The Average Isn't Normal[1]

Henry M. Cowles* and Joshua Knobe**

* Department of History, University of Michigan

** Program in Cognitive Science and Department of Philosophy, Yale University

Within contemporary science, it is common practice to compare data points to the *average*, i.e., to the statistical mean. Because this practice is so familiar, it might at first appear not to be the sort of thing that requires explanation. But recent research in cognitive science gives us reason to adopt the opposite perspective. Research on the cognitive processes involved in people's ordinary efforts to make sense of the world suggests that, instead of using a purely statistical notion of the average, people tend to use a value-laden notion of the *normal*. This finding about ordinary cognition gives us reason to rethink certain familiar facts about scientific practice. In particular, it suggests that the fact that scientists so often make use of the statistical average should be seen as a highly surprising fact, the sort of thing that calls out for explanation. To understand it, we turn to work in the history of science, and especially to work on the ways in which the practice of science changed over the course of the 19th century.

Looking at the practices at work in contemporary science, one obvious and seemingly unremarkable fact is that scientists often compare data points to the *average*, i.e., to the statistical mean. This practice plays an absolutely central role in everyday statistical analysis. Indeed, it is such a commonplace part of scientific practice that it is easy to find oneself taking it for granted and not regarding it as worthy of explanation or exploration.

But now suppose we look instead at research in cognitive science on the processes that take place within people's minds ordinarily as they are trying to make sense of the world. Strikingly, this research indicates that people do not ordinarily understand data points by comparing them to the statistical average. Instead, people seem to employ a more value-laden notion of the *normal* (Bear & Knobe, 2017; Hitchcock & Halpern, 2014; Hitchcock & Knobe, 2009; Icard et al., 2017; Kominsky et al., 2015; Wysocki, 2020). As we will see, people's ordinary notion of the normal differs in important respects from the notion of the statistical average. Existing research in cognitive science has therefore explored the ordinary notion of the normal as part of an attempt to understand questions about the mind (see, e.g., Bear et al., 2020; Egré & Cova, 2015; Knobe & Samuels, 2013).

[1] We are very grateful to Kathryn Tabb for valuable comments on a previous version.

Our aim in this paper is to bring together the literature on practices taking place in systematic science with this literature on processes taking place within the mind. We argue that findings about the mind can teach us that certain facts about scientific practices that might initially seem perfectly obvious or unremarkable should actually be seen as highly surprising. One of these facts, we argue, is the fact that scientists compare data points to the average. This practice represents a substantial departure from the processes found in ordinary cognition, and it is worth thinking in detail about how scientific practice came to depart from people's ordinary understanding in this respect.

We then suggest that this divergence between scientific practice and ordinary understanding is best understood from a historical perspective. Research in the history of science points to striking changes in scientific practices over the course of the 19th century. We will argue that the practice of comparing data points to the average arose as part of that much larger change. At the beginning of the 19th century, scientists were often relying on the ordinary notion of the normal, but by the end, things had changed considerably. Statistical and other innovations during that time period had made possible an importantly different set of practices.

If this claim does turn out to be correct, it does not answer the question about how to explain divergence between the practices of science and that processes occurring ordinarily in the mind. Instead, it *deepens* that question. One wants to know why the change first arose and how it is sustained within everyday scientific practice. We do not have definitive answers to these questions, but we will be suggesting some hypotheses that might be worth investigating.

Although we will be focusing very specifically on questions about the average and the normal, our inquiry is clearly related to much broader questions about the relationship between philosophy of science and philosophy of mind. Within existing philosophy of science, there has been an enormous amount of research on the role that value judgments play in contemporary scientific practice. The present paper takes up a new approach to that traditional question. We look at the role that value judgments play ordinarily in people's minds, then argue that the role they play in scientific practice is importantly different from the role they play in ordinary cognition. Ultimately, then, our goal is to shed light on contemporary scientific practice by looking at the ways in which it differs from ordinary cognition — in other words, by addressing the role of values in contemporary science *indirectly*, through attention to ordinary cognition today and scientific work in the past.

1. Two Approaches to the Study of Values in Science

Let's begin, then, by situating our exploration within a broader context in the history and philosophy of science. We will be concerned here with questions about the role of value judgments in the practice of science, but will be pursuing this question in a slightly different way from the one that has become most common within existing research. We therefore begin by drawing a distinction between the traditional approach to research on this topic (what we call the 'direct approach') and the approach we will be pursuing here (which we will call the 'indirect approach').

1.1. The Direct Approach

The direct approach aims to characterize the way that contemporary science actually works. Research conducted via this approach has uncovered numerous important features of contemporary science. These features have often come as a surprise to people who are not already intimately familiar with the workings of contemporary science, and in many cases, they would be extremely difficult to discern even for people who are engaged in the practice of science every day.

Work within this first approach has provided a series of powerful arguments for the claim that value judgments play an important role in contemporary science. Researchers have argued that value judgments play a role in deciding how much evidence one needs before arriving at a conclusion in a given domain (Hempel, 1965; Rudner 1953) that some of the concepts used in social science are irreducibly value-related (e.g., Dupre, 2012); that values (help) determine how one makes sense of scientific results (Barnes, 1977); that value judgments play a role in how scientists choose between different modes of explanation (Longino, 1990); and that epistemic and non-epistemic values cannot be cleanly separated and inevitably end up influencing each other (Kitcher, 2003). Work in this vein continues up until the present day. For a sophisticated recent discussion, see Douglas (2009).

Abstracting away from the details of the individual arguments, let's offer a brief characterization of the importance of the direct approach to values in contemporary science as a whole. Much of the work in this area provides arguments for the thesis:

(Value-Ladenness) Value judgments play an important role in contemporary science.

This thesis can be contrasted with an opposing opinion about the lack of a role for value judgements in science:

(Value-Freedom) Value judgments do not play an important role in contemporary science.

To the extent that one started out with the assumption of Value-Freedom, it would be quite surprising to read work motivating Value-Ladenness. The history of the philosophy of science suggests that this assumption was prevalent among practitioners as recently as the mid-twentieth century, and that the arguments for Value-Ladenness cited above were surprising and controversial when published (Zammito 2004). Indeed, insofar as many people presume or expect Value-Freedom, arguments for Value-Ladenness remain both surprising and important (Oreskes & Conway 2010).

We mention this direct approach to values in contemporary science only to distinguish it from the project we will be pursuing here. Our project will not contribute anything new to this first stream of research. We will not be arguing either that the contemporary scientific notion of the average is value-free or that it is value-laden. Instead, we pursue an indirect approach.

