press.
- Donaldson, L. (1996). *For positivist organization theory: Proving the hard core*. Thousand Oaks: Sage.
- Donaldson, L. (2003). Organization theory as a positive science. In H. Tsoukas & C. Knudsen (Eds.), *The Oxford handbook of organization theory: meta-theoretical perspectives* (pp. 39–62). New York: Oxford University Press.
- Dunnette, M. D. (1966). Fads, fashions, and folderol in psychology. *The American Psychologist*, *21*, 343–352.
- Farewell, V. T., Tom, B. D. M., & Royston, P. (2004). The impact of dichotomization on the efficiency of testing for an interaction effect in exponential family models. *Journal of the American Statistical Association*, *99*(467), 822–831.
- Feldman, S. P. (2004). The culture of objectivity: Quantification, uncertainty, and the evaluation of risk at NASA. *Human Relations*, *57*(6), 691–718.
- Ferraro, F., Pfeffer, J., & Sutton, R. I. (2005). Economics language and assumptions: How theories can become self-fulfilling. *The Academy of Management Review*, *30*(1), 8–24.
- Feyerabend, P. K. (1975). *Against method: Outline of an anarchistic theory of knowledge*. London: New Left Books.
- Fidler, F., Cumming, G., Burgman, M., & Thomason, N. (2004). Statistical reform in medicine, psychology and ecology. *The Journal of Socio-Economics*, *33*, 615–630.
- Finch, S., Cumming, G., & Thomason, N. (2001). Reporting of statistical significance in the Journal of Applied Psychology: Little evidence of reform. *Educational and Psychological Measurement*, *61*(2), 181–210.
- Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver & Boyd.
- Fuchs, S. (2001). What makes "sciences" scientific. In J. H. Turner (Ed.), *Handbook of sociological theory* (pp. 21–35). New York: Kluwer Academic/Plenum.
- Gans, J. S., & Shepherd, G. B. (1994). How are the mighty fallen: Rejected classic articles by leading economists. *Journal of Economic Perspectives*, *8*, 165–179.
- Gephart, R. P. Jr. (1986). Deconstructing the defense for quantification in social science: A content analysis of journal articles on the parametric strategy. *Qualitative Sociology*, *9*, 126–144.
- Gephart, R. P., Jr. (1988). *Ethnostatistics: Qualitative foundations for quantitative research*. Newbury Park: Sage.
- Gephart, R. P., Boje, D. M., & Thatchenkery, T. J. (1996). Postmodern management and the coming crises of organizational analysis. In D. M. Boje, R. P. Gephart Jr., & T. J. Thatchenkery (Eds.), *Postmodern management and organization theory* (pp. 1–18). Thousand Oaks: Sage.

- Gergen, K. J. (2001). Psychological science in a postmodern context. *The American Psychologist*, 56(10), 803–813.
- Ghoshal, S. (2005). Bad management theories are destroying good management practices. *Academy of Management Learning & Education*, 4(1), 75–91.
- Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences, Vol 1: Methodological issues* (pp. 311–339). Hillsdale: Erlbaum.
- Gigerenzer, G. (1998). We need statistical thinking, not statistical rituals. *Behavioral and Brain Sciences*, 21, 199–200.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33, 587–606.
- Gliner, J. A., Leech, N. L., & Morgan, G. A. (2002). Problems with null hypothesis significance testing (NHST): What do the textbooks say? *The Journal of Experimental Education*, 71(1), 83–92.
- Goldstone, J. A. (2004). Response: Reasoning about history, sociologically. *Sociological Methodology*, 34, 35–61.
- Gorski, P. S. (2004). The poverty of deductivism: A constructive realist model of sociological explanation. *Sociological Methodology*, 34, 1–33.
- Gujarati, D. N. (1988). *Basic econometrics* (2nd edn.). New York: McGraw-Hill.
- Guttman, L. (1985). The illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis*, 1, 3–10.
- Hacking, I. (1965). *The logic of statistical inference*. Cambridge: Cambridge University Press.
- Hagen, R. L. (1997). In praise of the null hypothesis statistical test. *The American Psychologist*, 52, 15–24.
- Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah: Erlbaum.
- Hedges, L. V. (1987). How hard is hard science, how soft is soft science? The empirical cumulativeness of research. *The American Psychologist*, 42(2), 443–455.
- Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin*, 88, 359–369.
- Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
- Hesford, J. W., Lee, S.-H., Van der Stede, W. A., & Young, S. M. (2007). Management accounting: a bibliographic study. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 3–26). Amsterdam: Elsevier.
