ORIGINAL PAPER

Statistics and Probability Have Always Been Value-Laden: An Historical Ontology of Quantitative Research Methods

Michael J. Zyphur[1] · Dean C. Pierides[2]

Abstract

Quantitative researchers often discuss research ethics as if specific ethical problems can be reduced to abstract normative logics (e.g., virtue ethics, utilitarianism, deontology). Such approaches overlook how values are embedded in every aspect of quantitative methods, including 'observations,' 'facts,' and notions of 'objectivity.' We describe how quantitative research practices, concepts, discourses, and their objects/subjects of study have always been value-laden, from the invention of statistics and probability in the 1600s to their subsequent adoption as a logic made to appear as if it exists prior to, and separate from, ethics and values. This logic, which was embraced in the Academy of Management from the 1960s, casts management researchers as ethical agents who *ought* to know about a reality conceptualized as naturally existing in the image of statistics and probability (replete with 'constructs'), while overlooking that S&P logic and practices, which researchers made for themselves, have an appreciable role in making the world appear this way. We introduce a different way to conceptualize reality and ethics, wherein the process of scientific inquiry itself requires an examination of its own practices and commitments. Instead of resorting to decontextualized notions of 'rigor' and its 'best practices,' quantitative researchers can adopt more purposeful ways to reason about the ethics and relevance of their methods and their science. We end by considering implications for addressing 'post truth' and 'alternative facts' problems as collective concerns, wherein it is actually the pluralistic nature of description that makes defending a collectively valuable version of reality so important and urgent.

Discussions about research ethics have placed significant emphasis on the avoidance of harm to individual participants, downplaying the importance of reflective thinking, or reflexivity, throughout the entire research process (Greenwood 2016). This has meant that the question of how ethics are embedded in methods beyond the context of data collection has received relatively less attention. The absence of a discussion about the relationship between methods and ethics is especially noticeable in the case of quantitative

methods, because quantitative researchers often claim their methods are value-neutral, or 'objective.'

In response to this situation, our paper uses historical inquiry to show how specific values came to be embedded in quantitative research practices, concepts, discourses, and its objects/subjects of study. Furthermore, we describe how these values were obscured over time to make quantitative methods appear as if they were value-free (Ezzamel and Willmott 2014; Wicks and Freeman 1998). We show how statistics and probability (henceforth 'S&P') served as a conduit to formalize an assumed priority of *logic-then-ethics* by putting logic first and explicit ethical considerations second—a consideration of means and ends, the right and the good, and what researchers or practitioners *ought* to do was yoked to a priority of S&P.

This relationship between logic and ethics has not changed markedly since the 1960s (for insight, see Augier and March 2011; Khurana 2007). Only recently have debates in business ethics started to show how the logic of S&P

✉ Michael J. Zyphur
mzyphur@unimelb.edu.au

Dean C. Pierides
d.c.pierides@stir.ac.uk

[1] Department of Management and Marketing, University of Melbourne, Parkville, VIC 3010, Australia

[2] University of Stirling, Stirling FK9 4LA, UK

uncritically relegates ethics to a secondary consideration, exposing how quantitative 'rigor' and 'best practices' are attempts to naively depict reality in a value-neutral way (see debate by Zyphur and Pierides 2017, 2019; Cortina 2019; Edwards 2019; Powell 2019; see exemplars by Aguinis et al. 2010; Edwards 2010; Köhler et al. 2017). Indeed, that the world is somehow naturally constituted in the image of S&P is so taken-for-granted in quantitative research that it remains implicit in most other discussions about ethics with respect to methods (e.g., Panter and Sterba 2011).

Our purpose in developing an historical analysis is thus twofold. First, we will show how S&P were manufactured over time to embody the ethics, interests, and institutions of different times in a manner that can prevent researchers from seeing how this normativity is embedded in their methods and the images of reality that they generate—a classic fact-value distinction that many fail to see as a product of history. For example, our analysis shows researchers how images of the reality they propose to merely represent are the products of methods they have made for themselves, and that *S&P methods generate images of reality, not the reverse*. Second, we use this historical account to find where quantitative researchers can intervene in the ethics of their methods. Whereas our other work uses classic pragmatism to confront the logic of S&P by providing examples and making concrete suggestions (Zyphur and Pierides 2017, 2019), here we explain the historical production of the ethical problems created by S&P to identify where quantitative researchers can collectively intervene in the ethics of their methods. By showing researchers how their research practices (and the 'observations' or 'facts' they generate) have always been value- and ethics-laden, our paper can be used to grapple with the ethics of quantitative practices using history, rather than abstractions about universal ethics that often feature in discussions of quantitative methods.

The approach we take is called 'historical ontology' (Hacking 2002). This inquiry examines knowledge practices and 'styles of reasoning' by identifying the conditions that allowed them to emerge, or their 'conditions of possibility.' An historical ontology, unlike a 'history of ontology,' is thus able to show how concepts and practices were made sensible—as things that seem reasonable and desirable. Our inquiry begins in the 1600s and ends with the adoption of S&P as research methods in the Academy of Management in the 1960s. We show how the values and conditions of different eras defined conceptions of the contents of reality available for study and how researchers *ought* to study it by developing specific methods that produce specific *kinds* of images of reality. Adopting S&P brought with it an ethical obligation to endorse narratives about reality existing 'objectively' in the image of S&P, rather than a recognition that specific practices *produce* images of reality that embody the logic of S&P. Historical inquiry into S&P is uniquely able to describe how this emerged and provides researchers with a new basis for intervening to modify their own practices.

Historical ontology does this by disrupting the automaticity of being a specific kind of person at a given point in time (e.g., a 'rigorous researcher' defined by S&P), offering an opportunity to see how any foundations for being in the world, knowing about the world, or talking about the world are historically created and built alongside the practical demands of specific situations and ethical imperatives. We propose that by working inside the history of their concepts and practices, and with recourse to our historical analysis, researchers can better understand what they are doing as they *act out* the foundations for their work by thinking, speaking, and otherwise practicing the research methods that are their craft. Those who teach quantitative methods, and those who learn them, can draw on this account to understand and address the value-laden nature of their concepts, methods, and inferences.

Our analysis is also relevant for researchers wanting to contextualize notions of rigor (e.g., Rynes et al. 2001), those wanting to address questionable research practices, and the troubling post-truth or alternative-fact problems with which scientists and political bodies now contend. We advocate for reflexive, ethics-oriented quantitative approaches that can at once counter claims that any research is free from values—always a sign that a rich history involving ethics is being obscured—while also motivating people to defend the helpful values that constitute their research methods. It is precisely because methods and facts are valued-laden that it is so important to produce them based on our collective concerns rather than a dogmatic defense of value neutrality—for example, by insisting on an outdated and ahistorical notion of objectivity. It is partly the pluralism of descriptions and their uses that makes defending a collectively valuable version of reality so important.

Our account begins with a short methodological note on historical ontology and proceeds with a chronological analysis of successive ethical problems for the quantitative community to address. Each section explores different conditions that persist to the present day, thus pointing to specific ethical problems for which interventions can be made. Although we encourage the quantitative community to collectively debate how best to intervene in each of these problems, we conclude by being upfront about our own commitments. We advocate for a return to classical pragmatism (Wicks and Freeman 1998), encouraging researchers to think in terms of 'relational validity' and to scientifically inquire into their beliefs, including those which are considered foundational (Zyphur and Pierides 2017).

An Historical Ontology of Quantitative Research Methods

Inquiring into S&P can be difficult because its practices need to be analyzed from outside their own logic. The problem, as Hacking notes (2002), is that S&P involves a 'style of reasoning' that stipulates what is real and how to know about it. By setting the terms for inquiry, a style of reasoning becomes self-vindicating over time rather than an object of inquiry in itself (Hacking 1990, 1991, 1992). For example, asking *why* research is done with S&P often invites a tautological response based on its own logic (e.g., "Why do we use *p*-values and Type-I/II errors? Because they estimate the probability of errors in inference").

This tautology can be avoided by inquiring about how S&P came to exist at all—an historical analysis. Yet, unlike historical work on epistemology (e.g., Rowlinson et al. 2014), understanding how S&P emerged as a style of reasoning requires an inquiry into ontology—what exists. For this, we use historical ontology, which "is history of the present, how our present conceptions were made" (Hacking 2002, p. 70). Historical ontology enables this by recognizing that "nothing, not even the ways I can describe myself, is either this or that but history made it so" (Hacking 1986, p. 37). The formation of concepts such as objectivity, fact, or truth can therefore be studied historically even if they seem timeless or 'natural.' As with everything made by people, S&P emerged in specific social, conceptual, and material contexts. To examine these is to describe conditions that made S&P possible, and in turn show how scholars made for themselves the *kinds* of images that S&P produce.

Styles of reasoning such as S&P rely on conceptions of what can be known, how to know about it, ways of being right and wrong, and how to speak and act through scientific practices (Hacking 2009). Thus, historical ontology allows us to investigate what constitutes notions of 'rigor' or 'best practices' of a style, together with the many ethical commitments that made adhering to them both possible and seemingly crucial or contentious. We now offer an historical ontology of S&P that draws attention to the objects, concepts, and contexts that forged an association between ethics, knowledge, and uncertainty to produce today's S&P and the ethical commitments that researchers are expected to uphold through their practices.

A New Reality and Researchers to Know It

How is it possible that researchers can quantify representations of nature (e.g., to conceive of 'models') and use S&P to address uncertainty in their knowledge about these quantifications (e.g., using *p* values) as though S&P are 'objective' rather than ethics-laden tools for generating images of reality? To do this required a series of fundamental changes in social and material relationships from the 1600s. Knowledge and opinion had to shift from being applied to separate *kinds* of things (Bromhead 2009; Byrne 1968; Shapin and Schaffer 1985; Shapiro 1983), to becoming a continuum of knowledge in which certainty and opinion were opposed. Where knowledge had previously been linked to deductive certainty, it came to be associated with opinion, which meant sense experience and uncertainty.

Defined by the church, knowledge before the 1600s had been certain and applicable to ideal forms, true 'essences,' or 'fundamental causes' such as triangles or spirits that were deductively 'demonstrated' without relying on sense experience or anything that *represented* an external reality. Knowledge centered on *resemblance* (e.g., walnuts could be a cure for brain maladies because they look like brains; Foucault 1970). Alternatively, opinion was a matter of the senses and rhetoric, going beyond resemblance by *representing* states of affairs, with representative legal testimony being a model for this (Shapin and Schaffer 1985). Thus, opinion could not be knowledge, and knowledge was owned by the church (Williams 2005).