1.2. The Indirect Approach

The indirect approach aims to shed light on the workings of contemporary science by exploring *other* ways of making sense of the world. While such work could take a number of forms, we will be focusing especially on cognitive science and history of science.

In a certain sense, the indirect approach also reveals surprising facts about contemporary science, but it does so in a very different way. It is not that the indirect approach teaches us new facts about how contemporary science works. Rather, the indirect approach takes facts we already knew about contemporary science and provides evidence that these facts should actually be seen as surprising. It does this by synthesizing work on adjacent modes of thinking to which contemporary science is frequently compared.

The best way to get a sense for this kind of research is not to describe it in general but rather to consider a prominent example. Consider work on *teleological explanation*. People tend to explain some things in terms of purposes while explaining other things in ways that do not involve purposes. In contemporary science, it is considered appropriate to explain the existence of certain objects in terms of purposes (chairs, buildings) but to explain others in ways that do not involve purposes (mountains, oceans). Research within cognitive science and history of science shows, however, that other modes of thought yield very different explanations of these phenomena. Cognitive scientists find that non-scientists seem drawn to give teleological explanations for phenomena that scientists explain non-teleologically. For example, one finds such explanations in children, in people with Alzheimer's, and even in trained scientists when they are forced to respond under extreme time pressure (Kelemen, Rottman & Seston, 2013; Lombrozo, Kelemen & Zaitchik, 2007). Similarly, historians of science have shown that teleology gave way to alternatives only recently, in the life sciences and elsewhere (Lenoir 1982; Riskin 2016). Taken together, this information leaves us with a very different understanding of the non-teleological explanations that are so often found in contemporary science. In other words, we might already have known that contemporary science attempts to explain the existence of mountains in ways that do not involve purposes (even if it sometimes fails). But after learning more about cognitive science and the history of science, we come to see that it is actually surprising that scientists use non-teleological explanations—and that we might want to figure out how, and why, they do so.

Our project adopts a similar approach to the use of the statistical average in contemporary science. We argue that it should be seen as surprising that value judgments do not play the same role in the use of this notion that they play in the notion of the normal that is so prominent both in people's ordinary cognition and in earlier scientific thinking. In other words, our aim is to take certain facts that you already knew about contemporary science and argue that these facts should be regarded as deeply surprising relative to facts about other ways of thinking.

Although our focus in what follows will be specifically on questions about the average and the normal, we will be drawing heavily on ideas from broader literatures on how contemporary science differs from other modes of thought. On one hand, we draw from the literature in cognitive science on the ways in which value judgments can impact a wide variety of different ordinary judgments. On the other, we draw from a broader literature in the history of science on the ways in which the role of value judgment in the practice of science changed over the course of the 19th century.

Recent work in cognitive science on the role of value judgments has shown that they have a far greater and more pervasive impact on their cognition than one might initially suppose. For example, one might initially think that people's moral values would impact their judgments about a certain range of questions (questions about how to live, questions about whether agents deserve praise or blame) but that there would be plenty of questions that people answer using cognitive processes that are unaffected by moral values (e.g., straightforwardly factual questions about whether one event caused another). Of course, one would not be surprised to find occasional exceptions to this generalization. There might be various unusual circumstances in which people's value judgments do end up impacting their intuitions about what seem like purely factual questions. Still, one might expect that, for the most part, the role of value judgment would be relatively circumscribed, with value judgments having an impact on the way people think about some questions but not others.

A large body of research over the past decade or so has challenged this assumption. This research suggests that people's moral judgments can actually impact their way of thinking about a wide variety of questions that might initially appear to be purely factual or value-neutral. Such questions include: whether one event caused another (Halpern & Hitchcock, 2014; Hitchcock, & Knobe, 2009; Kominsky, Phillips, Gerstenberg, Lagnado & Knobe, 2015), whether an agent acted intentionally (Knobe, 2010), whether an agent has knowledge of a proposition (Beebe & Buckwalter, 2010), whether an agent is happy (Phillips, De Freitas, Mott, Gruber & Knobe, 2017), whether an event is possible (Phillips & Cushman, 2017) or probable (Dalbauer & Hergovich, 2013)).

Considerable controversy remains about how to explain these effects (e.g., Alicke, Rose, & Bloom, 2011; Samland & Waldmann, 2016; Uttich & Lombrozo, 2010), but we will not be weighing in on that controversy here. For our purposes, the key point is not that these effects are best explained by one or another specific cognitive process, but rather just that the effects themselves truly do exist. This point has by now been amply confirmed by a wide-ranging research program in psychology and experimental philosophy.

Something similar is true of the work on values in the history of science. Over the last few decades, historians have shown just how important value judgments were in some of science's most canonical debates and discoveries. One of the most influential examples of this line of research is Shapin and Schaffer's *Leviathan and the Air-Pump*, which argued that a key debate in the Scientific Revolution was as much about the competing values of practitioners as it was about the theories and practices involved (Shapin and Schaffer 2011 [1985]). Similar arguments have been made about Enlightenment chemistry (Golinski 1992), Victorian physics (Smith 1999), and cognitive science (Cohen-Cole 2014), among many, many other topics. Much like work in philosophy of science, these historical books push back against the once-common assumption that modern science is completely value-neutral. Instead, historians have shown how values of many sorts played surprising roles in judgments about truth, evidence, causality, and other supposedly value-free domains.

Historians have also shown how the role of values has changed over time. While scholars hold different views about the exact nature of these changes, most seem to agree that a major shift in the ideal of value-neutrality among scientific practitioners took place during the nineteenth century. Explicit attempts to address the role of values *qua* values in science had increased markedly by the end of the nineteenth century (Proctor 1991). Alongside the value-neutral ideal arose what Theodore Porter has called "trust in numbers": deference to quantitative expertise that continues to characterize both science and politics today (Porter 1995b). The power of numbers was mirrored by the rise of what Lorraine Daston and Peter Galison have called "mechanical objectivity": efforts at self-denial among scientists that aimed to replace human judgment with observation aided by machines (Daston

& Galison 2007). Though these authors disagree about causes, they agree that the role of values in science fundamentally changed in the nineteenth century.