- Hill, T. L. (2003). Null hypothesis significance testing, the file drawer problem, and intercessory prayer: a Bayesian analysis of the effect (or effectiveness) of each. *Dissertation Abstracts International B*, 63(7-B). Accessed April 19.
- Hilmer, F. G., & Donaldson, L. (1996). *Management redeemed: Debunking the fads that undermine corporate performance*. East Roseville: Free Press Australia.
- Hoffman, A. J. (1997). *From heresy to dogma: An institutional history of corporate environmentalism*. San Francisco: Lexington.
- Hofmann, S. G. (2002). Fisher's fallacy and NHST's flawed logic. *The American Psychologist*, 57(1), 69–70.
- Hogben, L. F. R. S. (1968). *Statistical methods*. New York: Norton.
- Hubbard, R., Parsa, A. R., & Luthy, M. R. (1997). The spread of statistical significance testing: The case of the *Journal of Applied Psychology*. *Theory & Psychology*, 7, 545–554.
- Hunter, J. E. (1997). Needed: A ban on the significance test. *Psychological Science*, 8(1), 3–7.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd edn.). Thousand Oaks: Sage.
- Jackson, N., & Carter, P. (1993). 'Paradigm wars': A response to Hugh Willmott. *Organization Studies*, 14(5), 727–730.
- Jepperson, R. L. (1991). Institutions, institutional effects, and institutionalism. In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis* (pp. 143–163). Chicago: University of Chicago Press.
- Kabanoff, B. (1996). Computers can read as well as count: How computer-aided text analysis can benefit organisational research. In *Journal of Organizational Behavior (1986–1998)* (pp. 1–20). London: Chichester.
- Kaplan, A. (1964). *The conduct of inquiry*. San Francisco: Chandler.
- Kaufman, A. S. (1998). Introduction to the special issue on statistical significance testing. *Research in the Schools*, 5, 1.

- Kilduff, M., & Mehra, A. (1997). Postmodernism and organizational research. *The Academy of Management Review*, 22(2), 453–481.
- Kincaid, H. (1996). *Philosophical foundations of the social sciences*. New York: Cambridge University Press.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746–759.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington: APA.
- Kmetz, J. L. (1998). *The information processing theory of organization*. Aldershot: Ashgate.
- Knorr-Cetina, K. D. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge: Harvard University Press.
- Krueger, J. (2001). Null hypothesis significance testing: On the survival of a flawed method. *The American Psychologist*, 56(1), 16–26.
- Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd edn.). Chicago: University of Chicago Press.
- Labovitz, S. (1972). Statistical usage in sociology: Sacred cows and rituals. *Sociological Methods & Research*, 1(1), 13–37.
- Laing, R. D. (1959). *The divided self*. London: Tavistock.
- Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos, & A. Musgrave (Eds.), *Criticism and the growth of knowledge*. Cambridge: Cambridge University Press.
- Lambert, R. A. (2007). Agency theory and management accounting. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (Vol. 1, pp. 247–268). Amsterdam: Elsevier.
- Larsen, K. R., & Monarchi, D. E. (2004). A mathematical approach to categorization and labeling of qualitative data: The latent categorization method. *Sociological Methodology*, 34, 349–392.
- Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts* (2nd edn.). Princeton: Princeton University Press.
- Lazarus, D. (1982). Interrater agreement and acceptance rates in physics. *Behavioral and Brain Sciences*, 5(2), 219.
- Leamer, E. (1983). Let's take the con out of econometrics. *American Economic Review*, 73, 31–43.
- Lecoutre, M.-P., Poitevineau, J., & Lecoutre, B. (2003). Even statisticians are not immune to misinterpretations of Null Hypothesis Tests. *International Journal of Psychology*, 38(1), 37–45.
- Lewis, M. C., & Grimes, A. J. (1999). Metatriangulation: Building theory from multiple paradigms. *The Academy of Management Review*, 24(4), 672–690.
- Light, R. J., & Smith, P. V. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. *Harvard Educational Review*, 41, 429–471.
- Lindblom, C. E. (1987). Alternatives to validity: Some thoughts suggested by Campbell's guidelines. *Knowledge: Creation, Diffusion, Utilization*, 8, 509–520.
- Locke, E. A., & Becker, T. E. (1998a). Objectivism's answer to the sad, old world of subjectivism. *The Academy of Management Review*, 23(4), 658–659.