Yet, the end of the 1600s saw the church replaced by a new kind of empirical science, and 'facts' emerged as representational. The prediction of uncertain events was a key location for this new arrangement, as in the emerging 'lower' sciences like medicine, chemistry, and biology. To handle the non-deducible, non-demonstrable nature of their subject matter (e.g., causes and cures of disease studied by observation), they generated new kinds of uncertainty and fears of being wrong (Daston 2005). New ways to work with causes were used, such as observable 'signs' that *indicated* underlying causes based on relative frequencies (Hacking 2006). For example, rashes appeared before death by the plague, so they were *signs* of an *underlying* malady due to the *frequency* at which they predicted death. This representational view of observations linked opinion and knowledge by introducing uncertainty.

This is a beginning for S&P: observations were 'internal evidence,' pointing beyond the observed to 'underlying' causes (Foucault 1970; Hacking 2006). Causes and effects were brought into a domain that could not be deduced or demonstrated with certainty (Shapin and Schaffer 1985). Instead, evidence linked opinion and knowledge by connecting events-as-signs to underlying causes at some relative frequency (Hacking 2006), such that a knowledge of causes and their future effects became linked to observation and inference. Thus, opinion mingled with knowledge, and event frequencies were used to make inferences with evidence that was uncertain because it *represented* what could not be directly observed.

What had to happen in order for a person to know reality in this way? To ground this new kind of knowledge, two moves were made: scientists theorized the *kinds* of things they could know; and with a new theory of experience they conceived of themselves as knowers (Dear 2001). For this, figures like Galileo recast biblical texts as 'nature'; a material inscription of a deity's word—the 'book of nature' (Biagioli 2006). Descartes also recreated deific laws as 'laws of nature' or 'causes' with a mechanical and deterministic worldview (Henry 2004; Truesdell 1984). Thus, scientists could proceed with the view that their objects of study were natural, allowing access to their deity's laws without upsetting the church (Williams 2005).

To assist this, the modern 'fact' as a discrete, quantified unit of knowledge emerged with the model of double-entry accounting (Poovey 1998). The 'fact' was exported to the new sciences, making nature something that was pre-organized into discrete, quantitative, accountable units. Indeed, as Galileo noted in 1623, "[p]hilosophy is written in this all-encompassing book that is constantly open before our eyes… It is written in mathematical language" (2008, p. 183), allowing it to be theorized as full of knowable facts.

New facts had to be gathered, so a new theory of experience turned people into the kinds of things that could know facts. This took the church's form of *public* knowledge and placed it within individuals as a *private* experience of nature. For the first time, figures like Descartes cleaved a public ethical 'conscience' to theorize a private 'consciousness' (Hennig 2007), proposing that nature could be known by the 'mind.' The mind could know by observing its own ideas from sense impressions and, thus, just like facts, could represent a natural reality (Alanen 2003). This move retained a place for a religious soul, inventing the mind as a spiritual substance that still exists today as an object of study for psychologists and philosophers (see Caton 1973; Dewey 1929; Rorty 1979; Williams 2005).

To further ground their science, Descartes and others desired to separate 'natural' and 'mental' attributes of perceptions. They proposed 'primary' qualities like shape (natural) and 'secondary' qualities like color (mental), and science was only meant to investigate what was natural, which meant quantifiable (Garber 1992; Schouls 2000). This came with rhetorical devices that made nature appear to speak for itself (Shapin 1984; Shapin and Schaffer 1985)—the dry impartial or 'objective' language of science today. With the individualization of knowledge and a rise of experimentation, new ways to embed trust into scientific discourse were also invented, including review practices by the peerage, or 'peer review,' an explicit reference to noble, landed gentlemen (Biagioli 2002; Shapin 1984, 1994).

Consequently, when John Locke invented a distinctly modern empiricism at the end of the seventeenth century, he could cast researchers with minds to behold a nature made up by discrete quantifiable units (Mandelbaum 1964). Nature had laws or 'causes' expressed in hypotheses and tested by observations that *represented* a natural reality, only true when they *corresponded* with it (Shapiro 1983). This conflation of a theory of knowledge with a theory of perception came to dominate science (Rorty 1979), putting a focus on what would become rigorous methods (Schouls 1979). Uncertainty reigned because the mind separated a natural reality from knowledge of it. As Locke noted, only 'probable knowledge' or 'moral certainty' were possible (Osler 1970; Shapiro 1983), creating fears of being wrong (Daston 2005) that led to codes of conduct 'among gentlemen,' which were ethical guidelines for the practice and discourse of research (Shapin 1994, 2008). Thus, a researcher treating uncertainty in knowledge with representations of nature was made possible, allowing S&P to later emerge as ethical tools for doing research as if nature came pre-packaged in the image of S&P.

Understanding that a representational theory of nature and a correspondence theory of truth are contingent historical developments is inconvenient for quantitative researchers who prefer to claim that the S&P reality these produce is the only reality that exists. Quantitative researchers who recognize that it is an ethical problem to erase this history can start to intervene in the institutional arrangements that create the kind of researcher who defends this reality, thus remaking quantitative methods and quantitative researchers with ethics being a central concern. For this project to be viable, it needs to be coupled with an understanding of how probability emerged as a new kind of religion by incorporating the ethics of a deity.

The Emergence of Probability

Prior to the 1660s, probability meant something similar to probity or approbation—that something was probable meant that there were good arguments for it (Hacking 2006). In the absence of knowledge, which previously had to be certain, there was probability as a guide for thought and action. As modern probability emerged, it kept this justificatory and ethical status, associated with right action, good judgment, social justification, and 'reason' (Gigerenzer et al. 1989). To fit the new kinds of research emerging in the seventeenth century, ethics had to conform to quantification—as was the case with facts (see Daston and Vidal 2004).

Pascal began probability's quantification by studying dice games (Galavotti 2005), developing an "arithmetical triangle to determine how the stakes should be divided between two players playing for a set of games" (Pascal 1653/1952, p. 460). It was the first proven combinatorial mathematics for calculating "perfect indifference for [wagering parties]" (1654/1952, p. 481, 487). Gambling was common, so it was

important for gamblers to know the expected frequency of an event and 'choose' using probabilistic expectations—a 'choice' was a bet. For the first time, knowledge, events, and action could be linked to a coherent quantitative tool that frequently worked in relation to inherent uncertainty (Hacking 2006). Soon, a text on the new method appeared, deriving "my Expectation to win anything… the value of my Expectation," to guide action in games of chance (Huygens 1657/2010, pp. 2–3).

Thus, a quantified probability emerged having two faces; the expected frequencies of events *versus* what people could say they knew, what they should believe, or how they should act (Hacking 2006). In turn, expected frequencies of uncertain events could be linked with ethics—what one should believe, should decide, or should do (as in Leibniz 1678/2004). This two-faced nature of probability—expected frequencies of uncertain events versus knowledge, belief, or choice/action—was its hallmark (Hacking 2006). In turn, games of chance became *the* model for a probabilistic ethics as it evolved to guide belief and action.

Yet, to make the new quantified uncertainty useful, another major shift was required: the model of uncertainty from games of chance had to be applied to other things in the world that were uncertain, even if they had nothing in common with the games. This began in 1662 at the end of *Logic, or The Art of Thinking* (Arnauld and Nicole 1662/1996). Also called *The Port Royal Logic*, it was "[t]he most influential logic book after Aristotle and before the end of the nineteenth century" (Hacking 1975, p. 26). In a final section by Pascal, the terms 'probable' and 'probability' appear for the first time with a modern, ethical meaning that today researchers can understand as being an early model for their own present-day practices. To intervene in the production of a *probabilistic logic-then-ethics* priority requires an understanding of how probability came to dominate research methods, as well as the strange use of probability from games of chance as a model of daily life.

This relies on expected distributions of events, with J. Bernoulli deriving the first sampling distribution to infer expectations for any number of coin flips (a binomial). In his *Art of Conjecture* (1713/2006), he connected probability-as-approbation with probability-as-expectation: "Probability… is degree of certainty… Probabilities are assessed according to the *number* together with the *weight* of the *arguments* that prove or indicate…; *arguments* are either internal or external. Internal… arguments are taken from the topics—cause, effect, subject, associated circumstance, sign" (1713/2006, pp. 315–318). Here was a new way to treat uncertainty, linking internal evidence, signs, and causes. Yet, it came with a warning: "the probabilities of things can be reduced to calculation… From this is resulted that the only thing needed for correctly forming conjectures on any matter is to determine the numbers of these cases accurately and then to

determine how much more easily some can happen than others [such as public health outcomes]. But here we come to a halt, for this can hardly ever be done. Indeed it can hardly be done except in games of chance" (1713/2006, p. 326).

This warning went unheeded as probability swept across society and the emerging sciences (Daston 1987a; Hacking 2006; Pearson 1978). In law, it defined evidence 'beyond a reasonable doubt' and determined the number of jurors for a fair trial (Shapiro 1983); in government it helped define democracy in elections (Stigler 1986); in science, claims could only be made if large probabilities existed—reported as 'moral certainty' or 'facts' (Shapin and Schaffer 1985). Probability emerged as *the* way one *should* deal with uncertainty, creating an ethic of 'rationality' for action that was based on games of chance (Gigerenzer et al. 1989).

Probability's ethicality is in Pascal's famous wager; it bets on the existence of a deity that was certain when the seventeenth century began. With a basis in dice games, probability had emerged with an ethic of belief, knowledge, and action under uncertainty, with new ways to constitute ethical agency using probability. The result is that S&P as an ethic had become possible, creating foundations for a future quantitative logic. A research practice could thus be 'best' or 'rigorous' if it increased a quantified *probability* of being correct. However, the warning of J. Bernoulli was forgotten along the way. To take ethics seriously, quantitative researchers can intervene in their own practices and training by recognizing that probability derived from games of chance may have little or nothing to do with what is ethical in a given real-world situation. To understand this better requires further historical study of probability.

The Probability of Causes

With probability in place and increasingly used in the 1700s, it was modified to solve new problems (Gigerenzer et al. 1989). There were two issues here. First, with knowledge aimed at causes, the goal was to work 'inductively' from observations to deterministic laws or 'causes' (Daston 1995). Bernoulli's binomial was a 'deductive' way to estimate probabilities of events given a known chance setup—a sampling distribution. Thus, he could not 'invert' the deductive tool for inductive inference (Gigerenzer et al. 1989; Stigler 1986).