This shift was far-reaching, affecting everything from technical practice and training to primary education and the public understanding of science. At the beginning of the nineteenth century, it was still common for scientific thinkers to invoke value-laden notions of divine power; by the end of the century, this was much less common in scientific circles and at the research universities that emerged over this same period (Gregory 1992; Lenoir 1997). Of course, the relationship between science and values does not reduce to that between science and religion (Harrison 2015), but the two were caught up together in the nineteenth century. This is perhaps clearest in Max Weber's famous essay on "Science as a Vocation," in which he argued for the separation of science and religion and insisted that "whenever the man of science introduces his personal value judgment, a full understanding of the facts ceases" (Weber 1946 [1917]). Whatever one thinks of Weber's claim, it encapsulated a view that dominated scientific thinking by the early-twentieth century and continues to guide scientific practice and pedagogy today.

In short, existing research provides powerful evidence for the claim that there is some important difference between the role that value judgments play in contemporary science and the role they play in ordinary cognition and in the science of earlier periods. Our aim now is to apply this more general idea to the specific case of the average and the normal.

2. The Average and Normal

It is common practice within contemporary science to compare quantities to some measure of central tendency. One can compare quantities to the median or to the mode, but the most common approach is to compare quantities to the average (the mean). Thus, when one wants to make sense of the values of a given variable, one often proceeds by representing them in terms of the degree to which they are higher or lower than the average.

To illustrate, suppose that you have collected information about how much TV certain people watched yesterday. One way to represent this information would be just to list out the raw amounts:

4 hours

0.5 hours

2.5 hours

7 hours

Another approach, however, would be to represent each amount in comparison to the average. To do this, one computes the average and then subtract that amount from each data point. In the present example, that approach would involve transforming our four numbers as follows:

0.5 hours

-3 hours

-1 hours

3.5 hours

This transformation is referred to as 'centering,' and scientists use it all the time when trying to understand patterns of data.

Taking things just a little bit farther, we might then divide each number by the standard deviation. This involves transforming each number into a *z-score*:

0.2

-1.1

-0.4

1.3

This transformation is known as 'standardizing' or, in a telling phrase, 'normalizing.' It plays an absolutely foundational role in the field of statistics as currently practiced.

Before proceeding any further, it might be helpful to emphasize that the point we are making here is very different from the kinds of points one usually finds in philosophical work on values in science. Within the existing literature, there is a tendency to focus on aspects of scientific practice that are not at all obvious and can only be revealed through serious empirical and conceptual research (e.g., Douglas, 2009). By contrast, the point we are making is one that would usually be regarded as entirely uninteresting. All we are saying is that scientists often make use of a particular statistical procedure. This procedure is widely taught in introductory statistics courses, and scientists invoke it quite explicitly all the time. No real insight or sophistication is required to see it at work.

Our aim is to show that this seemingly uninteresting aspect of contemporary science should actually be regarded as highly surprising and worthy of further exploration. The argument relies on a comparison between contemporary science and two other modes of thought.

2.1. Normality in cognitive science

To begin with, we can ask about the relationship between the precise statistical concept of the average and people's ordinary ways of making sense of the world. One possible hypothesis would be that the precise statistical concept is best understood as providing a more formal version of the sorts of notions people in ordinary cognition. People might not ordinarily go through the steps necessary to precisely calculate the average, but they do have an intuitive sense that certain quantities are *normal*. For example, we might have an intuitive sense for the 'normal amount of TV.' Then we might classify other amounts in terms of their relationship to the normal ('a little bit less than normal,' 'far greater than normal'). One hypothesis would be that statistical concepts just give us a more precise, formal way of spelling out these ordinary notions. Thus, one might think that a statistical concept like a z-score of 2 is best understood as simply spelling out more precisely the very same thing we might express in a vague, intuitive manner by using an expression like 'much more than normal.'

Recent work in cognitive science shows that this hypothesis is mistaken. People's ordinary thought does seem to include a notion of the normal that in some ways resembles a statistical measure of central tendency, but this ordinary notion appears to differ from the average in one very important respect. Specifically, recent empirical studies suggest that the ordinary (as opposed to the scientific) understanding of normality is *value-laden* (Bear & Knobe, 2017; Egré & Cova, 2015; Hitchcock & Halpern, 2014; Hitchcock & Knobe, 2009; Icard, Kominsky & Knobe, 2017; Kominsky, Phillips, Gerstenberg, Lagnado & Knobe, 2015; Wysocki, 2020). People's intuitions about which quantity counts as normal are not simply sensitive to their statistical beliefs — "normal" is not simply a synonym for "average" — but are also sensitive to people's beliefs about the degree to which quantities would be good or bad from a more evaluative standpoint.

To get a sense for this phenomenon, consider again the case of amounts of TV. In a recent study, one group of participants were asked to guess the average amount of TV people watch per day (a statistical judgment), while another was asked about the ideal amount of TV to watch per day (a value judgment). Unsurprisingly, participants gave a quite high amount for the average and a much lower amount for the ideal. A third group of participants was then asked about the *normal* amount of TV to watch in a day. The results show that people's judgments about the normal were not simply identical to their judgments about the average. Rather, the perceived normal amount was intermediate between the average and the ideal. This pattern of judgments did not arise only for the case of TV; it arose systematically across a wide variety of quantities, including everything from amounts of exercise for a person to do in a week to percentages of students to be bullied in a middle school (Bear & Knobe, 2017). The pattern as a whole suggests that people's notion of the normal is shaped by a mixture of statistical and evaluative considerations.

This same basic effect has emerged in studies using many other methodologies. For example, another study showed that participants' judgments about whether a particular political view was normal do not depend only on the statistical prevalence of that view. Rather, each participant's judgment depends in part on whether that participant regards the view itself as good or bad (Wysocki, 2020).

Difficult questions arise about precisely how people integrate statistical judgments and value judgment into an overall judgment of normality. It is clear that people's ordinary notion of normality somehow brings together judgments about what is statistically average with judgments about what is evaluatively ideal, but it is not yet clear exactly how to understand this more integrated notion. Yet, regardless of how these difficult questions are resolved, it seems that the scientific concept of the average is importantly different from the ordinary notion of normality. Perhaps it can be shown that value judgments play some role in people's way of understanding the average, but they do not seem to play the role that is characteristic of normality judgments. To the extent that a person has the relevant knowledge, she can take a list of numbers and, without making any further value judgment, calculate the average using a straightforward mathematical procedure.