- Locke, E. A., & Becker, T. E. (1998b). Rebuttal to a subjectivist critique of an objectivist approach to integrity in organizations. *The Academy of Management Review*, 23, 170–175.
- Luhmann, N. (1990). *Die Wissenschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Lynch, M. (1993). *Scientific practice and ordinary action: Ethnomethodology and social studies of science*. Cambridge: Cambridge University Press.
- Lynne, G. D. (1999). Divided self models of the socioeconomic person: The metaeconomics approach. *The Journal of Socio-Economics*, 28(3), 267–288.
- MacKenzie, D. A. (1981). *Statistics in Britain, 1865–1930: The social construction of scientific knowledge*. Edinburgh: Edinburgh University Press.
- Mahoney, M. J. (1976). *Scientist as subject: The psychological imperative*. Cambridge: Ballinger.
- Malgady, R. G. (2000). Myths about the null hypothesis and paths to reform. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment* (pp. 49–62). Mahwah: Lawrence Erlbaum.
- McCloskey, D. N. (1998). *The rhetoric of economics* (2nd edn.). Madison: University of Wisconsin Press.
- McGuire, W. J. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1–47). Orlando: Academic Press.

- McNemar, Q. (1960). At random: Sense and nonsense. *The American Psychologist*, *15*, 295–300.
- McSwite, O. C. (1997). *Legitimacy in public administration: A discourse analysis. Advances in public administration*. Thousand Oaks: Sage.
- Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*, 195–244. Monograph Suppl. 1-V66.
- Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictors. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* Mahwah: Erlbaum.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, *83*, 340–363.
- Meyer, J. W., & Scott, W. R. (1983). *Organizational environments: Ritual and rationality*. Beverly Hills: Sage.
- Miller, P. (2007). Management accounting and sociology. In C. S. Chapman, A. G. Hopwood, & M. D. Shields (Eds.), *Handbook of management accounting research* (pp. 285–295). Amsterdam: Elsevier.
- Miner, J. B. (2003). The rated importance scientific validity, and practical usefulness of organizational behavior theories: A quantitative review. *Academy of Management Learning & Education*, *2*(3), 250–268.
- Mirchandani, R. (2005). Postmodernism and sociology: From the epistemological to the empirical. *Sociological Theory*, *23*(1), 86–115.
- Morris, M. W., Leung, K., Ames, D., & Lickel, B. (1999). Views from inside and outside: Integrating emic and etic insights about culture and justice judgment. *The Academy of Management Review*, *24*(4), 781–796.
- Morrison, D. E., & Henkel, R. E. (1970). *The significance test controversy: A reader*. Chicago: Aldine.
- Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
- Nelson, N., Rosenthal, R., & Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *The American Psychologist*, *41*, 1299–1301.
- Nelson, J. S., Megill, A., & McCloskey, D. N. (1987). Rhetoric of inquiry. In J. S. Nelson, A. Megill, & D. N. McCloskey (Eds.), *The rhetoric of the human sciences: Language and argument in scholarship and public affairs* (pp. 3–18). Madison: University of Wisconsin Press.
- Nicholson, N. (1997). Evolutionary psychology: Toward a new view of human nature and organizational society. *Human Relations*, *50*(9), 1053–1078.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*(2), 241–301.
- Nola, R. (2003). *Rescuing reason: A critique of anti-rationalist views of science and knowledge*. Boston: Kluwer Academic.
- Nunnally, J. C. Jr. (1960). The place of statistics in psychology. *Educational and Psychological Measurement*, *20*, 641–650.
- Nunnally, J. C. Jr. (1975). *Introduction to statistics for psychology and education*. New York: McGraw-Hill.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd edn.). New York: McGraw-Hill.
- Oakes, M. L. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Oliver, C. (1991). Strategic responses to institutional processes. *The Academy of Management Review*, *16*, 145–179.
- Orlitzky, M. (2011). Institutional logics in the study of organizations: The social construction of the relationship between corporate social and financial performance. *Business Ethics Quarterly*, *21*(3), 409–444.
- Orlitzky, M., & Benjamin, J. D. (2001). Corporate social performance and firm risk: A meta-analytic review. *Business & Society*, *40*(4), 369–396.
- Orlitzky, M., & Benjamin, J. D. (2003). The effects of sex composition on small-group performance in a business school case competition. *Academy of Management Learning & Education*, *2*(2), 128–138.
- Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. *Organization Studies*, *24*(3), 403–441.
- Park, S. H., & Gordon, M. E. (1996). Publication records and tenure decisions in the field of Strategic Management. *Strategic Management Journal*, *17*, 109–128.