Second, with representational theories of science, what probability represented was a question: 'relative frequency of events' or 'degree of knowledge or belief'? In mechanical views of nature unexpected events or unknown causes had to be deterministic, so probability had to be a 'measure' of knowledge or belief (Daston 1995; Gigerenzer et al. 1989; Hacking 2006; Kamlah 1987; Porter 1986). Thus, Hume's (1739) 'problem of induction' placed uncertainty in the

researcher and Hume proposed that chance events were different than probability, which measured degrees of knowledge (Hacking 1978). In turn, probabilistic reasoning from observations to causes was a part of the mind that could be mathematized even if reality was deterministic (Hume 1739, pp. 125–130).

Hume poked fun at causal certainty with this view of probability (Hume 1739, p. 124), noting that a 'cause' was not a feature of nature, but of *human* nature: people probabilistically reason about causes not to be certain, but to develop practical beliefs. A cause was a 'habit' of reasoning that became a model for science: with more observations, an increase in knowledge or belief about causes can inform action (Hume 1739, p. 130). Its logic also highlighted an intractable paradox that became foundational for S&P: an infinite number of observations would be needed to be certain about causes.

A first attempt to address Hume's problem is, today, called *Bayes' rule* (from Bayes 1763; see Zabell 1989), which was popularized by Laplace (1774/1986). The aim was "to determine the probability of causes of events… [It] is principally from this point of view that the science of chances can be useful" (Laplace 1774/1986, p. 364; see Dale 1999; Hald 2007; Pearson 1978). This rule led probability to guide inferences in a new way, for "the most important questions of life are indeed, for the most part, only problems of probability… the principal means of arriving at the truth—induction and analogy—are based on probabilities…" (Laplace 1825/1995, p. 1). The result is that, by the turn of the eighteenth century, probability from games of chance could be used to treat uncertainty in scientific inferences, and these inferences could be evaluated based on an ethic of minimizing uncertainty. The continuing question that researchers should be asking themselves, however, is whether a probability calculus has much bearing on whatever it is they are studying, and why they treat probability as if it were essential for science—a topic we now further historically excavate.

Laws and Limits of Truth, Knowledge, and Error

Although probability was used in many ways in the 1800s, in research it was often linked to measurement (Daston 1995; Stigler 1986). Precise measurements were needed to evaluate physical models and to standardize weights and measures, the latter of which was pivotal for organizing modern societies. To maximize precision, scientists often took multiple measurements, but usually chose the most trusted datum rather than averages due to fear that errors were additive (Stigler 1986), a practice that was to be transformed by

reasoning with probability (e.g., Lagrange 1770–1773/2009; see also Pearson 1978).

Measurements here were of physical objects that were known to exist (e.g., the location of planets). In turn, variation in measurements was construed as 'error.' The major innovation was applying probability to this error (Stigler 1986). The result was that researchers became ethically obligated to use probability for treating what was 'true' versus 'error'—using a logic of coin tosses, observations were analogized as uncertain events.

This resulted in a new way to treat uncertainty, with 'error laws' or 'laws of error' applied to observations (Porter 1986; Stigler 1986). By mapping a research practice such as measurement onto a chance process like a coin toss, it was possible to prove mathematically how often different levels of error would occur. With a theory of what was true (or a cause) as the opposite of error (or chance), uncertainty about true scores could be treated with laws of error. The deific 'laws' were distributions that were derived from games of chance. These became the first 'central limit theorem,' which modeled precision in measurement (Adams 2009; Fischer 2010). The point was that "[t]his artifice, extended to some arbitrary laws of [chances], gives a general method to determine the probability that the error of any number of observations will be contained in some given limits" (see Laplace 1809a/2011, p. 5).

In turn, a new way to reason emerged with an ethical imperative for research practices to maximize the 'true' and minimize 'error' (Porter 1986). This ethic did two things. First, it coupled what would become 'true scores' with averages, large sample sizes, and probability distributions. For example, a planetary "orbit should not be taken from single observations, but… from several so combined the accidental errors might… mutually destroy each other… [I]t will be proper to take the arithmetical mean… and afterward to free it from the mean error… [A]ccordingly, the probability to be assigned to each error will be expressed by a [probability] function" (Gauss 1809/2010, pp. 250–259). This function was a distribution, which at the time was a law of errors that defined what researchers could fear: "we will name it curve of probabilities… [and] we will observe that [the average] point is the one where the deviation from the truth, which we can fear, is a minimum" (Laplace 1809b/2010, p. 2).

Second, all of this emerged with a new tool that used probability to justify its use: "the most probable system of values of the quantities… will be that in which… *the sum of the squares of the differences between the actually observed and computed values multiplied by the numbers that measure the degree of precision, is a minimum*" (Gauss 1809/2010, p. 260). This was "the celebrated *method of least squares*" (De Morgan 1838/2010, p. 155), which still defines and justifies regression analysis. The result was that scientists became obligated to use averages, large samples, and estimation

tools like regression to reduce 'error.' More importantly, it had become possible for a new logic to emerge by ethically dictating a research practice which *uses a probabilistic justification* to access true scores and minimize error. By treating probability distributions as laws, it was possible for probability to reflect what could be expected in relation to truth when doing research, which at once was an ethical imperative and part of a larger foundation on which the social sciences would emerge.

At this point, perhaps a useful intervention for researchers would involve grappling with whether or not the model of planetary bodies and uncertainty in their location has much to do with the ethics of social scientific research practices. To get a grip on some of the issues involved requires historically understanding how this shift—from celestial bodies to local political and national bodies—was initiated and maintained as if this were unproblematic.

Statistics and the Social Sciences

Around the same time (1800–1850), political revolutions and collective movements put a focus on people and their welfare, or 'society' and 'the state' (Cohen 1987; Hacking 1987). This interest developed as censuses emerged on large scales (Hacking 1990, p. 118). Data analysis and inference were done by "statists" who studied society and the state with statistics (Hacking 2006, p. 102; Porter 1986, p. 24)—the first United States census asked four questions; by 1880 there were 13,010 (Hacking 1991).

These social numbers also emerged with distinctly modern notions of objectivity and subjectivity (Daston 1992, 1994; Daston and Galison 2007; Swijtink 1987). In turn, notions of objectivity were used in order to be convincing by using number-as-fact, while trying to hide any participation in the process of writing census questions, analyzing data, and drawing inferences based on the new ways to produce images of people and society (see Alonso and Starr 1987; Desrosières 1998; Poovey 1995, 1998; Porter 1986, 1995; Woolf 1989).

New data and analyses created new kinds of people and social entities because each new question created new ways to view and treat people (Hacking 1990, 1991, 1999, 2002). A good example is Quetelet's first 'average man' (1842/2010)—the first average of a social group imbued with a kind of real status. This led to a new idea of 'normal' by applying the normal curve to cluster people *as* a distribution (Foucault 1970, 1980, 2003, 2008). Thus, statistics emerged as standardization tools that enacted ethics of relative comparison—consider today's *Diagnostic and Statistical Manual of Mental Disorders* (APA 2013).

New sciences emerged borrowing ideas from the natural sciences to study the new 'social' and 'mental' attributes being 'measured' (Foucault 1970; Goldman 1983; Krüger et al. 1987; Ross 1991). The 'social physics' of Quetelet was pivotal. He applied the logic of astronomy to study 'laws' and 'forces' governing 'society' (Gigerenzer et al. 1989; Hacking 1990; Stigler 1986). Averages were true scores (as in the location of a planet) and variance around them was error governed by a law (i.e., the distribution around the average was a kind of physical error; Lécuyer 1987).

The idea was profound: people from the same province or country were natural kinds that defined a distribution, or a 'population' that was subject to one set of laws or causes. This application of probability was revolutionary: by clustering people together based on a local system of classification—itself always ethics- and value-laden—the cluster could act representationally to indicate not the researcher's classification scheme, but rather as an objective part of a social 'nature'—as in Descartes' primary qualities of physical objects.

With his tools, Quetelet created new ways to answer questions such as, "Are Human Actions Regulated by Fixed Laws?[,]… How the Laws relative to Man ought to be Studied and Interpreted[, and]… the Causes which Influence Man." This reinvented 'man' as "under the influence of regular and periodic causes, affecting not merely his physical qualities, but likewise his action… Now, these causes, and their mode of action, or the laws to which they give rise, may be determined by a close inquiry… with respect to his moral and intellectual qualities" (1842/2010, pp. 7–9). For the first time, social objects emerged that were *produced in the image of S&P*. This is a useful place for generating ethical interventions: it is crucial for quantitative researchers to understand that their objects of study and results are always based on classification, measurement, and analysis schemes they develop for themselves.

Again, simply because S&P research produces specific quantitative results does not mean that nature itself is pre-packaged in the form of S&P, even if the new social 'sciences of man' or 'human sciences' could show that social data were statistically regular in the aggregate by fitting probability distributions (Porter 1986; Stigler 1986). The result was that natural laws and causes could be known about 'social' and 'mental' 'structures' that emerged in large data. Yet, all of this borrowed probabilistic metaphors from astronomy, which were based on coin tosses. In turn, people were theorized as being subject to natural causes, with effects that were studied by treating variation as error (Hopwood et al. 2010). Thus, as the social sciences were formed, the focus was on explaining aggregates—whether psychophysical or macrosocial (Danziger 1990; Duncan 1984; Krüger et al. 1987). In turn, statistics tools were institutionalized as a way to understand true scores and causes that were theorized as existing apart from the S&P logic that helped create them (Hacking 1990).
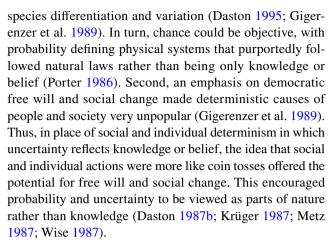
By proposing that people or social groups reflected unobserved true scores or causes in populations that could be studied with S&P, new ways to ethically evaluate research and other practices emerged (Foucault 2008; Hacking 1990). In sum, key parts of social science had become possible, with conceptual foundations, methods, and objects of study arriving to create the possibility of S&P with an ethic of representation for true score or causes—necessitating large samples and managing probabilistic error as a matter of ethics, thus creating the conditions for defining and narrowing the practices that would later be considered ethical or unethical. Again, here is a place for intervening in the generation and use of quantitative methods: researchers must place themselves within the research process, as ethical agents that actively produce images of people and society that are consistent with the ways they are theorized (Greenwood 2016). Whether based on psychologistic notions of a 'mind,' or economistic 'utility' or 'preference,' or sociological 'structures,' are the specific characterizations being produced with 'measurement' tools ethical?

