Socially-inclusive foundations of statistics: An autoethnography

Dan J. Spitzner *

July 28, 2020

Abstract

Though distinct, the practice of statistics and the scientific worldview in which it is almost always practiced are so entwined that they are often taken to be the same. Feminist and decolonization scholars, among others, have noted that the scientific worldview is control-centered, irresponsible, and socially exclusionary. This chapter, written as an autoethnography, documents my efforts as a statistician in creating a new alliance of statistical practice with an alternative, socially-inclusive worldview. Aided by the concept of community elicitation, I argue that Bayesian data-analysis methods hold promise for interfacing with socially-inclusive research approaches that include a quantitative component, such as mixed methods research.

KEY WORDS: Social inclusion, inclusive research, mixed methods research, Bayesian methods, community elicitation, probability.

ABBREVIATED TITLE: Socially-inclusive statistics

1 Introduction

Robin Wall Kimmerer (2013, pp. 345-346), who identifies herself as a mother, scientist, decorated professor, and member of the Citizen Potawatomi Nation, frames science as a process of revelation through rational inquiry. She makes a crucial distinction between scientific *practice* and the scientific *worldview*: On scientific practice, she writes, "the practice of doing real science brings the questioner into an unparalleled intimacy with nature," and suggests many scientists find the pursuit to be humbling and deeply spiritual. Kimmerer contrasts this with the scientific worldview, writing that it involves a cultural interpretation of science and technology to reinforce "reductionist, materialist economic and political agendas." She furthermore refers to the scientific worldview

*Dan Spitzner is Associate Professor, Department of Statistics, University of Virginia, P. O. Box 400135, Charlottesville, VA 22904-4135, USA. Email: spitzner@virginia.edu. The author is grateful to the Quantitative Collaborative at the University of Virginia for funding a teaching release that supported this work.

as an "illusion of dominance and control, the separation of knowledge from responsibility." Once making the distinction between practice and worldview, Kimmerer puts forward her dream of a new alliance, "of a world guided by a lens of stories rooted in the revelations of science and framed with an indigenous worldview—stories in which matter and spirit are both given voice."

These distinctions map well to my own experience of practicing my discipline within a problematic, dominant worldview. My practice is that of a statistician, a practice that I define as making *meaning* out of numerical data. Its subject, abstract and mathematical, does not cultivate the sort of intimacy of which Kimmerer writes; nevertheless, statistical practice can activate a fascination with the creation of its idiosyncratic conceptual pictures, and a curiosity about what it is capable of describing. My discipline is furthermore steeped in a philosophical stance that draws heavily from logical-empiricist, positivist, and post-positivist perspectives. That is the dominant stance that surrounds current-day statistical practice, and that stance is deeply couched within the same scientific worldview that Kimmerer describes, accurately, I think, as ultimately control-centered, irresponsible, and, I will add, socially exclusionary. This chapter documents a portion of my journey to reconcile these distinctions within my discipline, and, in particular, to dream in my own way of a new alliance.

1.1 Motivations

I have chosen to write this chapter as an autoethnographic account (see Ellingson and Ellis, 2008, for guiding practices of this method) primarily because I wish to avoid any perceived mantle situating me as representative to my discipline, but also to tell the story of my thinking optimistically through ideas that are contrary to my formal training. To my knowledge, there is no movement, nascent or otherwise, within statistics to shake off the scientific worldview, and to find another one. Accordingly, my journey has been a solitary one, at least within my discipline.

On the other hand, I draw substantial motivation from outside of statistics. I am particularly motivated by the sustained and increasing activity within mixed methods research, an approach to research that applies qualitative and quantitative methods in dialogue with each other. Meixner and Hathcoat (2018, p. 3) write of historical and current-day debates within mixed methods research around the "philosophical legitimacy of mixing qualitative and quantitative data," and furthermore illustrate a practical imperative to reconcile philosophical incongruency without resorting to a substance-free version of pragmatism. In addition, Meixner and Hathcoat (2020, p. 3) situate mixed methods research as "inquiry that advances the process and aims of both inclusive research and social inclusion." When viewed in this context, a contribution made by this chapter is its identification of a data-analysis methodology that holds promise for resolving tensions between qualitative and quantitative stances, and for use in assembling a framework for mixed methods research. Through its use in mixed methods research, I put this methodology forward as potentially suitable for the promotion of inclusive research.

Specifically, I hold up Bayesian data-analysis methods for this purpose. The advantage of these methods is due to their being equipped for incorporating background knowledge into quantitative analysis. To clarify this capability, in Section 2.1, below, I offer a minimally-technical demonstration of the Bayesian approach, which is sufficient for illustrating its core aspects. Subsequent discussion offers a reimagining of Bayesian epistemological foundations that decenters the scientific worldview, and builds a new center around a conception of knowledge that is based in community.