With this framework in place, let's consider again people's ordinary ways of comparing quantities to a standard. To say that a quantity counts as (for example) 'large' or 'small,' we need to compare the quantity in question to some standard (Kennedy, 2013). But what standard do people use? One obvious hypothesis would be that people compare quantities to the average. For example, it might be thought that people would see someone as watching a 'large amount of TV' when the amount that person watches is sufficiently above the average (say, anything higher than $z = 1$). However, existing studies suggest that this is not the case. Instead, people seem to regard a quantity as large when it is larger than their undifferentiated representation of the *normal* (Egré & Cova, 2015). Thus, the threshold one needs to surpass to count as watching a 'large amount of TV' is not the statistical average but a value-laden notion of the normal (Bear & Knobe, 2017). Similar effects have been observed for numerous other psychological phenomena. The notion of normality has been implicated in people's intuitions about prototypicality, causation, even the folk-biological concept of innateness. In all of these cases, studies show that people's intuitions are not shaped solely by statistical considerations. Rather, people's intuitions appear to be shaped in each case by a mixture of statistical and evaluative considerations (Barsalou, 1985; Bear & Knobe, 2017; Icard et al., 2017; Knobe & Samuels, 2013; Kominsky et al., 2015). Thus, the available evidence suggests that people's integrated statistical/evaluative notion of normality plays a pervasive role in cognition.

These facts about people's ordinary understanding give us reason to adopt a different view of the scientific practice of comparing data points to the average. This concept plays such an important role in our scientific practices that it is easy to take it for granted, and it might be difficult to see this concept as involving any kind of important innovation. However, existing studies suggest that the concept of an average actually involves a fundamental departure from people's ordinary mode of thought.

2.2. Normality in the History of Science.

To understand how the descriptive notion of the average departs from the evaluative notion of the normal, it is instructive to study not only how far apart they are in the present, but also how they came apart in the past. As with the case of cognitive science, history suggests that ideas about normality in contemporary science are relatively value-free and should be regarded as surprising, given the value-laden history of their usage and development in the nineteenth century.

Historians of science have highlighted stages in the history of normality, from the early development of "political arithmetic" in the seventeenth century (Deringer 2018) through the advent of classical probability in the eighteenth century (Daston 1988) to the rise of statistical thinking in the nineteenth century (Porter 1986). In 1800, depending on the field you were in and the questions you asked, it might have made sense to invoke value-laden ideas of "normality" when describing the natural world, but by 1900 a more value-neutral statistical notion of "the average" had taken over in most fields. This transformation occurred in different ways in different areas, but the general trend was to invoke a value-neutral meaning of "normal" as part of a more general movement to scrub values from science. Ian Hacking has called this process "the taming of chance": that is, the separation of statistical terms for describing the world from value judgments of (and value-laden terms for) whatever was being described (Hacking 1990).

It can be instructive to review a specific case. Evaluative and descriptive meanings of "normality" were intertwined in the field of medicine during the nineteenth century. Physicians held "the normal state" to be a healthy one—bodies operating at a temperature and a rhythm conducive to the functions of whatever organ or organism was being observed. The opposite of "the normal," as Georges Canguilhem famously showed, was "the pathological," and in this context it was clear that one was to be preferred and one avoided (1991 [1943]). In medicine, this value-laden meaning has stuck with "normality." This persistence is illustrated by the common question: "Doc, is this normal?" A positive answer is calming, a negative one cause for concern. According to Canguilhem, this value-laden sense of "normality" is inevitable in both medicine and physiology. "It seems to us," he concluded, "that physiology has better to do than to search for an objective definition of the normal, and that is to recognize the original normative character of life."

Even if the value-laden meaning of "normal" played an important role in medicine through the twentieth century (Wellman 1958) up to the present (Manrai 2018), its importance and role changed during the nineteenth century. The rise of "scientific medicine" meant, in part, aspirational identification with an approach to the "normal" that was increasingly statistical alone. Whether or not this was achieved (Tiles 1993), normality's changing meaning produced tension and debate in the changing medical landscape of the nineteenth century.

One way to trace this is in the development of so-called "normal curves" and their application to human affairs. The use of "normal curves" originated with Pierre-Simon Laplace and Johann Carl Friedrich Gauss, two mathematicians who plotted the distribution of observations of a given data point in order to reduce error and arrive at the true value. The original "normal curve" was thus a plot of human errors, not of natural phenomena. Gradually, however, this pattern was reimagined as a part of nature itself—the distribution of observational errors was transposed onto the things that were being observed, natural and human alike (Hacking 1990).

It was Francis Galton who argued that the "normal curve" captured something out there in the real world, and in his hands descriptive statistics was joined to probability theory and the mathematical prediction of complex patterns (Porter 1986). This marriage of statistics and mathematics, achieved during the nineteenth century, seemed to help scientists separate the descriptive and evaluative senses of "normal." If "the normal heart" still meant the one that pumped blood like it should, it also became possible for it to mean something else: the average heart, which in a given population (under specific stresses) might actually be a poor pump indeed.

Importantly, scientific authors recognized early on that the term "normal" blurred the very boundary they sought to shore up. This applied both to particular values on distribution curves and to those curves themselves. Thus, Galton could say in 1895 that "it is only a few curves that are symmetrical and conform closely to the normal law of facility of error," concluding: "When the conformity between the observations and the normal law ceases to be close, the latter must be applied warily" (Galton 1895:319). He was identifying a zeal for "normalizing" that was only beginning. Karl Pearson echoed his wariness when discussing "the *normal* curve, which name, while it avoids an international question of priority, has the disadvantage of leading people to believe that all other distributions of frequency are in one sense or another 'abnormal'" (Pearson 1920). Both Galton and Pearson worried about how descriptive and evaluative notions of "normal" ran together.

Here we see the emergence of statistical "curve fitting" in two competing senses. First, there is the sense pursued by Pearson: the effort to fit statistical curves to the data at hand. Second, however, there is another sense: the effort to fit the data themselves to an idealized "normal" curve. This latter sense is controversial, of course. Its most persistent critic was Michel Foucault, a student of Canguilhem. Foucault argued that the term "normal" carried with it a value-laden weight that had the power to shape the contexts in which it was used, even if those who wielded the term insisted it had only statistical meaning. Over the course of the nineteenth century, this power increased as ideas of "normality" were enforced in places like hospitals, schools, and—most famously—prisons (Foucault 1977 [1975]). It was to Foucault that scholars like Porter and (more explicitly) Hacking looked as they unearthed the moral history of "normal" in statistics and everyday life.

Subsequent work in this vein has proven Pearson's worries correct. Confusion about what "normal" means continued well into the twentieth century. Public opinion polls, for example, built "the average American" out of values associated with American-ness in the early- to mid-twentieth century (Igo 2007). Other, related developments confirm this continued confusion. The rise of "normal controls" in biomedical research, for example, were *meant* to be "average" but came with their own assumptions about ideal body types and the generalizability of male bodies (Stark forthcoming). Such studies highlight how difficult it has been to separate the evaluative and descriptive aspects of "normality"—which has led some, echoing Canguilhem, to insist that there is no such distinction in any meaningful sense (Sholl 2017).