The Birth of Frequentism

As social sciences developed in the 1800s, there was a new interest in formalizing probability, which had always been two-faced: *statistical*, concerning relative frequencies; and *epistemic*, concerning beliefs or knowledge (Hacking 2006). In the 1830s, these were defined with two calculi: 'deductive' or 'direct' as a relative frequency; and, 'inductive' or 'inverse' as a probability of causes with Bayes' rule (e.g., de Morgan 1838/2010, pp. 30, 53).

By the 1840s, the two types of probability started to become two approaches to science. This helped to initiate modern notions of objectivity and subjectivity, ushering in a critique of Bayes' rule and the inverse probability of causes for using 'prior' knowledge or beliefs, rather than 'objective' events (Daston 1994, 1995; Kamlah 1987; Porter 1986; Strong 1978). This critique began as part of a larger effort to create a 'philosophy of science' meant to manage research in the 1800s (Dewey 1929; Rorty 1979). In turn, philosophers were empowered to invent and work with concepts like objectivity and subjectivity to manage science (Daston 1992, 1994; Daston and Galison 2007), allowing a critique of Bayes' rule and inverse probability. Two features of the nineteenth century helped this critique work.

First, science embraced indeterminism with statistical theories of nature. Maxwell and Boltzmann invented statistical mechanics, describing emergent properties of particles by drawing on Quetelet's social physics of stability in aggregates and random error as a law in a population (Daston 1995; Gigerenzer et al. 1989; Porter 1994). Darwin also used Quetelet's ideas to theorize chance as a cause of

species differentiation and variation (Daston 1995; Gigerenzer et al. 1989). In turn, chance could be objective, with probability defining physical systems that purportedly followed natural laws rather than being only knowledge or belief (Porter 1986). Second, an emphasis on democratic free will and social change made deterministic causes of people and society very unpopular (Gigerenzer et al. 1989). Thus, in place of social and individual determinism in which uncertainty reflects knowledge or belief, the idea that social and individual actions were more like coin tosses offered the potential for free will and social change. This encouraged probability and uncertainty to be viewed as parts of nature rather than knowledge (Daston 1987b; Krüger 1987; Metz 1987; Wise 1987).

In sum, uncertainty could now be treated as a part of nature. In this context, the first book with a frequentist theory of probability that owed much to Quetelet's ideas of random variation in homogenous 'populations' was written by Venn in 1866 (see 1866/2006). In turn, probability as knowledge or belief began a descent into obscurity in what would become S&P practices, and "textbooks of the Laplacian type (i.e., Bayesian) became rarer and finally disappeared" (Kamlah 1987, p. 111; McGrayne 2011; Schneider 1987). This was the start of a revolution that created a statistics wherein uncertainty was natural or 'objective.' In turn, it became possible to adopt a frequentist theory of probability based on ethical grounds; that this was what one *should* do to objectively reflect laws of nature. As we show next (see also Petit 2013), it took work to institutionalize a 'population' and a 'sample' as natural features of reality, rather than ethic- and value-laden objects conceived by and for researchers.

Variance, Correlation, and Regression

Although nature was now a chancy thing, a social–physics model of reality restricted research to true scores or causes as averages, treating variation as error. Evolutionary theory rejected this (Ariew 2007), suggesting that 'chance' determined variation but that it was not error. Darwin legitimized chance as a topic for study by noting that the parts of a distribution otherwise considered erroneous could be very valuable (e.g., high intelligence).

Darwin's half-cousin, Galton, helped make variation itself an object of inquiry in studies of *Hereditary Genius: An Inquiry into its Laws and Consequences* (1869/2010), which became today's 'cognitive ability' (Hacking 1990). To understand such traits, ideas of co-relation or 'correlation' among parents and offspring emerged with notions of natural laws governing 'reversion to the mean' or 'regression' across generations (Galton 1889/2010, pp. 95–137). Along with mathematically savvy researchers like K. Pearson, Galton used a

variety of ways to argue that variation was a legitimate topic for study and prediction (Stigler 1986).

It then became possible to start talking about relationships among social variables and 'accounting for variance,' even though variation could still be tied to notions of chance. As Porter notes, "[the law of error's] reinterpretation as a law of genuine variation, rather than of mere error, was the central achievement of nineteenth century statistics" (1986, p. 91). Yet, the way this new kind of statistics was used reveals a feature of its logic that would follow it into the future: Darwin's and Galton's point was that some variations were better than others. The goal of correlation, regression, or anything else that treats variation became prediction and control, due to valuing the ends of a distribution (e.g., high intelligence; see MacKenzie 1981). To do this while appearing scientific required cleansing numbers of their ideological and value-laden origins by talking about natural laws or causes and forces that were facts, not values (Daston and Vidal 2004; Douglas 2009; Porter 1995; Shapin 2008, 2010).

It is therefore telling to examine the first estimation of an effect as a regression coefficient using the least squares method that was developed almost 100 years earlier to estimate the unknown locations of planets, treating a 'causal effect' as if it were a material object in an unknown location that could be represented in a regression equation. Using census data, Yule justified an argument against social welfare by estimating an effect purportedly showing that social welfare caused poverty (Yule 1897, 1899; see Stigler 1986).

This inference resulted as a 'partial regression' coefficient (Yule 1897, p. 833), which could 'control' for confounding factors by keeping them 'constant' (Yule 1899, p. 262; see also Freedman 2005). Thus, by the close of the nineteenth century, it was possible to speak of objectivity and statistical laws and apply these ideas to 'estimate' causal effects among variables using a tool developed for estimating the locations of material objects. Also, all of this could be used to *represent* the many things that were invented in the 1800s, such as 'society,' mental and social 'structures,' and their 'unobserved' 'causes.' This is a crucial place for ethical interventions in quantitative research: statistical estimation tools are touted as representing an underlying social reality, with an ethical agenda in which statistical estimates *ought* be used to represent 'true effects' theorized as constituting reality. However, of course, such representations only appear *as the results of S&P logics and practices*, which should prompt the consideration of a more practice-based understanding of any S&P reality.

Disciplining Statistics

As the twentieth century dawned, the use of statistics was fragmented across its applications (Stigler 1999). To discipline statistics and to professionalize it, it needed to become

a singular thing, organized by 'foundations.' This is to say that statistics had to become a *style of reasoning* that could be called objective by being stripped of its diversity, context, and social origins, so that it appeared valid for any problem because it could determine the ways a problem was constructed and understood (Abbott 1991). Put differently, objectivity would have to come from statistics itself rather than the contexts, problems, and practices where it was applied (Hacking 1992). This was done by inventing and normalizing things such as statistical theory, inference, and hypothesis testing. For this, various ideological, personal, and technical battles had to be fought and then forgotten, two of which stand out for ethical analyses (Gigerenzer et al. 1989; Howie 2002; Stigler 1999).

The first battle was between the generations of K. Pearson and Fisher. Pearson was older, still using inverse probability with Bayes' rule, and as editor of *Biometrika* he had critiqued Fisher but would not publish Fisher's response (Howie 2002). Fisher sought revenge and spent years on a frequentist logic that banished Bayesian probability and helped statistics emerge as a discipline (see Fisher 1921, 1922a, b, 1925c; see also Aldrich 1997, 2008; Zabell 1989, 1992). He worked to ensure that all aspects of research would have to conform to his frequentist probability logic by connecting what had been dispersed scientific practices: research design, data analysis, and statistical inference (Howie 2002).

Three aspects of Fisher's logic helped it work. First, it had a familiar ethics, with design, analysis, and inferential tools that *should* be used based on probabilistic arguments about estimating true scores or true causal effects (see 1922a, 1925c). Second, it was justified by being mapped onto models and experiments: correlational data could produce true effects if models were specified correctly, which meant controlling for confounds using regression (1922a, 1925c); experiments could produce true effects by making the world act like a coin toss with random assignment and frequentist hypothesis tests (1925a, 1935a). Third, Fisher gave expected distributions for his statistics (e.g., 1915, 1921, 1922b, 1924, 1925a, 1928, 1929), facilitating a logic of inference with a null hypothesis and p values of 0.05 or less for significance tests (1925b, 1935a). To complete the offering, Fisher published tables of p values that any researchers could use (e.g., Fisher and Yates 1943). This made Fisher a lot of money while making his tools accessible in a way that S&P had never been before.

In sum, Fisher produced a coherent logic that allowed what could be called objective tests of underlying mechanisms or causes *of any kind*, pronouncing "empiricism is cleared of its dangers if we can apply a rigorous and objective test," by which he meant his frequentist descriptions (Fisher 1922a, p. 314). In turn, statistics was disciplined by offering many of today's notions of 'rigor' and 'best

practices' that researchers feel ethically compelled to obey without knowing much about their history. Thus, as statistics was disciplined, so was the researcher: to be 'rigorous' meant using Fisher's frequentist logic that linked design, analysis, and inference without regard to the context of a specific research question.

After Fisher's logic had reached this status, the second battle was over what statistics would do. Fisher emphasized scientific inference and knowledge (1935b), but Neyman and E. Pearson wanted a logic for practical decision-making and behavior (1933; Neyman 1934). In brief, Fisher cast statistics as scientific tools for generalizing from a sample to an uncertain world for the purpose of generating knowledge. It was an "inductive logic… reasoning from the sample to the population from which the sample was drawn, from consequences to causes, or in more logical terms, from the particular to the general" (Fisher 1955, p. 69).

The Neyman and E. Pearson approach formalized action via statistical hypotheses that were not meant to create abstract 'knowledge,' for "[w]ithout hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour with regard to them" (1933, p. 291). The idea was "[t]he problem of testing a statistical hypothesis occurs when circumstances force us to make a choice between two courses of action… to accept a hypothesis H means only to decide to take action A rather than action B" (Neyman 1950, p. 259), such as the problem of deciding whether to retool a factory based on sampled products on an assembly line. This treated hypothesis tests not as acts of inductive inference, but as decision-making to avoid errors in behavior and outcomes: Type-I/II errors.

In the debate between these positions, there was vitriol. Fisher described the logic of Neyman and E. Pearson as "the phantasy of circles rather remote from scientific research" (1956, p. 100). Neyman called Fisher's procedures "worse than useless" (Hacking 1965, p. 99; see also Neyman 1956, pp. 289, 292). In turn, statistics inherited a combination of Fisher's logic of modeling and experiment, with Neyman and E. Pearson's hypothesis tests to decide among two options and avoid Type-I/II errors (a null and alternate hypothesis today), all of which relied on a logic of sampling, representation, and correspondence that produced the possibility of errors due to multiple hypothesis tests (see Wald 1945, 1950).