As for my situation, I am a white, cis-gendered, heterosexual male, who spent all of my childhood in a rural town in the northeast United States. My partner and I raise a blended family of four children, one of whom has Down's syndrome, a topic that I touch on later in this article. As an undergraduate student, I studied mathematics and statistics at a public university, while participating in arts-related extracurricular activities. I have maintained interests in both the technical and creative, but have at times felt it straining to have my feet in two distinct cultural streams; I have long sought to achieve some genuine degree of integration.

My time studying at a high-ranking mathematical statistics doctoral program left me inspired by the mathematical depth of probability theory, and curious to understand why theories of statistical inference do not achieve similar depth. Later, as a young faculty academic, I became involved, and still am, in the community of Bayesian statisticians, having been attracted by that community's stewardship of statistical theory and principled development of statistical methodology. In time, as this chapter explains, it became clear to me that the answers to many of my questions regarding statistical theory have to do with the influence of the scientific worldview. To the extent that I was able, my attention moved away from applications in the hard sciences and engineering, toward those in the social sciences, but I continue to be dismayed by the utter dominance of positivist-type thinking in quantitative research. In this article, I share my wonderings on how quantitative methodology could potentially be informed by the perspectives of a socially-inclusive worldview, and my hope to foster interactions with those who would envision a place for quantitative methodology within such perspectives.

1.2 Socially-inclusive research and worldview

Socially-inclusive research and its associated worldview can refer to a variety of concepts, themes, and objectives. Nind (2014, p. 1) describes inclusive research as encompassing a range of research approaches, including "participatory, emancipatory, partnership and user-led research," among others. Koikkalainen (2011, p. 2) offers that the term "social inclusion" may refer synonymously to such notions as "unity, cohesion, civic engagement, togetherness, or bridging the gap between 'us' and 'the other'," or it can refer to focused actions that are designed to foster social integration of disadvantaged groups. Cappo and Verity (2014, p. 27) are guided by social-inclusion concepts in devising policy changes in South Australia, which aim to remove barriers to "secure housing, learning and employment, health and other services, social support and connections" that prevent people and communities from "living out their active citizenship." Thus, a socially-inclusive worldview involves a wide-ranging set of guiding concepts, as well as goals that would potentially be achieved through public policy (see also Liamputtong, 2020).

A prominent theme in social inclusion is the importance of community. Several authors are explicit on this point: Tua and Banerjee (2019, p. 109) succinctly define social inclusion as "the level of community participation and interpersonal relationships that people experience, as individuals and as groups." Simplican, Leader, Kosciulek, and Leahy (2015, p. 18) build these ideas into an ecological model for social inclusion, defining social inclusion as “the interaction between two major life domains: interpersonal relationships and community participation.”

A socially-inclusive worldview may also be understood by what it stands against. Allman (2013, p. 1) defines social inclusion in terms of contrasts, indicating that it is to “consider facets of social equality and inequality, social integration and stratification, social mobility as it relates to social inclusion and exclusion, and the functional contribution of the periphery relative to the social core.” His work also highlights sociological mechanisms of exclusion, including exclusion hierarchies, ostracism, and stigmatism. Koikkalainen (2011) reinforces this confrontational thread running through ideas around social inclusion, commenting that early use of the term was in opposition to approaches to research that demonstrate social exclusion.

The themes of community, opposition to social exclusion, and other guiding concepts of social inclusion are prominent within the realignment of statistical practice that I offer below, and crucial to distinguish it from traditional modes.

In Section 2, I cast Bayesian methodology as a community endeavor by situating it within Mercier and Sperber’s (2011, p. 60) “argumentative” theory of reasoning, which posits that “reasoning has evolved and persisted mainly because it makes human communication more effective and advantageous,” and furthermore strengthens the notion that even isolated individuals are woven into the fabric of community. This gives rise to a concept for describing the processes of inquiry that I call “community elicitation.” As I discuss in Section 2.3, traditional Bayesian reasoning is individualistic, and contrasts with community elicitation by its consistency with what Mercier and Sperber describe as the “classical view” that the main function of reasoning is to “enhance individual cognition” (p. 59) and to correct the mistakes of intuition.

In Section 3, I lay out my precise vision for realignment. There, I examine major themes of the scientific worldview that are particularly relevant to statistical practice; along the way I discuss contemporary currents in statistics that suggest an acceleration of exclusionary practices. I explore these aspects through the twin lens of decolonization of knowledge and feminist epistemology of science, drawing most heavily on Haraway (1988), Harding (2008), and de Sousa Santos (2018). I subsequently dismiss these themes in favor of those of a socially-inclusive worldview. Following this, in Section 4, I offer related perspectives on mathematical meaning and evaluating the processes of inquiry.

2 Bayesian methodology

In this section I aim to convey a basic sense of the Bayesian approach. Although statistical argumentation is necessarily mathematical, I keep the technical level of this section as low as possible, requiring of the reader only familiarity with the concept of an incidence rate (*e.g.*, of a disease) and an elementary understanding of histograms. The demonstrations I go through, below, highlight elements of statistical practice that open channels to a socially-inclusive worldview, which, moreover, contrast with elements of traditional Bayesian thinking. For conciseness, I focus in this section on quantitative analysis; my intent, however, is for the demonstrations to represent one portion of a broader mixed methods research context.