And yet, something *did* change in the nineteenth century. This can be illustrated by the gradual success of applying statistical methods to the interpretation of human affairs. At the start of the nineteenth century, the idea that human behaviors—including crime and suicide—were regular or predictable seemed to challenge the idea of free will (Porter 1986). In part, this was because laws like those governing astronomical phenomena were thought to be divine expressions of an ideal order (hence the very notion of "law"). How could this hold true of suicide? Adolphe Quetelet broke this taboo in the 1830s, but he did so by applying such laws only to "the average man" (*l'homme moyen*), not to individual men. This idea, perhaps ironic, shielded Quetelet and his "social physics" from charges of immorality and blasphemy (Porter 1985, Porter 1995a). But the door was open for purportedly value-free tools to be used on the ultimate value-laden object: society itself.

By the end of the century, Emile Durkheim could dispense with Quetelet's circumspection. In his canonical study, *Suicide* (2002 [1897]), he attributed regularities in this tragic behavior to the "normal" distribution of happiness and family values, with "normal" encompassing both positive and negative emotions. While this may *seem* value-laden—and in certain ways, it surely was—it is important to note that Durkheim was able to describe a "normal" rate of suicide without running afoul of ideas about divine laws or seeming to imply that suicide was "good." It was a distribution, he could argue, and nothing more. What is surprising is that, eventually, people agreed.

Galton took this one step further. If Quetelet and Durkheim had made it possible to include both good and bad under the "normal" umbrella, Galton claimed to identify "normal" ranges that were *only* bad. As one of the founders of the eugenic movement, Galton saw something like "normal intelligence" as decidedly less than ideal. Being "only average" was something to be improved upon (Hacking 1990). Thus, by the end of the nineteenth century it was possible not only to imagine "the normal" as a descriptive, rather than evaluative, claim, but even to imagine it as evaluative *in a new sense*—as less than you might hope for, assuming you wanted to fall further down the curve.

The point is that, by 1900, the meaning of "normal" had bifurcated. While "normal" still implied—and indeed, still implies (Metzl and Kirkland 2010)—a value judgment when it came to one's bodily health, it could also indicate something like "average." Then, as now, the task was to distinguish between the two in scientific contexts, such that statistical tools could be applied for descriptive purposes without introducing (or seeming to) value judgments that, by then, seemed to have no place in science. While we can identify breaches of these standards today, it is worth noting their emergence as a means of indirectly analyzing the role of values in contemporary science.

2.3. Summary and interim conclusion

Within contemporary science, we find a practice of "normalizing" data that involves comparing each data point to the mean. This aspect of contemporary science might at first seem perfectly straightforward, but we have argued that it contrasts sharply with what is found in other modes of thought. Both in people's ordinary intuitions and in earlier periods in the history of science, we find a notion of normality that is determined by a mixture of statistical and evaluative considerations. The idea of comparing each data point to a level determined by purely statistical considerations, we have suggested, is best understood as a striking innovation, first introduced in the 19th century.

So then, what implications might this claim have for the study of contemporary science? One possible answer would be that we could use it as an additional piece of evidence within the kind of inquiry that is already quite well-established in the philosophy of science literature. As noted above, research in this area has uncovered numerous subtle ways in which value judgments play a role in scientific practice. In keeping with this tradition, we might now try to uncover a subtle yet important respect in which contemporary science actually *does* use a value-laden notion of normality. For example, existing work in the philosophy of psychiatry has wrestled with how to understand the notion of normality used within psychiatry and whether this notion is best understood as a value-laden one (see, e.g., Washington, 2016). Perhaps a study of value-laden notions of normality within other modes of thought could help to illuminate some of the difficult issues that arise here.

Now, one possible view would be that it is only when we begin investigating these more subtle aspects of scientific practice that our inquiry truly becomes philosophically interesting. After all, it was perfectly obvious all along that scientists often normalize their data by comparing each data point to the mean. The more interesting question, one might think, is whether scientists sometimes depart from this straightforward statistical procedure and begin employing a more value-laden notion.

There is certainly something right in this view, but we have been trying to show that there is also something right in going against it. In a certain sense, it is not surprising at all that scientists sometimes make use of a value-laden notion of normality. If we discover, e.g., that the practice of psychiatry involves comparing each person to a value-laden standard that is intermediate between the statistical average in the prescriptive ideal, we are basically just discovering that psychiatrists do the same thing that people do all the time. They are human, after all.

As such, it is remarkable that scientists even come close to normalizing data using statistics alone. For example, it should be seen as remarkable that scientists are able to consider a person who watches four hours of TV per day and think of the quantity not as "higher than the normal" but rather "about average." In doing this, they are engaged in a practice that involves a very serious departure from ordinary cognition and earlier forms of scientific thought. Somehow scientists are able to use a set of procedures that make it possible for them to compare each data point to a standard that is derived simply by considering the statistical distribution of the data and does not take into account the degree to which various quantities are good or bad.

3. Explaining use of the statistical average

We have been arguing for a change in perspective about what should be taken for granted and what calls for explanation. An obvious view would be that we can simply take for granted the fact that people sometimes compare data points to the average, and that the only thing that calls for explanation is the fact that people sometimes compare data points to a value-laden notion of the normal. We have argued against that view. In its place, we argued for a view on which the fact that people ever compare data points to the statistical average is what calls for explanation. We now ask what that explanation might be.

The thing that requires explanation here is a certain sort of divergence between the practices found in systematic scientific research today and the processes found in people's more intuitive, ordinary cognition or in past scientific research. We will be taking up the effort to explain that divergence on two different levels. On a *historical* level, we want to understand how scientific practice diverged in this way from people's ordinary understanding. Then, at a more *contemporary* level, we want to understand the processes that sustain that divergence today.

Clearly, the questions we are taking up here touch on some extremely abstract issues about the relationship between complex practices sustained by large-scale institutions and the ordinary intuitions generated by cognitive processes that take place within individual human minds. Our question is about one specific case in which the two appear to diverge, and in what follows, we will be focusing entirely on this one specific case. Still, a close examination of this one case has some potential to shed light on the more general issue.

While there are many strands of existing research—both historical and cognitive—that can help explain the divergence in this specific case, no single explanation dominates. In what follows, we explore a number of different plausible accounts of this divergence that could contribute to the indirect approach to values in contemporary science toward which we are pointing.