Within the battles, there was often agreement on key points: "[t]he statistician is concerned with a population… which for some reason… cannot be studied exhaustively" (Neyman 1937, p. 347); and "[s]tatistics may be regarded as (i) the study of populations, (ii) as the study of variation" (Fisher 1925b, p. 1). By focusing on drawing inferences from 'samples' to 'populations,' Quetelet's notion of a collection of social entities as a natural kind persisted, including his 'average man' governed by natural laws or causes (Gigerenzer et al. 1989). By the end of World War II it had become possible to draw on the tools and discourse of frequentist S&P. These purported to offer access to true scores and true causal effects via logics of modeling and experiment, which were coupled with inferential tools that could be used by simply finding p values in a table to decide among two competing options: a null or alternate hypothesis. Also, specific ideas such as Type-I/II error were put into place, almost fully forming the basis for the probabilistic ethics of 'rigor' and its 'best practices.'

Here again is a place for ethical intervention, by considering whether most social sciences, and certainly the study of business and organizations, can benefit very much from simple yes/no decisions to make inferences about a reality produced by looking at the world through a lens of S&P. Every real problem addressed in a real situation is specific, and therefore abstract logics involving hypothetical 'true scores' or 'populations' will very rarely be relevant for figuring what to do in a specific situation. Unfortunately, without access to case-based training about how to reason, or casuistry, quantitative researchers are often forced to control every problem as if it were the same: something in need of applying S&P.

The Inference Revolution

The emerging social sciences gave little attention to the internal debates in statistics. Instead, there was a desire to standardize research practices to appear scientific and objective (Gigerenzer et al. 1989), for which they began adopting the new statistics with an interest in producing a single way to make scientific inferences that could be deployed irrespective of the context or research problem (see Abbott 1998; Gigerenzer 2004; Krüger et al. 1987). Key parts of this process involved ways of working with what were viewed as unobservable, 'underlying,' or 'latent' aspects of reality. For this, tools from twentieth century philosophy of science were also used, including descriptive practices designed for relativistic and quantum physical entities that could not be observed and therefore required logical stand-ins to work with representational theories of meaning and correspondence theories of truth. Bertrand Russell's 'constructs' were designed for this purpose and were passed from logical positivists such as Feigl and Carnap to researchers such as Cronbach and Meehl, who formalized various practices of justification loosely called 'construct validity' (Hacking 1999, p. 44).

With the ability to treat an 'underlying' or 'latent' reality of non-objects, a statistical 'inference revolution' created a hybrid of the approaches of Fisher, Neyman, and E. Pearson that today is called null hypothesis significance testing (Gigerenzer et al. 1989; Gigerenzer and Murray 1987). The hybrid took Fisher's logic and combined it with fears

of Type-I/II errors. Interestingly, Neyman and E. Pearson's practice of accepting hypotheses with an applied focus on behavior did not take hold, perhaps due to the influence of philosophers such as Popper who emphasized falsification (e.g., 1959/2002), but also because Fisher's approach was focused on typical forms of inductive inference for building scientific knowledge—applicable to the interests of social inquiry meant to be scientific.

The adoption of this hybrid was most rapid in psychology. As noted by Rucci and Tweney (1980) and Sterling (1959), between 1934 and 1940 there were only 17 papers using the hybrid, but by 1955 a full 80% of published work reported the technique and editors of journals began using *p* values to measure *p*ublishability (e.g., Melton 1962, p. 553). Against notable critiques (for discussion see Gigerenzer 1998; Gigerenzer et al. 1989), the hybrid spread alongside ethical approaches to research and a new logic of standardization for doing good social science: 'best practices' and 'rigor' that ignored context and practical problems in the name of true scores and true causal effects to generalize from samples to populations—so that social science could seem more scientific and 'objective.' In turn, it became possible for psychology to claim that its methods were more 'rigorous,' based on a set of ethical commitments associated with adherence to the logic and practices of S&P. In combination with the use of 'constructs' as if these were natural features of the world—a rather obvious irony—researchers now had S&P tools that they would claim were value-free or 'objective.'

Rigor in Management Research

The context of adopting S&P in management research has been well studied (e.g., Augier and March 2011; Cooke and Alcadipani 2015; Khurana 2007; Schlossman and Sedlak 1985; Schlossman et al. 1987; Tadajewski 2009; Weatherbee 2012). This research acknowledges influential reports by the Ford Foundation (Gordon and Howell 1959) and the Carnegie Corporation (Pierson 1959), noting problems with business schools before 1960. We extend this to the ethics of S&P by describing how the inference revolution provided the grounds for treating research ethics as an afterthought of the logic of S&P.

With statistics as a discipline and the hybrid in place, a crisis emerged that used S&P to shape the methods *and* content of management research. The crisis was a set of mutually reinforcing conceptions of what management was, what management education should be, what its researchers should study, the methods they should use, and how they should use them—with implications for how researchers ethically constituted themselves. As in early foundations for S&P (e.g., Arnauld and Nicole 1662/1996; Laplace 1774/1986; 1825/1955), these conceptions constituted managers *and* researchers as ethically obligated to use:

> powerful analytical tools which would contribute to more rational decision-making… [T]here is need for… a better understanding of the interrelationships among the variables with which the business manager must deal … Whether the aim is to improve our understanding of business behavior… or to develop better techniques and rules for decision-making, it is clear that business research needs to… utilize a more sophisticated methodology (Gordon and Howell 1959, pp. 384–385).

This was only possible because S&P had emerged as a style of reasoning—a self-vindicating way to make decisions whether for management itself or for research about management.

Quantitative training was needed: "The general direction which the doctoral program in the best schools will take in the years ahead is beginning to be clear… [T]he behavioral sciences will be stressed…; so will training in quantitative methods," creating a "need of the doctoral candidate for sophisticated, research-oriented training in statistics" (Gordon and Howell 1959, pp. 407–409). The idea was that programs needed "more advanced doctoral sequences in quantitative methods, the behavioral sciences (built especially from social and individual psychology… including quantitative method)" (Pierson 1959, pp. 347–349).

William Starbuck experienced this first hand (personal communication, March 6th, 2014). Those promoting statistical tools "really believed that they wanted to revolutionize American business education… Mathematical modeling and lab experiments and getting survey data and doing analyses of the data—those were all things that to them were one big conglomerate… [S]cience became the religion." By constituting management and research with the new logic of S&P and the hybrid, the stage was set for quantitative methods to take their present-day form, informed especially by psychology.

The new *Journal of the Academy of Management* illustrates this (e.g., McFarland 1960). In 1960 the AOM president noted, "[t]he challenge ahead is to improve the standards of our academic work. Fortunately, we have the Ford and Carnegie Reports to assist us in reappraising our performance and in indicating those activities which need our attention," (Towle 1960, p. 150). From 1961: "If management is to have a claim to continue as an independent discipline on the university level, it must be understood to have developed out of or rest on other such disciplines and it must itself be research-oriented on the basic assumption that it is itself capable of cumulative accretion" (Bornemann 1961, p. 136). Only a year later it was noted that this accretion began with 'objectivity': "the growth of controlled objective

research in administrative and organizational behavior is a tide that will not be stemmed… it seems inevitable that there can be only one direction of change—towards the researcher" (Shull 1962, p. 125). Only one year later, Miner (1963, p. 138) notes the effects of this tide: "it is apparently [in management] that the influx of psychologists has been greatest" (See also Bernthal 1960; Halff 1960; Moore 1960). Their influence is still felt at journals like *Organization Research Methods*, which has editors with psychological pedigree and requires that submissions follow guidelines in the *Publication Manual of the American Psychological Association*, replete with recommendations that use S&P with the hybrid.

Interestingly, not a single article in what is now *AMJ* used today's 'rigor' until the Ford and Carnegie reports were published and retraining had commenced. Golembiewski (1961) first mentions analysis of variance and statistical significance while discussing psychological research, followed by House (1962) who published the first paper with a modern style of S&P, complete with hypotheses, a results section, and *p* values used to 'confirm' hypotheses. Perhaps ironically, it was only 13 years later that he also published an article that described emerging tensions between rigor and relevance (see House 1975). Despite this tension, rigor as S&P emerged as a source of legitimacy for making particular claims based on specific research practices, often divorced from the context of specific situations and problems (e.g., Aguinis et al. 2010; Edwards 2010). This would also be true of rigor's 'best practices,' which often serve to inhibit contextualization. Yet, as rigor came to dominate quantitative work, relevance and contextualization would be casualties of a focus on law-like 'causes' and statistical aggregates. Also ignored would be the history of values and ethics in S&P, allowing it to appear timeless as it colonized journals—by 2012, *AMJ* had an average of 99 *p* values in *each* article (Gigerenzer and Marewski 2015).

In sum, after the 1960s it is no surprise that S&P came to dominate quantitative research—in organization science and its core disciplines, especially psychology. As our historical ontology shows, the ethical basis for working with S&P is the same imputed reality that emerges from its 'best practices.' Differences between objects of study are taken to be the result of natural laws and causes, which can be represented by a particular kind of ethical agent who uses statistics with hypothesis tests that involve Type-I/II errors and an ethics-laden uncertainty that is probabilistic, alongside 'constructs' that are also taken to be natural rather than socially produced. This promotes myopic understandings of research ethics while at the same time demotivating free thinking, criticism, and scholarly inquiry into this strange state of affairs.

Instead of inquiry, we find now-tired forms of uncertainty modeled on games of chance, with critiques of research

practices based on concerns regarding what a researcher *ought* to do in order to estimate true scores and/or true causal effects while probabilistically avoiding errors in inference. It should come as no surprise, then, that ethics is given expression within a theorized individual knower—the researcher—who can be blamed for making good or bad choices between 'best' or 'questionable' research practices (e.g., Honig et al. 2017). Yet, our historical ontology shows that the logical foundations of both kinds of practices are the same: a natural quantitative reality made in the image of S&P, which offers researchers little choice in what is 'rigorous' or a 'best practice.' These are all of course value-laden and matters of ethics, whether or not researchers are willing to contend with their historical production.

Conclusion

Our historical ontology offers a new understanding of S&P, including notions of rigor and its best practices, by showing how quantitative researchers came to enact a priority, or a hierarchy, in which the logic of S&P comes first and then explicit ethical concerns second. Yet, what this *logic-then-ethics* relationship tends to miss is that practices of S&P are contingent. Notions of 'rigor' and 'best practices' are products of ethical problems, including values, institutional arrangements, and socio-material conditions of different eras, many of which derive from concerns about a deity's laws or 'causes.' In contrast, by focusing on *ethics-and-logic*, the history of S&P becomes important for at least two reasons that we now discuss. The first reason relates to research methods, whereas the second relates to using historical analysis for understanding ethics in a wider variety of research and social issues.