STEP 1: assemble a "prior" distribution.	The *prior distribution* is a probabilistic description of what is known and what is uncertain about the phenomenon under study.
STEP 2: develop a "data-generating" distribution.	The *data-generating distribution* is a probabilistic description of what is known and what is uncertain about how data are generated under each plausible circumstance determined in STEP 1.
STEP 3: use Bayes's Theorem to calculate the "posterior" distribution.	The *posterior distribution* is a probabilistic description of what is known and what is uncertain about the phenomenon under study, in the light of newly-measured data.

Table 1: *The three steps of the Bayesian approach.*

2.1 What is Bayesian methodology?

As I introduce Bayesian methodology, it is helpful to have a simple, uncontroversial example on hand for reference. The following simple scenario is one I offer to students in my introductory statistics classes:

Example 1. Suppose a certain health disorder is known from vast amounts of historical data to be prevalent in 5% of the general population. A health assessment is known to report incidences with 80% accuracy among patients where the disorder is present, and to report non-incidences with 90% accuracy among patients where the disorder is absent. Related terminology would characterize the assessment as having a "sensitivity" of 80%, a "false negative rate" of 20%, a "specificity" of 90%, a "false positive rate" of 10%. Suppose the health assessment is applied to a generic patient, and it reports an incidence of the disorder. Subsequently, a mathematical result known as Bayes's Theorem implies that there is a 30% probability the report is correct when taking the low prevalence of the disorder into account. □

With this example in mind, refer to the three core steps of a Bayesian approach laid out in Table 1. Translating to Example 1, the *prior distribution* is identified by the statement of 5% prevalence of the disorder in the general population. To be clear, the term "distribution" is probabilistic terminology referring to the weighting of probabilities across circumstances that are thought possible; so, *e.g.*, the prior distribution has 5% probability the disorder is present, and 95% that it is absent. It is common that a probability distribution would be displayed graphically as a histogram, as is done in the discussion around Example 2, below.

The data in Example 1 are the conclusions of the health assessment (incidence or non-incidence), and the *data-generating distribution* is identified by the statements of the assessment's 80% sensitivity and 90% specificity. This probability distribution is of a special type that is defined "conditionally" on other information; so, in actuality, it defines separate versions of the distribution across all circumstances covered by the prior distribution: in the example, one version of the distribution is conditional on the presence of the disorder (80% incidence to 20% non-incidence), and the other is conditional on its absence (10% incidence to 90% non-incidence).

The *posterior distribution* is sometimes said to represent an "update" of the prior, given new data; this interpretation reflects that Bayes's Theorem conceptualizes the posterior distribution as a combination of the prior and data-generating distributions. As with the data-generating distribution, the posterior distribution is defined "condi-

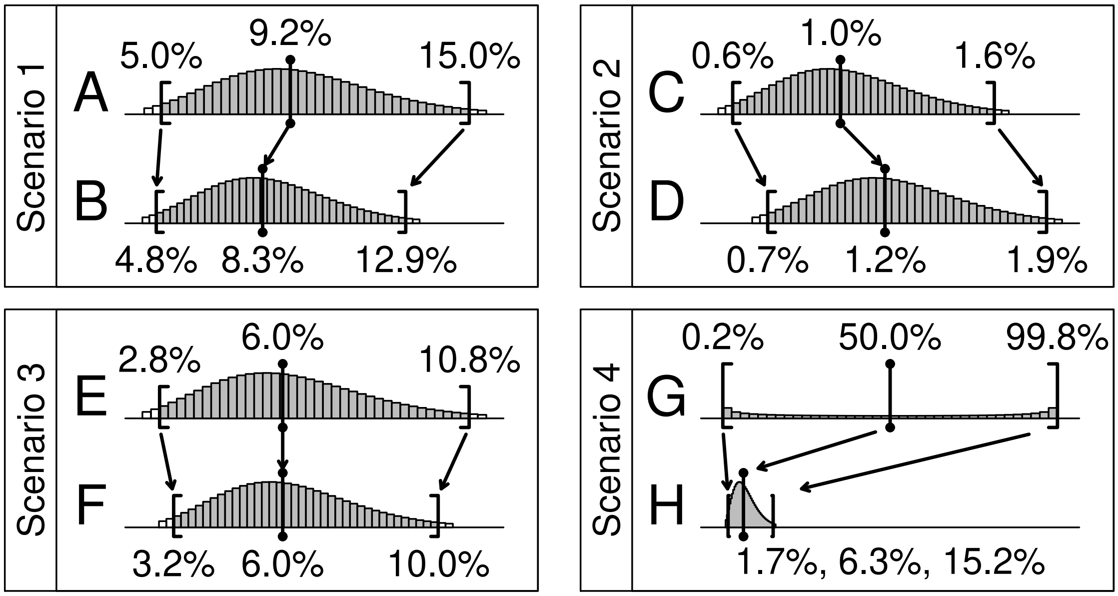

Figure 1: *Bayesian and Bayesian-like analyses in four scenarios. Histograms B, D, F, and H are summaries of the posterior distributions associated with the prior distributions respectively represented by Histograms A, C, E, and G, each determined using the 3-in-50 prevalence in newly recorded data.*

tionally" on other information, which, in this case, is the newly-measured data. In Example 1, the stated posterior distribution (30% incidence to 60% non-incidence) is calculated given the condition of the health assessment's report of an incidence; if the health assessment had reported a non-incidence, then a different posterior distribution would have been stated.