3.1. Historical Explanations

Our first question is historical. As our case study shows, scholars have documented a change over the course of the nineteenth century—a shift in invocations of "the normal," away from a value-laden notion to something more like the statistical average. However, this account was purely descriptive. That is, our claim was simply that, with respect to normality, the role of values that one finds at the beginning of the nineteenth century differs from the one found at century's end. But one might pose a deeper question: *why* did scientific practice change in this way?

We have already seen two approaches to answering such a question. One is political: Porter, for example, argued that the rise of quantification was less about its scientific superiority and more about the *political* power of statistical arguments. Applied to explain the changing meaning of "the normal," this approach would link increasing references to the statistical mean to the rise of a "trust in numbers" within political culture. The other is epistemological: Daston and Galison's "mechanical objectivity" stemmed from the impact of Kant's account of the limits of human knowledge. In order to extend this claim to our specific case, Daston and Galison might link the rise of the statistical mean (and the fall of more value-laden notions of normality) to a broader distrust of the individual as the source of knowledge of ultimate causes. Both approaches would link the shift in normality we have documented to shifts in political or epistemological attitudes in the broader culture.

Of course, one could also look more narrowly at how practices and pedagogy themselves were changing in this period. For example, the relative reduction in value-ladenness could be linked to technological shifts as well. The development and spread of tabulating machines, mechanical calculators, and other aids to computation during the nineteenth century would have made the statistical mean of large datasets easier to compute and, eventually, practically ready to hand. Indeed, even the *dream* of such machines played a role in how the human mind was understood and the value placed on a range of scientific practices (Jones 2016). The spread of what Ursula Klein has called "paper tools"—including specific inscription practices and mathematical formulae shared by communities of researchers—no doubt helped make arriving at the statistical mean not only easier but also a matter of course as researchers performed specific computations without thinking twice (Klein 2001). Such technological shifts explain how value-laden notions of normality could give way to more statistical notions *without* requiring recourse to the kinds of conscious rejection of values implied by the political and epistemological shifts traced above.

Changes in *publishing* would also have contributed to the shift we are exploring. From the rise of the scientific journal as a specific site for publishing (Csiszar 2018) to the development of peer review and other mechanisms for standardizing practices in the field (Baldwin 2015), practices such as comparing data points to the mean and—eventually—computing z-scores would have become expected elements of the scientific paper that was gradually taking more or less its modern form. The same goes for calls for replicability in the sciences. As replication became central to adjudicating matters of trust and truth, notions of "the normal" that depended upon unarticulated values would have been less acceptable than those (supposedly) shorn of those values and limited to comparisons one's colleagues could repeat with their own pen and paper (Cantor and Shuttleworth 2004, Fyfe

2012). As science came to be seen as "out there," embodied in journals and groups but not in individuals, it also came to be seen as natural—and, thus, value-free (Cowles 2017). It is easy to see how the standardization of publishing forms would have naturalized procedures like comparing data to a statistical mean and denaturalized the reliance on value-laden notions of "the normal."

The same goes, to take one final example, for changes in scientific training in particular and science education in general. Scholars have shown how a new secularism in university education (Reuben 1996) and debates over the place of science in such a context (Jewett 2012) paved the way for the decline in value-ladenness we have sketched. From the rise of state-based science curricula in the nineteenth century to the so-called "general science" movement in the early-twentieth century, there emerged a felt need for the kind of standardized practices that anyone could replicate on the way to gaining familiarity with scientific fields or specific scientific credentials (White 2003, Rudolph 2005a, Rudolph 2005b). One notion that gained traction during this shift was the idea of a single, shared scientific method based on hypothesis testing (Cowles 2020), which went on to structure how basic scientific norms have been taught to children ever since (Rudolph 2019). Something similar may well have occurred with regard to the question of comparing data to the average, such that—by the turn of the twentieth century—the performance of such computations became definitional for rigorous science. Such an explanation could also help explain why this practice became for practitioners but relatively rare among others: by *identifying* science with such a practice, other ways of imagining "the normal" would be allowed to persist in areas (including everyday life) not held to the same standard.

In presenting these three specific hypotheses from the work of others, we certainly don't mean to suggest that we already have in hand a complete and accurate explanation of the change over time in the role that value judgments play in science in general or the rise of statistical notions of the average in particular. Rather, we present these hypotheses to illustrate the sorts of ideas that might be pursued in further work.

3.2. Contemporary Explanations

Another question arises about how to understand the ways in which this departure from other modes of thought is sustained within the day-to-day practice of science. Given that people's ordinary intuitive mode of thought seems to involve comparing data points to the normal, how is it possible for scientific research to proceed in this very different way, that involves comparing data points to the statistical average?

One possible view would be that this phenomenon is to be explained in large part in terms of cognitive processes taking place *within the minds of individual scientists*. Within existing work on other phenomena, it has often been suggested that even trained scientists retain the same intuitions found in non-scientists but that they are able to override these intuitions and rely instead on a process of more controlled conscious reasoning (Kelemen, Rottman & Seston, 2013; Shtulman & Valcarcel, 2012). A similar process might explain scientists' thinking about the average and the normal. Perhaps scientists have an intuitive tendency to compare data points to the normal but they are sometimes able to override this tendency and instead compare data points to the statistical average. Of course, this type of overriding can only take us so far. Scientists are human beings, and they will inevitably be guided in many cases by more ordinary modes of thought. Still, even if scientists only occasionally override their ordinary way of applying value judgments, this occasional occurrence should be seen as a remarkable and deeply important aspect of scientific practice.

A second possible view would be that divergences from ordinary thought are sustained in large part by the use of *technology*. Contemporary scientific work involves a complex interplay between individual scientists' minds and external technologies. Indeed, scientific work always has. But the nature of that relationship has shifted over time, and specific practices that were once done by hand have been offloaded to technologies like computers, with major consequences for the nature of scientific thinking. An obvious hypothesis would be that it is in part this interplay that more generally makes possible certain kinds of divergence from more ordinary modes of thought. Perhaps this type of hypothesis can also explain people's use of the statistical average

At the most basic level, there is the fact that people do not need to calculate the average in their heads; they can do the calculations using a pencil and paper. Thus, a person might determine that a given data point is above the average, but the process used to make that determination does not take place entirely within the person's own head. Rather, the determination is a product of a interplay between processes taking place within the person's head and processes taking place in an external technology (the pencil and paper). In this way, the process of comparing data points to the statistical average is quite different from the process usually used to determine whether something falls above or below the normal.