First, should S&P and its logic be used to produce and regulate research in the form of 'rigor' or 'best practices'? Although trust in research reporting practices is important for any community of researchers, in light of our historical ontology we must critically evaluate the endorsement of 'best practices' *when uncertainty is treated only on the terms of S&P*. One practical response might be to overturn best practices through exploratory or abductive research (for example by following Schwab and Starbuck 2017). However, our analysis leads to a much wider-ranging consideration of how ethics and quantitative realities are entangled, because breaking from existing best practices will most likely lead to new best practices that have not disrupted the underlying priority of logic over ethics which we identified previously.

Existing features of S&P enable the production of knowledge, but only by imposing limits on how logic and ethics are constructed, and then, construed as separate. Building on our historical ontology, these limits can be reconsidered by recognizing how S&P demand these imperatives:

(1) representational theories of meaning that are combined with correspondence theories of truth; (2) constituting the world as a kind of thing that is statistical and arrives as a bundle of quantitative facts; (3) taking what is probable as ethically virtuous and assuming that a probability calculus derived from games of chance is relevant for addressing social problems and uncertainty in research; and, 4) treating populations (and 'constructs') as natural kinds and treating anything observed as a sample (or 'measurement').

None of these imperatives are essential for science. Indeed, they can be problematic by limiting the ways that researchers understand ethics and the practical aspects of handling 'reality' in the process of research. Although we leave many implications of this insight for future work, it is important to point out how any notion of reality is itself value-laden (see also Ezzamel and Willmott 2014; Wicks and Freeman 1998). As our analysis shows, the 'real' and the 'right' often derive from similar locations by answering to the concerns, concepts, and technologies available to people from specific locations and eras. Whether it is organizational researchers attempting to produce knowledge on terms that can be understood in their local system of 'peer review,' or it is scholars from centuries past who needed to address a deity and the church, any notion of reality is informed by the present logical *and* ethical conditions of those who enact it.

It is therefore warranted to question whether uncritical and decontextualized uses of S&P offer substantive or evaluative tools, either for adequately constructing and testing representations, or for adequately constructing an ethics of research. As an alternative to the dubious *logic-then-ethics* relationship, researchers can stop treating ethics as afterthought and approach ethics with a seriousness that we think it deserves. Though there are many ways to do this (e.g., following Ezzamel and Willmott 2014; Martela 2015; American Psychiatric Association 2013; Wicks and Freeman 1998; Zyphur and Pierides 2017), we are upfront about our commitment to the classical pragmatist John Dewey's understanding of inquiry.

The pragmatist approach to quantitative research that we would encourage is unapologetic about its empirical specificity and contextual situatedness. It should be pursued via Deweyan inquiry (esp. Dewey 1938), by taking a particular situation or problem as its starting point. To be clear, our present paper does not constitute a starting point for research, but it does provide an historically informed rationale for dismissing a universalistic mandate that quantitative 'rigor' and its 'best practices' should necessarily be the starting points, the means or the ends for quantitative research—if not to motivate more scholarly and scientifically informed approaches, then at least to avoid a kind of *logic-then-ethics* priority.

If researchers were to understand and do quantitative research this way, problems associated with rigor-relevance tensions and a lack of reflexivity could be better addressed, because rigor would no longer stand in the way of being relevant *and* ethical (for concerns, see Hardy et al. 2001; Rynes et al. 2001). History conditions the nature of research, the reality towards which it is addressed, and the researchers who enact it (see Casler and du Gay 2019). In this configuration, inquiry becomes a continuous process that also constructs researchers whose actions are consequential (Dewey 1938). Although we do not have space to fully sketch an overview, future work can use our historical analysis to better understand how to refashion 'rigor' and research so that it better speaks to contemporary problems, with an understanding of how deities and institutions of the past shape what is purported to be 'natural' or 'objective.' Specifically, we think it is time for a fundamentally different theoretical approach for doing and evaluating S&P, one that is based in pragmatism.

The second reason our analysis is important is that it shows how S&P are not timeless or ahistorical, while simultaneously serving as an example of how to treat concepts or ontologies that obscure their value-laden nature by appearing to have no history. Our analysis helps explain *how* human practices have produced an ethical spectrum of practices—which now appear to have 'best' on the one end, and 'questionable' on the other—by grounding quantitative research methods in a logic of S&P at the expense of ethics-oriented inquiry. Other historical inquiries make related points (e.g., Kuhn 2012), but our analysis specifically shows how science and research involve *ways of reasoning* or *ways of working among researchers* that are at once value-laden and constituted by historical conditions.

This scientific work is about the ways one constitutes oneself as a researcher through research practices that involve handling objects such as 'populations,' 'samples,' 'causes,' 'laws,' 'errors,' 'true scores,' 'unbiased effects,' 'data,' 'reality,' 'truth,' 'rigorous research,' and anything else that takes part in S&P. Researchers can use our historical ontology as a source and an inspiration for further inquiry into how researchers are constituted as specific kinds of people through normative and technical training which involves the formation of these objects (again, see Casler and du Gay 2019). The focus of such historical analysis makes it possible to break down the fact-value or epistemology-ethics distinctions that make ethical analyses of scientific practices inaccessible for some researchers. We hope that our paper, in addition to our other work, motivates historically informed analysis of this kind.

To this end, our paper offers an entry point for researchers, especially quantitative researchers, who want to understand and address 'post truth' or 'alternative facts' problems. When criticisms of truth, facts, or scientific research are made in bad faith, one problem for science is that attempting to battle these agendas by offering more facts or 'greater'

truths can be dysfunctional because the problems involved are not only about conceptions of objectivity and adherence to them. As we have shown, the very concept of a fact or a notion of truth is constituted by the ways that social groups value and arrange their own practices, discourses, and objects/subjects of study. Thus, confronting efforts to undermine scientific facts or truth should be done by collective efforts to organize around the values that bind people together in the name of creating better futures, whether through quantitative methods or other means.

Indeed, this was always what facts and notions of truth were meant to enable through science. The making of scientific facts was not, and is not, about having blind faith in objectivity or the existence of a reality that is merely 'out there waiting for us to study it.' Facts have always been and continue to be the products of hard work that comes about from a pragmatic effort to actively generate the tools, conceptual or otherwise, for making a better tomorrow. Some researchers cling to existing research methods based on a faulty assumption that doing otherwise may invite 'anything goes' approaches to conceptualizing reality (e.g., Cortina 2019; Edwards 2019). In our view, it is better to understand the value-laden nature of quantitative methods and the facts they generate in order to defend them based on our collective concerns—this confronts the problem on its own terms—rather than using outdated narratives about, for example, objectivity. Indeed, it is partly the pluralistic nature of description that makes defending a collectively valuable version of reality so important.
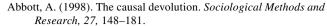
Compliance with Ethical Standards

Conflict of interest The authors declare no conflicts of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abbott, A. (1991). The order of professionalization: An empirical analysis. *Work and Occupations, 18,* 355–384.

Abbott, A. (1998). The causal devolution. *Sociological Methods and Research, 27,* 148–181.

Adams, W. J. (2009). *The life and times of the central limit theorem*. Providence, RI: American Mathematical Society.

Aguinis, H., Werner, S., Abbott, J. L., Angert, C., Park, J. H., & Kohlhausen, D. (2010). Customer-centric science: Reporting significant research results with rigor, relevance, and practical impact in mind. *Organizational Research Methods, 13,* 515–539.

Alanen, L. (2003). *Descartes's concept of mind*. Cambridge, MA: Harvard University Press.

Aldrich, J. (1997). Fisher and the making of maximum likelihood 1912-1922. *Statistical Science, 12,* 133–220.

Aldrich, J. (2008). R. A. Fisher on Bayes and Bayes' theorem. *Bayesian Analysis, 3,* 161–170.

Alonso, W., & Starr, P. (Eds.). (1987). *The politics of large numbers*. New York: Russell Sage.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Ariew, A. (2007). Under the influence of Malthus's law of population growth: Darwin eschews the statistical techniques of Adolphe Quetelet. *Studies in the History and Philosophy of Biology and Biomedicine, 38,* 1–19.

Arnauld, A., & Nicole, P. (1662/1996). *Logic or the art of thinking*. Cambridge, UK: Cambridge University Press.

Augier, M., & March, J. G. (2011). *The roots, rituals, and rhetorics of change: North American business schools after the second world war*. Stanford, CA: Stanford University Press.

Bayes, T. P. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London, 53,* 370–418.

Bernoulli, J. (1713/2006). *The art of conjecturing (together with letter to a friend on sets in court tennis)* (E. D. Sylla, Trans. & Ed.). Baltimore, MD: Johns Hopkins University Press.

Bernthal, W. F. (1960). Integrating the behavioral sciences and management. *The Journal of the Academy of Management, 3,* 161–166.

Biagioli, M. (2002). From book censorship to academic peer review. *Emergences: Journal for the Study of Media and Composite Cultures, 12,* 11–45.

Biagioli, M. (2006). *Galileo's instruments of credit: Telescopes, images, secrecy*. Chicago, IL: University of Chicago Press.

Bornemann, A. (1961). The development of the teaching of management in the school of business. *The Journal of the Academy of Management, 4,* 129–136.

Bromhead, H. (2009). *The reign of truth and faith: Epistemic expressions in the 16th and 17th century English*. Berlin: Mouton de Gruyter.

Byrne, E. F. (1968). *Probability and opinion: A study of the medieval presuppositions of post-medieval theories of probability*. The Hague: Matinus Nijhoff.

Casler, C., & du Gay, P. (2019). Stances, paradigms, personae. *Studi di Sociologia, 1,* 69–80.

Caton, H. (1973). *The origin of subjectivity*. New Haven, CT: Yale University Press.

Cohen, I. B. (1987). Scientific revolutions, revolutions in science, and a probabilistic revolution 1800-1930. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 23–44). Cambridge, MA: MIT Press.

Cooke, B., & Alcadipani, R. (2015). Toward a global history of management education: The case of the Ford Foundation and the São Paulo School of Business Administration, Brazil. *Academy of Management Learning & Education, 14,* 482–499.

Cortina, J. (2019). On the whys and hows of quantitative research. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04195-8.

Dale, A. I. (1999). *A history of inverse probability: From Thomas Bayes to Karl Pearson* (2nd ed.). New York: Springer.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge, UK: Cambridge University Press.