Example 1 is uncontroversial because it alludes to a situation wherein historical data on disorder prevalence and accuracy of the health assessment are richly available. The following is a less data-rich example, which highlights an approach to assembling a prior distribution that is important for later discussion.

Example 2. As a parent of a child with Down's syndrome, I am acquainted with a number of other children with Down's syndrome, including two young boys who also live with an Autism spectrum disorder (ASD). This has led me to become personally curious about the comorbidity of Down's syndrome and ASDs. In what follows I use my explorations of this topic to set up a simple, hypothetical example of Bayesian data-analysis in practice.

Reviewing literature, I observe that Moss and Howlin (2009) are motivated by the biomedical question of whether a genetic syndrome associated with intellectual disability, such as Down's syndrome, plays a causal role or is an additional risk marker for ASDs. A low rate of ASDs among individuals with Down's syndrome, at or below 1%, which is the rate of ASDs in the general population, is consistent with the causal explanation implied in a theory that the affectionate personality profile associated with Down's syndrome is protective against autistic-type behaviors. Startin *et al.* (2020) state an interest in variation among the patterns of comorbidity over lifespan, for the purpose of informing "clinical services, individuals, and their families and carers." These two studies, and others, report comorbidity rates far higher than 1%.

However, pinpointing the comorbidity rate is challenging, due partly to variations in the screening and diagnostic criteria applied across studies. Nærland, Bakke, Storvik, Warner, and Howlin (2017) cite a range of values, between 5% to 39% comorbidity, which roughly agrees with the range of values reported in Moss and Howlin (2009). A data set examined in Startin *et al.* (2020) suggest distinct comorbidity rates across three different age groups, taking the values 5.7% for older children, 13.5% for younger adults, and 2.5% for older adults.

With this knowledge on hand, now consider a hypothetical new study carried out using the three steps of the Bayesian approach. For concreteness, let us say our hypothetical researchers collect new information on 50 individuals with Down's syndrome, under a sound study design, and subsequently determine, without errors, an incidence of ASD for 3 of these individuals. Hence, the comorbidity rate in just the newly-measured data is 6% prevalence.

In Step 1 of the analysis, as defined in Table 1, the researchers compare diagnostic and screening criteria between their study and those cited above, as well as the age distribution of the population being examined. Based on these comparisons, they determine that the range of comorbidity rates between 5% and 15% describes a sensible combination of the available information, which puts most weight on the younger age groups reported in Startin *et al.* (2020), while not ignoring the remainder of available information. This "background" range of comorbidity rates is interpreted to mark the middle 95% of a prior probability distribution whose complete form is derived through the use of conventional mathematical tools. Specifically, the "beta-binomial" probability model, which has been developed for use with incidence data, is applied to provide a suitable mathematical form for both the prior distribution of Step 1 and data-generating distribution of Step 2. (A more sophisticated analysis would modify the beta-binomial probability model to take into account the ages of the fifty individuals actually measured, but that would needlessly complicate the example, which is intended primarily for illustration.) Readers interested in technical aspects of statistical modeling and Bayesian analysis may find the references provided in Section 5 to be helpful.

One result of working with the beta-binomial probability model is a more extensive set of summary values associated with the prior distribution. Through mathematical tailoring of this model so that its middle 95% of prior probability matches the 5% to 15% range, a median comorbidity rate of 9.2% emerges. Histogram A, in the top-left panel of Figure 1, displays this prior distribution in graphical form, where the middle 95% of probability is marked by square brackets and the median is marked by vertical dumbells. In Step 3, application of Bayes's Theorem "updates" the prior distribution in light of having measured a 3-in-50 prevalence rate in new data, to produce a posterior distribution with median 8.3% comorbidity and middle 95% of probability within the range 4.8% to 12.9% comorbidity. This posterior distribution is displayed as Histogram B in the top-left panel of Figure 1. Our hypothetical researchers list the posterior summary values in a statistical report of their findings: the median value, 8.3%, gives a single "best estimate" of comorbidity, while the middle 95% posterior values, 4.8% to 12.9%, offers a "most credible" range of comorbidity rates. □