- Pearce, J. L. (2004). Presidential address: What do we know and how do we really know it? *The Academy of Management Review*, *29*(2), 175–179.
- Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale: Erlbaum.

- Peters, D. P., & Ceci, S. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5, 187–195.
- Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *The Academy of Management Review*, 18(4), 599–620.
- Pfeffer, J., & Sutton, R. I. (2006). *Hard facts, dangerous half-truths, and total nonsense: Profiting from evidence-based management*. Boston: Harvard Business School Press.
- Platt, J. R. (1964). Strong inference. *Science*, 146, 347–353.
- Popper, K. R. (1962). *The open society and its enemies: Volume II: The high tide of prophesy: Hegel, Marx, and the aftermath*. London: Routledge & Kegan Paul.
- Popper, K. R. (1969). *Conjectures and refutations*. London: Routledge.
- Popper, K. R. (1972). *The logic of scientific discovery*. London: Hutchinson.
- Popper, K. R. (1983). *Realism and the aim of science*. London: Hutchinson.
- Porter, T. M. (1992). Quantification and the accounting ideal in science. *Social Studies of Science*, 22, 633–652.
- Powell, T. C. (2001). Competitive advantage: Logical and philosophical considerations. *Strategic Management Journal*, 22, 875–888.
- Priem, R. L., & Butler, J. E. (2001). Is the resource-based “view” a useful perspective for strategic management research? *The Academy of Management Review*, 26, 22–40.
- Reed, M. (1990). From paradigms to images: The paradigm warrior turns post-modernist guru. *Personnel Review*, 19(3), 35–40.
- Rindskopf, D. M. (1997). Testing “small,” not null, hypotheses: Classical and Bayesian approaches. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 319–332). Hillsdale: Erlbaum.
- Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *The American Psychologist*, 65(1), 1–12.
- Rorty, R. (1989). *Contingency, irony, and solidarity*. New York: Cambridge University Press.
- Rorty, R. (1991). *Objectivity, relativism, and truth: Philosophical papers*, Vol. I. Cambridge: Cambridge University Press.
- Rorty, R. (1997). *Truth, politics, and post-modernism*. Assen: Van Gorcum.
- Rosenau, P. M. (1992). *Post-modernism and the social sciences: Insights, inroads, and intrusions*. Princeton: Princeton University Press.
- Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *The American Psychologist*, 44, 1276–1284.
- Royall, R. (1997). *Statistical evidence—a likelihood paradigm*. London: Chapman & Hall.
- Rozeboom, W. W. (1997). Good science is abductive, not hypothetico-deductive. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 335–391). Hillsdale: Erlbaum.
- Rynes, S. L., Bartunek, J. M., & Daft, R. L. (2001). Across the great divide: Knowledge creation and transfer between practitioners and academics. *Academy of Management Journal*, 44(2), 340–355.
- Sastry, M. A. (1997). Problems and paradoxes in a model of punctuated organizational change. *Administrative Science Quarterly*, 42, 237–275.
- Scandura, T. A., & Williams, E. A. (2000). Research methodology in management: Current practices, trends, and implications for future research. *Academy of Management Journal*, 43, 1248–1264.
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *The American Psychologist*, 47, 1173–1181.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training and researchers. *Psychological Methods*, 1, 115–129.
- Schmidt, F. L., & Hunter, J. E. (1997). Eight common but false objections to the discontinuation of significance testing in the analysis of research data. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 37–64). Hillsdale: Erlbaum.
- Schmidt, F. L., & Hunter, J. E. (2002). Are there benefits from NHST? *The American Psychologist*, 57(1), 65–66.
- Schmidt, F. L., Hunter, J. E., & Urry, V. E. (1976). Statistical power in criterion-related validity studies. *Journal of Applied Psychology*, 61, 473–485.
- Schofer, E. (2004). Cross-national differences in the expansion of science, 1970–1990. *Social Forces*, 83(1), 215–248.
- Schwab, D. P. (1999). *Research methods for organizational studies*. Mahwah: Erlbaum.
- Schwab, A., & Starbuck, W. H. (2009). Null-hypothesis significance tests in behavioral and management research: We can do better. *Research Methodology in Strategy and Management*, 5, 29–54.

- Scott, W. R. (1994). Institutions and organizations: Toward a theoretical synthesis. In W. R. Scott & J. W. Meyer (Eds.), *Institutional environments and organizations: Structural complexity and individualism* (pp. 55–80). Thousand Oaks: Sage.