Daston, L. (1987a). The domestication of risk: Mathematical probability and insurance 1650-1830. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 238–260). Cambridge, MA: MIT Press.

Daston, L. (1987b). Rational individuals versus laws of society: From probability to statistics. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 295–304). Cambridge, MA: MIT Press.

Daston, L. (1992). Objectivity and the escape from perspective. *Social Studies of Science, 22,* 597–618.

Daston, L. (1994). How probability came to be objective and subjective. *Historia Mathematica, 21,* 330–344.

Daston, L. (1995). *Classical probability in the enlightenment*. Princeton, NJ: Princeton University Press.

Daston, L. (2005). Scientific error and the ethos of belief. *Social Research, 72,* 1–28.

Daston, L., & Galison, P. (2007). *Objectivity*. Brooklyn, NY: Zone Books.

Daston, L., & Vidal, F. (Eds.). (2004). *The moral authority of nature*. Chicago, IL: University of Chicago Press.

De Morgan, A. (1838/2010). *An essay on probabilities: And their application to life contingencies and insurance offices*. Charleston, SC: BiblioLife, LLC.

Dear, P. (2001). *Revolutionizing the sciences: European knowledge and its ambitions, 1500-1700*. London: Palgrave.

Desrosières, A. (1998). *The politics of large numbers: A history of statistical reasoning*, C. Naish (trans.). Cambridge, UK: Harvard University Press.

Dewey, J. (1929). *The quest for certainty*. New York: Minton, Balch, & Co.

Dewey, J. (1938). *Logic: The theory of inquiry*. New York: Henry Holt and Company.

Douglas, H. E. (2009). *Science, policy, and the value-free ideal*. Pittsburgh, PA: University of Pittsburgh Press.

Duncan, O. D. (1984). *Notes on social measurement: Historical and critical*. New York: Russell Sage.

Edwards, J. R. (2010). Reconsidering theoretical progress in organizational and management research. *Organizational Research Methods, 13,* 615–619.

Edwards, J. R. (2019). The peaceful coexistence of ethics and quantitative research. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04197-6.

Ezzamel, M., & Willmott, H. (2014). Registering 'the ethical' in organization theory formation: Towards the disclosure of an 'invisible force'. *Organization Studies, 35,* 1–27.

Fischer, H. (2010). *A history of the central limit theorem: From classical to modern probability theory*. New York: Springer.

Fisher, R. A. (1915). The frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika, 10,* 507–521.

Fisher, R. A. (1921). On the "probable error" of a correlation coefficient deduced from a small sample. *Metron, 1,* 3–32.

Fisher, R. A. (1922a). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, A, 222,* 309–368.

Fisher, R. A. (1922b). The goodness of fit of regression formulae and the distribution of regression coefficients. *Journal of the Royal Statistical Society, 85,* 597–612.

Fisher, R. A. (1924). The distribution of the partial correlation coefficient. *Metron, 3,* 329–332.

Fisher, R. A. (1925a). Applications of "Student's" distribution. *Metron, 5,* 90–104.

Fisher, R. A. (1925b). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd Ltd.

Fisher, R. A. (1925c). Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society, 22,* 700–725.

Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society, A, 121,* 654–673.

Fisher, R. A. (1929). Moments and product moment of sampling distributions. *Proceedings of the London Mathematical Society, Series, 2*(30), 199–238.

Fisher, R. A. (1935a). *The design of experiments*. Edinburgh: Oliver & Boyd Ltd.

Fisher, R. A. (1935b). The logic of inductive inference. *Journal of the Royal Statistical Society, 98,* 39–54.

Fisher, R. A. (1955). Statistical methods and scientific induction. *Journal of the Royal Statistical Society: Series B, 17,* 69–78.

Fisher, R. A. (1956). *Statistical methods and scientific inference*. Edinburgh: Oliver & Boyd.

Fisher, R. A., & Yates, F. (1943). *Statistical tables for biological, agricultural and medical research*. Edinburgh, UK: Oliver & Boyd Ltd.

Foucault, M. (1970). *The order of things: An archaeology of the human sciences*. New York: Random House.

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings*. New York: Random House.

Foucault, M. (2003). *Abnormal: Lectures at the college de France 1974-1975*. New York: Graham Burchell.

Foucault, M. (2008). *The birth of biopolitics Lectures at the college de France 1978-1979*. New York: Graham Burchell.

Freedman, D. A. (2005). Linear statistical models for causation: A critical review. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 2, pp. 1061–1073). Chichester, UK: Wiley.

Galavotti, M. C. (2005). *Philosophical introduction to probability*. Palo Alto, CA: CSLI.

Galileo, G. (2008). *The essential Galileo* (M. A. Finocchiaro, Ed.). Indianapolis, IN: Hackett Publishing.

Galton, G. (1869/2010). *Hereditary genius: An inquiry into its laws and consequences.* Memphis, TN: General Books.

Galton, F. (1889/2010). *Natural inheritance*. Charleston, SC: Nabu Press.

Garber, D. (1992). Descartes' physics. In J. Cottingham (Ed.), *The Cambridge companion to Descartes*. Cambridge, UK: Cambridge University Press.

Gauss, C. F. (1809/2010). *Theory of the motion of the heavenly bodies moving about the sun in conic sections* (C. H. Davis, Trans.). Charleston, SC: Nabu Press.

Gigerenzer, G. (1998). We need statistical thinking, not statistical rules. *Behavioral and Brain Sciences, 21,* 199–200.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33,* 587–606.

Gigerenzer, G., & Marewski, J. N. (2015). Surrogate science: The idol of a universal method for scientific inference. *Journal of Management, 41,* 421–440.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, MI: Erlbaum.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance. How probability changed science and everyday life*. Cambridge, UK: Cambridge University Press.

Goldman, L. (1983). The origins of British 'social science': Political economy, natural science and statistics, 1830-1835. *The Historical Journal, 26,* 587–616.

Golembiewski, R. T. (1961). Management science and group behavior: Work-unit cohesiveness. *Journal of the Academy of Management, 4,* 87–99.

Gordon, R. A., & Howell, J. E. (1959). *Higher education for business*. New York: Columbia University Press.

Greenwood, M. (2016). Approving or improving research ethics in management journals. *Journal of Business Ethics, 137,* 1–14.

Hacking, I. (1965). *Logic of statistical inference*. Cambridge, UK: Cambridge University Press.

Hacking, I. (1975). *Why does language matter to philosophy?*. Cambridge, UK: Cambridge University Press.

Hacking, I. (1978). Hume's species of probability. *Philosophical Studies, 33,* 21–37.

Hacking, I. (1986). The archaeology of Foucault. In D. C. Hoy (Ed.), *Foucault: A critical reader* (pp. 27–40). Oxford, UK: Blackwell.

Hacking, I. (1987). Was there a probabilistic revolution 1800-1930? In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 45–55). Cambridge, MA: MIT Press.

Hacking, I. (1990). *The taming of chance*. Cambridge, UK: Cambridge University Press.

Hacking, I. (1991). How should we do the history of statistics? In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect* (pp. 181–195). Chicago, IL: Chicago University Press.

Hacking, I. (1992). Statistical language, statistical truth and statistical reason: The self-authentification of a style of scientific reasoning. In E. McMullin (Ed.), *The social dimensions of science* (Vol. 3, pp. 130–157). Notre Dame, IN: University of Notre Dame Press.

Hacking, I. (1999). *The social construction of what?*. Cambridge, UK: Harvard University Press.

Hacking, I. (2002). *Historical ontology*. Cambridge, UK: Harvard University Press.

Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference* (2nd ed.). Cambridge, UK: Cambridge University Press.

Hacking, I. (2009). *Scientific reasoning*. Taipei: NTU Press.

Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fisher, 1713-1935*. New York: Springer.

Halff, J. F. (1960). Applying the scientific method to the study of management. *The Journal of the Academy of Management, 3,* 193–196.

Hardy, C., Phillips, N., & Clegg, S. R. (2001). Reflexivity in organization and management theory: A study of the production of the research 'subject'. *Human Relations, 54,* 531–560.

Hennig, B. (2007). Cartesian conscientia. *British Journal for the History of Philosophy, 15,* 455–484.

Henry, J. (2004). Metaphysics and the origins of modern science: Descartes and the importance of laws of nature. *Early Science and Medicine, 9,* 73–114.

Honig, B., Lampel, J., Siegel, D., & Drnevich, P. (2017). Special section on ethics in management research: Norms, identity, and community in the 21st century. *Academy of Management Learning & Education, 16,* 84–93.

Hopwood, N., Schaffer, S., & Secord, J. (2010). Seriality and scientific objects in the nineteenth century. *History of Science, 48,* 251–284.

House, R. J. (1962). An experiment in the use of management training standards. *Journal of the Academy of Management, 5,* 76–81.

House, R. J. (1975). The quest for relevance in management education: Some second thoughts and undesired consequences. *Academy of Management Journal, 18,* 323–333.

Howie, D. (2002). *Interpreting probability: Controversies and developments in the early twentieth century*. Cambridge, UK: Cambridge University Press.

Hume, D. (1739). *A treatise of human nature: Being an attempt to introduce the experimental method of reasoning into moral subjects* (Vol. 1). London: Jon Noon.

Huygens, C. (1657/2010). *De ratiociniis in aleae ludo. or, the value of all chances in games of fortune; … mathematically demonstrated*. Farmington Hills, MI: Gale ECCO.

Kamlah, A. (1987). The decline of the Laplacian theory of probability: A study of Stupf, von Kries, and Meinong. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 91–116). Cambridge, MA: MIT Press.

Khurana, R. (2007). *From hired aims to hired hands: The social transformation of American business schools and the unfulfilled promise of management as a profession*. Princeton, NJ: Princeton University Press.

Köhler, T., Landis, R. S., & Cortina, J. M. (2017). From the editors: Establishing methodological rigor in quantitative management learning and education research: The role of design, statistical methods, and reporting standards. *Academy of Management Learning & Education, 16,* 173–192.

Krüger, L. (1987). The slow rise of probabilism: Philosophical arguments in the nineteenth century. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 59–89). Cambridge, MA: MIT Press.

Krüger, L., Gigerenzer, G., & Morgan, M. S. (Eds.). (1987). *The probabilistic revolution: Ideas in the sciences* (Vol. 2). Cambridge, MA: MIT Press.

Kuhn, T. S. (2012). *The structure of scientific revolutions* (5th ed.). Chicago, IL: University of Chicago Press.