The statistical analysis described in Example 2 is represented in the panel of Figure 1 labelled Scenario 1. The three remaining panels illustrate variations of the analysis, all carried out on the same data but using different prior distributions. The variations are intended to set up discussion of nuance and aspects of controversy around the Bayesian approach. Scenario 2 depicts a Bayesian analysis derived from a prior distribution that describes background knowledge assembled from outdated literature. Its prior distribution, summarized in Histogram C, reflects since-debunked claims that ASDs are no more prevalent among individuals with Down's syndrome than in the general population. Scenarios 3 and 4 illustrate two strategies for implementing a mimic of Bayesian analysis that *ignores* background information. Histogram E is a summary of a prior distribution with median comorbidity set to 6%, deliberately matching the comorbidity rate of the newly-measured data; the width of its middle 95% of probability set to 8%, which these hypothetical researchers determine is consistent with the maximum extent of uncertainty in the comorbidity rate that they are willing to accept. Histogram G is a summary of a prior distribution that is deduced as an abstract representation of a situation of ignorance about the comorbidity rate. It is determined by a line of mathematical argumentation known as "Jeffreys's general rule" (see, *e.g.* Berger, 2006, sec. 2.2), one of many that have been put forward by statisticians who pursue a theory of Bayesian methodology that would give preference to the practice of setting the prior distribution to an analytically-derived configuration over collecting background knowledge. Far from serving as, perhaps, a remedy in extraordinary cases where responsible inquiry would face insurmountable hurdles to the collection of background information, the winds of exclusion have made Scenario 4's abstractions into the dominant mode of Bayesian practice (*cf.* Mayo, 2018, sec. 6.3).

2.2 Argumentation and community elicitation

Though each four of the scenarios indicated in Figure 1 resemble Bayesian analysis, some of these scenarios do not meet conventionally-accepted standards of sound inquiry. For instance, Scenario 3 is clearly problematic for ignoring actual background knowledge, and instead defining prior uncertainty to achieve a desired outcome. Scenario 2 is problematic as well, for using outdated background knowledge. Scenario 4 is controversial for assembling a prior distribution abstractly, in ignorance of the research context. The analysis of Scenario 1 might reach the other side of peer-review and become published, at which point it could potentially inspire additional discussion across a broad community.

The process alluded to in these hypothetical criticisms is one of *argument* and *counterargument*. Peer-reviewers would expect the authors to provide arguments in support of the decision they made in formulating their analysis. Otherwise, *e.g.*, the flaws of even Scenario 3 could not be detected. This and the broader process of argument and counterargument, where authors, by disseminating their work, make arguments to their communities and other members respond with counterarguments, is influential to the development of a community's conventions of good practice and preferred lines of argumentation.

The general process of specifying a prior probability distribution is commonly referred to as "elicitation," which, in simpler language, roughly translates to "bringing forth." When a Bayesian statistician brings forth a prior probability distribution, they have the conventions and preferences of their community in mind, as is necessary to prepare an argument for the validity of their insights. In this way, meaningful elicitation, because it takes place within the argumentative dynamic of a community, is a social process. Accordingly, I use the term *community elicitation* to refer to this social interpretation of how a prior probability distribution is determined. At times in what follows, especially in Section 4.2, I also use the term generically, in reference to the community-centered process of bringing forth general knowledge.

Existing literature contains occasional clear examples of a prior distribution's community elicitation. The demographic study described in Daponte, Kadane, and Wolfson (1997) use Bayesian ideas in an attempt to forecast what the Iraqi Kurdish population from 1977-1990 would have looked like had the repression of the Kurds since 1977 not occurred. In that analysis, information is expressed using probability distributions; it is compiled from data gathered from various surveys, censuses, reports, and other data sources, and rigorously assembled from fertility, mortality, and migration rates, specific to time, age, and rural/urban status.

This study has served for me as an inspirational anchor as I have strived to clarify my vision of the community elicitation concept. Importantly, Daponte *et al.* (1997, p. 1256) remark that "Making one's beliefs explicit using probability distributions allows other demographers to observe exactly how one views the sources of uncertainty in the phenomenon." They state that this is important to "enhance communication among demographers," as it allows that "others can then know on what they agree or disagree," and because the "reasons given for particular probability distributions can be an important source of insight." While the authors' expressed interest in communication points to their valuing of healthy community, the latter assertions speak to criteria for assessing the quality of the knowledge that is learned; *i.e.*, the criteria are epistemic in character.

I regard two specific criteria as especially helpful for guiding discussions of the quality of the knowledge and general effectiveness of community elicitation: *accessibility*, which is helpful for addressing the mechanisms that allow community elicitation to function, and *thoroughness*, which is helpful for assessing the richness of the arguments produced. For example, to help a reader to "know on what they agree or disagree," by grounding comments in familiar conventions, improves access; offering underlying reasons for a choice made during inquiry adds an additional layer to knowledge, making it more thorough. Accessibility might also include such considerations as parsimony, to encourage thoughtful crafting of descriptions for ease of communication and digestibility, or the use of devices such as varied repetition, which would create multiple access points that reach across the perspectives of a diverse audience. Thoroughness might include such considerations as precision, to encourage thoughtful attention to any gaps in description. Whatever a community's formulation of accessibility and thoroughness, I would hope that neither criterion would be equated with such simplistic considerations as the *amount* or *intensity* of argumentation, for those do not take into account that community elicitation can be degraded, or frozen, by argumentation that is disingenuous, cynical, or that otherwise adds to dialectical noise, whether deliberately or non-deliberately introduced. A special challenge for communities that aspire for social inclusion as well as effective community elicitation is to develop conventions and practices for filtering out noise while not excluding voices that would add to knowledge.