But of course, in much work within contemporary science, the reliance on external technology is far more extensive. For example, to calculate a correlation coefficient, one needs to compare each data point to the average, but it is not as though scientists typically go through this process using pencil and paper. Almost always, the process is executed by a computer, and on many occasions, the scientists aren't thinking at all about the actual computations the computer is carrying out. The result is a striking divergence between what is happening within science and what is happening within the heads of individual scientists. Science relies on a procedure that involves comparing data points to the average, but this does not mean that there needs to be any process at all in which individual scientists compare data points to the average.

Third, and perhaps most importantly, there is the *social* character of science. Scientific progress is not usually the product of an individual scientist working in isolation. Instead, scientists typically work in teams, and there is usually a structure set up such that each team will not be successful unless its work is accepted by other teams. An individual scientist may have her own cherished values, but she will also be embedded in a larger structure that is set up in such a way that her work cannot succeed without the blessing of various other people, many of whom will have quite different values.

Suppose now that each individual scientist has certain values but that different scientists have different values. Each individual scientist might show a tendency to develop an understanding of the normal that is deeply informed by her own values, but as long as other scientists do not share those values, the social character of science may lead to an outcome in which her published work is not simply a reflection of her own tendencies. Thus, even if each individual scientist shows a tendency to use something like the ordinary notion of the normal, science as a whole might be drawn toward practices that more closely approximate a purely statistical notion of the average.

Finally, we might consider hypotheses that combine a number of these factors together. One of the most appealing such hypotheses would be that (a) there is a mechanism within the minds of individual scientists that allows them to use a purely statistical notion of the average but that (b) this mechanism is created or sustained by a process that requires either external technology or the social character of science.

To give a simple analogy, studies have shown that people who frequently do arithmetic by using an abacus eventually acquire the ability to do arithmetic in their minds using what is called a "mental abacus" (Frank & Barner, 2012). In much the same way, it might be that people first acquire the ability to think in terms of a purely statistical average either by using technology or by relying on the social character of science. However, as people continue using this method, it might be that they eventually come to be able to conceptualize things in a purely statistical way even without relying in the moment on either extra technology or external social connections. The extent to which scientists actually are able to do this is, of course, an open and very interesting empirical question.

4. Conclusion.

Our inquiry has been concerned with the practice of comparing data points to the statistical average. We argued that evidence from cognitive science and from the history of science gives us reason to regard this practice as highly surprising. The fact that this practice exists at all should be seen as something that calls for explanation, and we have sketched a number of possible ways to explain it.

Although we have focused very narrowly just on questions about the use of the statistical average, the approach introduced here could potentially be applied to numerous other problems. To give just one example, consider the notion of *essence*. Work on contemporary science suggests that scientists might make use of a notion of 'causal essentialism,' in which essences are understood as hidden factors that causally explain observable features (Putnam, 1975). But within work in the cognitive science of ordinary judgments and in the history of science, there is growing evidence of a more value-laden notion of essentialism (for cognitive science evidence, see Bailey, Knobe, & Newman, forthcoming; Gelman & Rhodes, 2012; Newman & Knobe, 2018; for historical evidence, see Daston & Galison, 2007; Dear 2014; Hacking 2007; Müller-Wille, 2013). Thus, there is at least some reason to think that we might face a real question as to why the understanding of essence at work in contemporary science departs from people's ordinary understanding of essence.

More generally, existing research has furnished us with an enormous amount of information about the role of values in contemporary science and also, separately, about the role of values in the processes at work ordinarily in people's minds. A key task now will be to bring those two literatures together. Looking across a whole range of different concepts, we need to explore the ways in which the role of value in contemporary science might depart from the role of value in more ordinary thinking. Ultimately, then, our inquiry into the average and the normal is perhaps best understood as just one case study in what is sure to be a far broader phenomenon.

References

Alicke, M. D., Rose, D., & Bloom, D. (2011). Causation, norm violation, and culpable control. *Journal of Philosophy*, 108, 670-696.

Bailey, A. H., Knobe, J., & Newman, G. E. (forthcoming). Value-based essentialism: Essentialist beliefs about social groups with shared values. *Journal of Experimental Psychology: General.*

Baldwin, M. C. (2015). *Making Nature: The History of a Scientific Journal*. University of Chicago Press, USA.

Barnes, B. (1977). *Interests and the Growth of Knowledge*. Routledge, UK.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 629.

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition*, 167, 25-37.

Bear, A., Bensinger, S., Jara-Ettinger, J., Knobe, J., & Cushman, F. (2020). What comes to mind? *Cognition*, 194, 104057.

Beebe, J. R., & Buckwalter, W. (2010). The epistemic side-effect effect. *Mind & Language*, 25, 474-498.

Canguilhem, G. (1991 [1943]). *The Normal and the Pathological*. Zone Books, USA.

Cantor, G. N., and Sally Shuttleworth, eds. (2004). *Science Serialized: Representations of the Sciences in Nineteenth-Century Periodicals*. MIT Press, USA.

Cowles, H. M. (2017). "History Naturalized." *Historical Studies in the Natural Sciences* 47(1), 107–16.

———————. (2020). *The Scientific Method: An Evolution of Thinking from Darwin to Dewey*. Harvard University Press.

Csiszar, A. (2018). *The Scientific Journal: Authorship and the Politics of Knowledge in the Nineteenth Century*. University of Chicago Press.

Dalbauer, N., & Hergovich, A. (2013). Is what is worse more likely?—The probabilistic explanation of the epistemic side-effect effect. *Review of Philosophy and Psychology*, 4, 639-657.

Daston, L. (1988). *Classical Probability in the Enlightenment*. Princeton University Press, USA.

Daston, L. & Galison, P. (2007). *Objectivity*. Zone Books, USA.

Dear, P. (2014). Darwin's Sleepwalkers: Naturalists, Nature, and the Practices of Classification. *Historical Studies in the Natural Sciences, 44*, 297-318.

Deringer, W. P. (2018). *Calculated Values: Finance, Politics, and the Quantitative Age*. Harvard University Press, USA.

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

Dupré, J. (2012). The inseparability of science and values. In *Drunk on Capitalism. An Interdisciplinary Reflection on Market Economy, Art and Science* (pp. 13-24). Springer, Dordrecht.

Durkheim, E. (2002 [1897]). *Suicide: A Study in Sociology*. Routledge, UK.

Egré, P., & Cova, F. (2015). Moral asymmetries and the semantics of many. *Semantics and Pragmatics*, 8, 13-1.