- Scott, W. R. (1995). *Institutions and organizations*. Thousand Oaks: Sage.
- Scott, W. R. (2001). *Institutions and organizations* (2nd edn.). Thousand Oaks: Sage.
- Sedlmeier, P., & Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, *105*, 309–316.
- Seth, A., Carlson, K. D., Hatfield, D. E., & Lan, H.-W. (2009). So what? Beyond statistical significance to substantive significance in strategy research. *Research Methodology in Strategy and Management*, *5*, 3–28.
- Shapin, S. (1994). *A social history of truth: Civility and science in seventeenth-century England*. Chicago: University of Chicago Press.
- Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle and the experimental life*. Princeton: Princeton University Press.
- Shaver, J. P., & Norton, R. S. (1980). Randomness and replication in ten years of the *American Educational Research Journal*. *Educational Researcher*, *9*, 9–15.
- Sober, E. (2005). Is drift a serious alternative to natural selection as an explanation of complex adaptive traits? *Royal Institute of Philosophy Supplement*, *56*, 125–154.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*, 173–180.
- Stinchcombe, A. L. (1965). Social structure and organizations. In J. G. March (Ed.), *Handbook of organizations* (pp. 142–193). Chicago: Rand McNally.
- Stinchcombe, A. L. (1968). *Constructing social theories*. Chicago: University of Chicago Press.
- Suchman, M. C. (1995). Managing legitimacy: Strategic and institutional approaches. *The Academy of Management Review*, *20*, 571–610.
- Sutton, R. I., & Staw, B. M. (1995). What theory is not. *Administrative Science Quarterly*, *40*, 371–384.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks: Sage.
- Tashakkori, A., & Teddlie, C. (2002). *Handbook of mixed methods in the social and behavioral sciences*. Thousand Oaks: Sage.
- Thompson, P. (1993). Postmodernism: Fatal distraction. In J. Hassard & M. Parker (Eds.), *Postmodernism and organizations* (pp. 183–203). Thousand Oaks: Sage.
- Tracey, T. J. G. (2000). Issues in the analysis and interpretation of quantitative data: Deinstitutionalization of the null hypothesis test. In S. D. Brown & R. W. Lent (Eds.) *Handbook of counseling psychology* (3rd edn.) (pp. 177–198). New York: Wiley.
- Tsang, E. W. K., & Kwan, K.-M. (1999). Replication and theory development in organizational science: A critical realist perspective. *The Academy of Management Review*, *24*(4), 759–780.
- Tsoukas, H., & Knudsen, C. (2003). Introduction: The need for meta-theoretical reflection in organization theory. In H. Tsoukas & C. Knudsen (Eds.), *The Oxford handbook of organization theory: Meta-theoretical perspectives* (pp. 1–36). Oxford: Oxford University Press.
- Turner, J. H. (2002). *Face-to-face: Towards a sociological theory of interpersonal behavior*. Stanford: Stanford University Press.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
- Van Maanen, J. (1988). *Tales of the field: On writing ethnography*. Chicago: University of Chicago Press.
- Van Maanen, J. (1995a). Fear and loathing in organization studies. *Organization Science*, *6*(6), 687–692.
- Van Maanen, J. (1995b). Style as theory. *Organization Science*, *6*(1), 133–143.
- Van Maanen, J. (1998a). Different strokes: Qualitative research in the Administrative Science Quarterly from 1956–1996. In J. E. Van Maanen (Ed.), *Qualitative studies of organizations* (pp. ix–xxxii). Thousand Oaks: Sage.
- Van Maanen, J. E. (1998b). *Qualitative studies of organizations*. Thousand Oaks: Sage.
- Wainer, H. (1999). One cheer for null hypothesis significance testing. *Psychological Methods*, *4*(2), 212–213.
- Weick, K. E. (1977). Editorial note. *Administrative Science Quarterly*, *22*, 138–139.
- Weick, K. E. (1989). Theory construction as disciplined imagination. *The Academy of Management Review*, *14*, 516–531.

- Weick, K. E. (1995). What theory is not, theorizing is. *Administrative Science Quarterly*, 40(3), 385–390.
- Weick, K. E. (1996). Drop your tools: An allegory for organizational studies. *Administrative Science Quarterly*, 41, 301–313.
- Weick, K. E. (2001). *Making sense of the organization*. Malden: Blackwell.
- Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance: How the standard error costs us jobs, justice, and lives*. Ann Arbor: University of Michigan Press.