2.3 Bayesian individualism

The reader may find my above description of community elicitation curious for its familiarity. Is not engagement with one's intellectual community what researchers already do? Is not careful discussion of background information already an expectation of sound research? The answer to each of these questions is, to a large extent, resoundingly yes. However, careful reading of the statistics literature on elicitation can engender a sense that elicitation is supposed to be about something else.

A strong current in this literature is that elicitation has to do with the judgments of individual experts. As one example where individual expert-judgement could be warranted within its context, Goldstein (2006, p. 404) describes a software-development application in which Bayesian methods are used to implement a set of testing procedures for a complex computer system during its development. In that example, background knowledge is collected from software testers who are already familiar with the software that is to be tested, and whose expertise "lies in the informed nature of the prior beliefs that they hold." In other words, the source of information for compiling any prior probability distribution is the beliefs of individuals. The use of expert judgement, in this case, is seemingly warranted from the economic context, where evaluation of "success" in developing software is largely measured in terms of financial utility. The use of individual experts, and quantitative methodology itself, is less a means to gain knowledge than it is to implement a strategy for the company's success.

The mode of inquiry alluded to in Goldstein's example is pragmatic, a stance that I carefully evaluate in Section 4.2. Among its most disconcerting aspects is its capacity to reduce human actors to abstract containers, to be filled with whatever theory is needed to justify a strategy's implementation. Traditional Bayesian inference, whose foundations draw heavily on economic abstractions, makes just such a move: one influential Bayesian insists upon a vision of the individual expert as a "rational man," an entity that "always acts as if he had a probability distribution over the states of the world and a utility function describing the utility to him of each decision in each state; he chooses his decision by maximising the expected utility" (Lindley, 1958, p. 192). This quote is in reference to a famous normative theory for rational decision-making developed by Savage (1954). Many Bayesians recognize that the "rational man" ideal and related normative theories are, at best, models (*cf.* Howson and Urbach, 1989, sec. 3.c). Yet, they remain very influential. O'Hagan *et al.* (2006, p. 21) assert that these normative theories imply that "probability is the uniquely scientific way to represent uncertainty," which indirectly asserts that Bayesian ideas are suitable not only as guidelines for economic strategy, but as a foundation for scientific inquiry in general.

Adding to the strong sense in traditional Bayesian thinking that knowledge is not a phenomenon of community, but one of individuals, O'Hagan *et al.* (2006, p. 97) use the following language, which is common in the Bayesian literature: "To express uncertainty about an unknown quantity, we need to elicit the expert's probability distribution for that quantity." The suggestion, here, is that knowledge is brought forth by *extracting* it from a person's mind. Accordingly, many Bayesian methodologists draw lessons from psychological experimentation on human judgement of uncertainty. For instance, O'Hagan *et al.* (2006) define a goal of reducing "bias" in elicitation, whose motivation derives from an influential series of such experiments. Building on this individualist perspective, a substantial body of literature is concerned with the combining of knowledge that is elicited from multiple experts in a process sometimes referred to as "belief aggregation" (which I compare with community elicitation in Section 4.2).

2.4 Traditional mistrust of background knowledge

Often accompanying traditional Bayesianism's imperative to elicit and make use of a prior distribution is a tendency to not trust the knowledge that it describes. This is expressed in mathematical results that describe a phenomenon known as "stable estimation" (see Edwards, Lindman, and Savage, 1963), which asserts that the influence of the prior distribution on statistical results becomes negligible in the presence of a suitably large amount of data. Among studies that take a Bayesian approach, it is not infrequent to find a stable-estimation argument put forward as partial support for the trustworthiness of the study's statistical reporting. Implied is an underlying message that the prior distribution does not *really* matter, provided there is new data at hand.

To illustrate the stable-estimation phenomenon, let us reconsider Scenarios 1 and 2 of Figure 1. Stable estimation implies that if the same prevalence rate, 6% comorbidity, is recorded in a larger sample, then, between the scenarios, the two posterior distributions would come closer to agreement, despite the two corresponding prior distributions being in quite different configurations. For instance, if 6% prevalence were recorded in a sample of 35,000 individuals (*i.e.*, 2,100-in-35,000 rather than 3-in-50 prevalence) then both posterior medians would be near 6% (6.0% for Scenario 1 and 5.8% for Scenario 2) and both middle 95% probability ranges would have widths around one-half of a percentage point (between 5.8% and 6.3% for Scenario 1 and 5.6% and 6.1% for Scenario 2). If the sample was even larger, so too would the posterior distributions be more similar.

An interesting, subtle reframing of stable estimation has it describing a process by which the knowledge of two Bayesian statisticians eventually merge, once a sufficient amount of data becomes available. It is characterized in Diaconis and Freedman (1986, sec. 3) as a process of "intersubjective agreement." This version is an important example of an appeal to eventual *agreement* as an epistemological basis for meaning in data analysis, and a step for traditional Bayesianism in the direction of viewing quantitative inquiry as having a social component. In this chapter, however, I mean to primarily emphasize the importance of *disagreement* in deepening knowledge.