Foucault, M. (1977 [1975]). *Discipline and Punish: The Birth of the Prison*. Pantheon Books, USA.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, 141, 134.

Fyfe, A. (2012). *Steam-Powered Knowledge: William Chambers and the Business of Publishing, 1820-1860*. University of Chicago Press, USA

Galton F. (1895). A new step in statistical science. *Nature*. 51: 319.

Gelman, S. A., & Rhodes, M. (2012). Two-thousand years of stasis: How Psychological Essentialism Impedes Evolutionary Understanding. *Evolution challenges: Integrating research and practice in teaching and learning about evolution*, 1-26.

Hacking, I. (1990). *The Taming of Chance*. Cambridge University Press, USA.

————————. (2007). Natural Kinds: Rosy Dawn, Scholastic Twilight. *Royal Institute of Philosophy Supplement* 61, 203-239.

Halpern, J. Y., & Hitchcock, C. (2014). Graded causation and defaults. *The British Journal for the Philosophy of Science*, 66(2), 413-457.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy,* 106, 587-612.

Hempel, C. (1965). "Science and Human Values." In *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, 81–96. New York: The Free Press.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy,* 106(11), 587-612.

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, 161, 80-93.

Igo, S. E. (2007). *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. Harvard University Press, USA.

Jewett, A. (2012). *Science, Democracy, and the American University: From the Civil War to the Cold War*. Cambridge University Press, UK.

Jones, M. L. (2016). *Reckoning with Matter: Calculating Machines, Innovation, and Thinking about Thinking from Pascal to Babbage*. University of Chicago Press, USA.

Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, *142*(4), 1074.

Kennedy, C. (2013). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge.

Kitcher, P. (2003). *Science, truth, and democracy*. Oxford University Press.

Klein, U. (2001). "Paper Tools in Experimental Cultures." *Studies in History and Philosophy of Science Part A* 32(2): 265–302.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, 33, 315-329.

Knobe, J., & Samuels, R. (2013). Thinking like a scientist: Innateness as a case study. *Cognition*, 126(1), 72-86.

Kominsky, J. F., Phillips, J., Gerstenberg, T., Lagnado, D., & Knobe, J. (2015). Causal superseding. *Cognition*, 137, 196-209.

Lenoir, T. (1982). *The Strategy of Life: Teleology and Mechanics in Nineteenth-Century German Biology*. University of Chicago Press, USA.

Lombrozo, T., Kelemen, D., & Zaitchik, D. (2007). Inferring design: Evidence of a preference for teleological explanations in patients with Alzheimer's disease. *Psychological Science*, 18, 999-1006.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Manrai A.K., C.J. Patel, and J.P.A. Ioannidis. (2018). "In the Era of Precision Medicine and Big Data, Who Is Normal?" *Journal of the American Medical Association* 319(19):1981–1982.

Metzl, J. and A. Kirkland. (2010). *Against Health: How Health Became the New Morality*. NYU Press, USA.

Müller-Wille, S. (2013). Systems and How Linnaeus Looked at Them in Retrospect. *Annals of Science* 70 (3): 305-317.

Newman, G. E., & Knobe, J. (2018). The essence of essentialism. *Mind & Language*.

Oreskes, N., & Conway, E.M. (2010). *How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press, USA.

Pearson, K. (1920). Notes on the History of Correlation. *Biometrika*, 13(1), 25-45.

Phillips, J., & Cushman, F. (2017). Morality constrains the default representation of what is possible. *Proceedings of the National Academy of Sciences*, 114, 4649-4654.

Phillips, J., De Freitas, J., Mott, C., Gruber, J., & Knobe, J. (2017). True happiness: The role of morality in the folk concept of happiness. *Journal of Experimental Psychology: General*, 146, 165.

Prasada, S., & Dillingham, E. M. (2006). Principled and statistical connections in common sense conception. *Cognition*, 99(1), 73-112.

Porter, T.M. (1985). The Mathematics of Society: Variation and Error in Quetelet's Statistics. *The British Journal for the History of Science, 18*(1), 51-69.

—————. (1986). *The Rise of Statistical Thinking, 1820–1900*. Princeton University Press, USA.

—————. (1995a). "Statistical and Social Facts from Quetelet to Durkheim." Sociological Perspectives 38(1): 15–26.

———. (1995b). Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press, USA.

Putnam, H. (1975). The meaning of 'meaning'. Philosophical papers, 2.

Quetelet, A. (1835). *Sur l'homme et le développement de ses facultés.*

Reuben, J.A. (1996). *The Making of the Modern University: Intellectual Transformation and the Marginalization of Morality.* University of Chicago Press, USA.

Riskin, J. (2016). *The Restless Clock: A History of the Centuries-Long Argument over What Makes Living Things Tick.* University Of Chicago Press, USA.

Rudolph, J. L. (2005a). "Epistemology for the Masses: The Origins of 'The Scientific Method' in American Schools." *History of Education Quarterly* 45(3), 341–76.

———. (2005b). "Turning Science to Account: Chicago and the General Science Movement in Secondary Education, 1905–1920." *Isis* 96(3), 353–89.

———. (2019). *How We Teach Science: What's Changed, and Why It Matters*. Harvard University Press.

Rudner, R. (1953). The scientist qua scientist makes value judgments. *Philosophy of Science*, 20(1), 1-6.

Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, 156, 164-176.

Sholl, J. "Nobody is Normal." *Aeon* (31 January 2017).

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, 124, 209-215.

Stark, L. J. (Forthcoming). *The normals: A people's history of human experiment*. University of Chicago Press, USA.

Tiles, M. (1993). "The Normal and the Pathological": The Concept of a Scientific Medicine." *British Journal of the Philosophy of Science* 44 (4), 729-742.

Uttich, K., & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116(1), 87-100.

Washington, N. (2016). Culturally unbound: Cross-cultural cognitive diversity and the science of psychopathology. *Philosophy, Psychiatry, & Psychology*, 23(2), 165-179.

Weber, M. (1946 [1917]). "Science as a Vocation." in *Science as a Vocation*, trans. H.H. Gerth and C.W. Mills. The Free Press, USA.

Wellman M. (1958). "The concept of normal in medicine." *Canadian Medical Association Journal*, *79*(1), 43–44.

White, Paul. (2003). *Thomas Huxley: Making the "Man of Science."* Cambridge University Press, UK.

Wysocki, T. (2020). Normality: A two-faced concept. *Review of Philosophy and Psychology, 11*, 689-716.