Lagrange, J. L. (1770–1773/2009). *Memoir on the utility of the method of taking the mean among the results of several observations in which one examines the advantage of this method by the calculus of probabilities, and where on selves different problems related to this material* (R. J. Pulskamp, Trans.) http://cerebro.xu.edu/math/Sources/Lagrange/on%20mean.pdf.

Laplace, P.-S. (1774/1986). Memoir on the probability of the cause of events (S. Stigler, Trans.). *Statistical Science*, *1*, 364–378.

Laplace, P.-S. (1809a/2011). *Mémoire sur les approximations des formules qui sont fonctions de très-grands nombres, et sur leur application aux probabilités* (R. J. Pulskamp, Trans.). http://www.cs.xu.edu/math/Sources/Laplace/approximations%20of%20formulas%201809.pdf.

Laplace, P.-S. (1809b/2010). *Supplement au mémoire sur les approximations des formules qui sont fonctions de très-grands nombres* (R. J. Pulskamp, Trans.). http://www.cs.xu.edu/math/Sources/Laplace/supplement%201809.pdf.

Laplace, P.-S. (1825/1995). *Philosophical essay on probability* (A. I. Dale, Trans.). New York: Springer.

Lécuyer, B.-P. (1987). Probability in vital and social statistics: Quetelet, Farr, and the Bertillons. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 317–335). Cambridge, MA: MIT Press.

Leibniz, G. (1678/2004). Leibniz on estimating the uncertain: An English translation of De incerti aestimatione with commentary (W. D. C. de Leo and J. Cussens, Trans.). *The Leibniz Review, 14*, 31–53.

MacKenzie, D. (1981). *Statistics in Britain, 1865-1930: The social construction of scientific knowledge*. Edinburgh, UK: Edinburgh University Press.

Mandelbaum, M. (1964). *Philosophy, science, and sense perception*. Baltimore, MD: Johns Hopkins Press.

Martela, F. (2015). Fallible inquiry with ethical ends-in-view: A pragmatist philosophy of science for organizational research. *Organization Studies, 36,* 537–563.

McFarland, D. E. (1960). The emerging revolution in management education. *The Journal of the Academy of Management, 3,* 7–15.

McGrayne, S. (2011). *The theory that would not die*. New Haven, CT: Yale University Press.

Melton, A. W. (1962). Editorial. *Journal of Experimental Psychology, 64,* 553–557.

Metz, K. H. (1987). Paupers and numbers: The statistical argument for social reform in Britain during the period of industrialization. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 337–350). Cambridge, MA: MIT Press.

Miner, J. B. (1963). Psychology and the business school curriculum. *Academy of Management Journal, 46,* 284–289.

Moore, D. G. (1960). Behavioral science and business education. *The Journal of the Academy of Management, 3,* 187–191.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified random sampling and the method of purposive selection (with discussion). *Journal of the Royal Statistical Society, B, 97,* 558–606.

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London, A, 236,* 333–380.

Neyman, J. (1950). *First course in probability and statistics*. New York: Henry Holt.

Neyman, J. (1956). Note on an article by Sir Ronald Fisher. *Journal of the Royal Statistical Society: Series B, 18,* 288–294.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society, A, 231,* 289–337.

Osler, M. J. (1970). John Locke and the changing ideal of scientific knowledge. *Journal of the History of Ideas, 31,* 3–16.

Panter, A. T., & Sterba, S. K. (Eds.). (2011). *Handbook of ethics in quantitative methodology*. New York: Taylor & Francis.

Pascal, B. (1653/1952). Treatise on the arithmetical triangle. In R. M. Hutchins (Ed.), *Great books of the Western world: Pascal* (Vol. 33, pp. 447–473). Chicago, IL: University of Chicago Press.

Pascal, B. (1654/1952). Correspondence with Fermat on the theory of probabilities. In R. M. Hutchins (Ed.), *Great books of the Western world: Pascal* (Vol. 33, pp. 474–487). Chicago, IL: University of Chicago Press.

Pearson, E. S. (1978). *The history of statistics in the 17th & 18th centuries*. London: Charles Griffin & Co.

Petit, V. (2013). An object called population. In V. Petit (Ed.), *Counting populations, understanding societies* (pp. 53–87). Dordrecht: Springer.

Pierson, F. C. (1959). *The education of American businessmen: A study of university-college programmes in business administration*. New York: McGraw-Hill.

Poovey, M. (1995). *Making a social body: British cultural formation, 1830-1864*. Chicago, IL: Chicago University Press.

Poovey, M. (1998). *A history of the modern fact: Problems of knowledge in the sciences of wealth and society*. Chicago, IL: Chicago University Press.

Popper, K. (1959/2002). *The logic of scientific discovery*. London: Routledge.

Porter, T. M. (1986). *The rise of statistical thinking, 1820-1900*. Princeton, NJ: Princeton University Press.

Porter, T. M. (1994). From Quetelet to Maxwell: Social statistics and the original of statistical physics. In I. B. Cohen (Ed.), *The natural sciences and the social sciences* (Vol. 150, pp. 345–362). New York: Springer.

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Powell, T. (2019). Can quantitative research solve social problems? Pragmatism and the ethics of social research. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-019-04196-7

Quetelet, L. A. J. (1842/2010). *A treatise on man and the development of his faculties* (R. Knox, Trans., T. Smibert, Ed.). Charleston, SC: Nabu Press.

Rorty, R. (1979). *Philosophy and the mirror of nature*. Princeton, NJ: Princeton University Press.

Ross, D. (1991). *The origins of American social science*. Cambridge, UK: Cambridge University Press.

Rowlinson, M., Hassard, J., & Decker, S. (2014). Research strategies for organizational history: A dialogue between historical theory and organization theory. *Academy of Management Review, 39,* 250–274.

Rucci, A. J., & Tweney, R. D. (1980). Analysis of variance and the "second discipline" of scientific psychology: A historical account. *Psychological Bulletin, 87,* 166–184.

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,* 340–355.

Schlossman, S., & Sedlak, M. (1985). The age of autonomy in American management education. *Selections, 1,* 15–26.

Schlossman, S. L., Sedlak, M. W., & Wechsler, H. S. (1987). *The 'new look': The Ford Foundation and the revolution in business education*. Los Angeles, CA: Graduate Management Admissions Council.

Schneider, I. (1987). Laplace and thereafter: The status of probability calculus in the nineteenth century. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 191–214). Cambridge, MA: MIT Press.

Schouls, P. A. (1979). *The imposition of method: A study of Descartes and Locke*. Oxford, UK: Oxford University Press.

Schouls, P. A. (2000). *Descartes and the possibility of science*. Ithaca, NY: Cornell University Press.

Schwab, A., & Starbuck, W. H. (2017). A call for openness in research reporting: How to turn covert practices into helpful tools. *Academy of Management Learning & Education, 16*(1), 125–141.

Shapin, S. (1984). Pump and circumstance: Robert Boyle's literary technology. *Social Studies of Science, 14,* 481–520.

Shapin, S. (1994). *A social history of truth: Civility and science in seventeenth-century England*. Chicago, IL: University of Chicago Press.

Shapin, S. (2008). *The scientific life: A moral history of a late modern vocation*. Chicago, IL: University of Chicago Press.

Shapin, S. (2010). *Never pure: Historical studies of science as if it was produced by people with bodies, situated in time, space, culture, and society, and struggling for credibility and authority*. Baltimore, MD: Johns Hopkins University Press.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton, NJ: Princeton University Press.

Shapiro, B. J. (1983). *Probability and certainty in seventeenth-century England: A study of the relationships between natural science, religion, history, law, and literature*. Princeton, NJ: Princeton University Press.

Shull, F. (1962). The nature and contribution of administrative models and organizational research. *The Journal of the Academy of Management, 5,* 124–138.

Sterling, R. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—Or vice versa. *Journal of the American Statistical Association, 54,* 30–34.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, UK: Harvard University Press.

Stigler, S. M. (1999). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, UK: Harvard University Press.

Strong, J. V. (1978). John Stuart Mill, John Herschel, and the 'probability of causes'. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1,* 31–41.

Swijtink, Z. G. (1987). The objectification of observation: Measurement and statistical methods in the nineteenth century. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 261–292). Cambridge, MA: MIT Press.

Tadajewski, M. (2009). The politics of the behavioral revolution. *Organization, 16,* 733–754.

Towle, J. W. (1960). Opportunities ahead for the Academy of Management. *The Journal of the Academy of Management, 3,* 147–154.

Truesdell, C. (1984). *An idiot's fugitive essays on science: Methods, criticism, training, circumstances.* New York: Springer.

Venn, J. (1866/2006). *The logic of chance*. Mineola: Dover Publications.

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics, 16,* 117–186.

Wald, A. (1950). *Statistical decision functions.* New York: Wiley.

Weatherbee, T. G. (2012). Caution! This historiography makes wide turns: Historic turns and breaks in management and organization studies. *Management & Organizational History, 7,* 203–218.

Wicks, A. C., & Freeman, R. E. (1998). Organizational studies and the new pragmatism: Positivism, anti-positivism, and the search for ethics. *Organization Science, 9,* 123–140.

Williams, B. (2005). *Descartes: The project of pure inquiry*. New York: Routledge.

Wise, M. N. (1987). How do sums count? One the cultural origins of statistical causality. In L. Krüger, L. J. Daston, & M. Heidelberger (Eds.), *The probabilistic revolution* (Vol. 1, pp. 395–425). Cambridge, MA: MIT Press.

Woolf, S. (1989). Statistics and the modern state. *Comparative Studies in Society and History, 31,* 588–604.

Yule, G. U. (1897). One the theory of correlation. *Journal of the Royal Statistical Society, 60,* 812–854.

Yule, G. U. (1899). An investigation into the causes of changes in pauperism in England, chiefly during the last two intercensal decades, *I. Journal of the Royal Statistical Society, 62,* 249–295.

Zabell, S. (1989). R. A. Fisher on the history of inverse probability. *Statistical Science, 4,* 247–263.

Zabell, S. (1992). R. A. Fisher and fiducial argument. *Statistical Science, 7,* 369–387.

Zyphur, M. J., & Pierides, D. C. (2017). Is quantitative research ethical? Tools for ethically practicing, evaluating, and using quantitative research. *Journal of Business Ethics, 143,* 1–16.

Zyphur, M. J., & Pierides, D. C. (2019). Making quantitative research work: From positivist dogma to actual social scientific inquiry. *Journal of Business Ethics.* https://doi.org/10.1007/s10551-019-04189-6.