One facet of disagreement that is widely discussed among Bayesian practitioners is disagreement between a specified prior distribution and newly measured data. For instance, disagreement of this sort is alluded to in Scenario 2 of Figure 1, where the recorded prevalence rate of 6% is far from not only the prior median of 1%, but *any* comorbidity rate that has non-negligible prior probability. Such broad deviation would perhaps motivate our hypothetical researchers to return to the literature in search of updated background information, and possibly revise their prior one closer to that of Scenario 1. Disagreement of this sort has given rise to statistical methodology known as "model checking" (*cf.* Gelman *et al.*, 2013, ch. 6), the use of which is typically recommended as a routine step of good data-analysis practice.

Less widely discussed, however, is the beneficial impact of potential disagreement among the stakeholders of a study and the communities they form. Notice in Example 2 the diversity in *positionality* of the relevant background literature: the conceptual framing of questions asked by Moss and Howlin (2009) is biomedical, while that of Startin *et al.* (2020) is of clinical care. The differences in these studies' positionality give rise to questions about how the two positions interface in society, and shifts perspective to a deeper level of meaning around the topic under study, one where quantitative reasoning is rendered less relevant because of its abstraction. In other words, shining light on the differences between communities opens a path for a pursuit of knowledge that speaks more directly to human and societal meaning. With this in mind, I like to imagine a reversal of the stable-estimation argument, which could be expressed as follows: the relevance of statistical results becomes negligible when suitable attention is put to uncovering layers of perspective in background knowledge. Implied is an underlying message that data-gathering and analysis is scarcely more than a therapeutic exercise, provided that a sufficiently strong connection to community is maintained.

The ground between a stable-estimation argument and its reverse is such that neither data-gathering and analysis nor differences in perspective are necessarily meaningless. In this middle ground, mixed methods research is appealing for offering a framework within which to synthesize and move forward with all pieces uncovered while collecting background knowledge. The use of Bayesian methods, because of the crucial role it assigns to prior information, is appealing for creating a seamless integration of quantitative elements within that framework.

3 Realigning statistical practice

The "scientific worldview" might more precisely be labelled as "Western science" or "the epistemologies of the North;" it refers to a collection of generally shared tenets of associated philosophies. De Sousa Santos (2018, p. 6) and Warren (2015, sec. 1.2), the former drawing on the latter, offer lists of these tenets, which I combine, paraphrase, and condense as follows: (*i.*) A commitment to rationalism, and its absolute priority as rigorous knowledge. (*ii.*) A distinction between the knower and the known, wherein the knower as a rational being. (*iii.*) A belief in fundamental dualisms, such as objectivity *versus* subjectivity, and absolutism *versus* relativism, together with a requirement that objectivity is socially and politically neutral. (*iv.*) Truth, conceived of as the representation of reality. (*v.*) Universalism as a criterion for confirming, or otherwise assessing, a state of reality, referring to a condition that does not depend on any specific social, cultural, or political context. In this section, I briefly reexamine and propose alternatives to three themes among these tenets that are especially relevant to statistical practice: objectivity, confirmation, and universalism.

3.1 Objectivity

On the theme of objectivity, and its requirement of social and political neutrality, de Sousa Santos (2018, p. 40) describes its exclusionary effects as follows: "Neutrality is an ideological device in a society divided between oppressors and oppressed. In such a society, to remain neutral amounts to being on the side of the powerful and the oppressors." Under division by the scientific worldview, the rhetorical action of a claim of objectivity in argumentation is to dare the reader to offer a counterargument, where they risk stigmatization as one less capable of dispassionate assessment, or perhaps general loss of legitimacy with the community. It is an attempt to shut down argumentation, *i.e.*, as Haraway (1988) characterizes it, a "power move." Similarly, Harding (2008, ch. 1, sec. 2, para. 2) writes of exclusionary effects of the scientific worldview within science itself, noting that the "very standards of objectivity" have been incompetent to identify systemic gender bias in the sciences. She furthermore specifies that feminist criticisms of the scientific worldview "focus not on the prejudices of individuals (unpleasant as those can be for their targets), but rather on the assumptions, practices, and cultures of institutions, and on prevailing philosophies of science."

My own discipline of statistics has been rattled by infuriating and heartbreaking instances of power misuse. In December 2017, Kristian Lum, a young academic who was known for her methodological work on biases in predictive policing software (see Lum and Isaac, 2016), posted a blog entry (Lum, 2017) detailing her experience of being sexually harassed at a large international meeting of Bayesian statisticians in 2010, one that I had, in fact, attended. This revelation exposed a strain of misogynistic culture within the Bayesian community, which, in an article in *The Guardian* (Levin, 2017), was connected to larger patterns of sexual misconduct across statistics and data science, wherein "serial harassers rarely face consequences." What Harding's comments imply is that effective change from these destructive patterns toward inclusion and equity within Bayesian and scientific communities requires reexamination of foundational philosophies, and reconsideration of the methodologies they deem legitimate; both are efforts that go beyond intensified recruitment and initiatives for cultural change, which is where professional institutional bodies tend to stop.

In my dream of a realignment of statistical practice with a socially-inclusive worldview, I dismiss objectivity's rhetoric of neutral declaration in favor of a conception of objectivity as an earnest, vulnerable call to one's community. De Sousa Santos (2018, p. 44) writes of alternative conceptions of objectivity, wherein "objectivity is always intersubjectivity, indeed, self-conscious intersubjectivity," hence knowledge is cocreated. It is cocreated by all participants in community elicitation, which, as Nind (2014, p. 5) reminds us, could extend beyond those with "researcher-only expertise" to "people being researched." Under a socially-inclusive worldview, a call for objectivity is to be recognized as no more than an author's attempt to persuasively anchor an argument to community conventions. For instance, in Example 2, our hypothetical researchers might argue that it is "objective" to incorporate past statistical reporting of comorbidity rates because it is consistent with an agreed-upon role for empirical measurement. Or, in a different scenario, an author might attempt to anchor their arguments to a theoretical framework, or mode of thinking, that has been ascribed validity within the community. Aspirationally, the environment for elicitation would be one in which an author claiming objectivity would always be clear about their positionality, about their own understanding of the conventions invoked, and would be open to the possibility that not all members of the community are satisfied with them.

3.2 Confirmation

On the theme of confirmation, statisticians tend to be especially conscious of its objectives and specific philosophies associated with them, which they are often hired to apply. Confirmatory objectives pose a particular challenge for socially-inclusive practice, since confirmation is itself an exclusionary assertion, or, at any rate, its conclusions risk being taken as justification for rejecting responsibility to some of inquiry's stakeholders. Consider, for instance, Karl Popper's confirmation theory of "falsification," which, in coordination with a "frequentist" interpretation of probability (see Mayo, 2005), underlies the most widely taught statistical methodology for confirmation. In Popper (1962, p. 39), he is careful to point out that his interest is not in criteria for meaning, but in "drawing a line (as well as this can be done) between the statements, or systems of statements, of the empirical sciences, and all other statements.” That is, his purpose for falsification is to create a separate category for a system of knowing—he labels it “empirical science”—to the exclusion of all others. A distinct layer of interpretation positions this outlying category as *rising above* (*i.e.*, more meaningful than) all others; subsequently, communities that hold reverently to empirical science have a tool with which to position *themselves* as rising above all others, and to furthermore position other communities’ ways of knowing, if not the communities themselves, as sub-human and disposable.

Traditional Bayesian thinkers avoid Popper’s “demarcation” approach, as well as issues of meaning, in favor of treating confirmation as a problem of decision-making. As I have already discussed in Section 2.3, the existing, disappointing solution is the model of a “rational man” who makes decisions by “maximising the expected utility,” with its peculiar focus on individualism. Yet, confirmation, even from the perspective of decision-making, remains problematic to socially-inclusive practice because of its basic binary, specifically because of a tendency for the decision made to be used as reason to avoid responsibility to those it excludes, or to issues that are lost by the yes-no framing. For example, in legal justice systems this plays out as inattention to the rights of an offender, as documented in Moore and Mitchell (2011), and resistance to the adoption of restorative practices that carefully attend to, as Polite (2018) describes, an “offense’s roots and resulting wounds” (p. 103).

In my dream of a realignment of statistical practice with a socially-inclusive worldview, I dismiss the binary of confirmation in favor of accessible and thorough description of what is assumed and what is learned. As I discussed in Section 2.2, Bayesian methodology, when practiced in the context of community elicitation, is appealing in decision-making for its transparency and extensive capacity to identify the assumptions of the decision-maker. See also Spitzner (2020) for discussion of methodology related to this particular aspect of Bayesian decision-making.

3.3 Universalism

Universalism underlies a frequent goal of statisticians, which is to generalize the results of a data analysis. When teaching beginning students, I often cover a traditional “sampling” model for generalization by which patterns noted in sample data are to generalize to the population from which the sample is randomly drawn. Further abstraction of this model leads to ideas supporting universal generalization to “truth” or “states of nature.”

Among criticisms of universalism, I find those expressed in Haraway (1988) to be particularly helpful. Haraway (1988, p. 581) calls out a version of universalism that manifests not through conceptual models but through vision. She writes of endless technological enhancements to vision (*e.g.* in the form of “artificial intelligence-linked graphic manipulation systems”) that render it a feast of “unregulated gluttony,” putting the myth of “seeing everything from nowhere” into ordinary practice, a performative feat that she calls a “god trick.” The result is a “false vision promising transcendence of all limits and responsibility” (p. 583). This criticism is particularly relevant today, to data science, an impressively fast-growing, newly-branded discipline that is adjacent to statistics. (See Mayer-Schönberger and Cukier, 2013, for an enthusiastic description.) Its methods, integrated with sophisticated computational architectures, are the backbone of well-known big-technology companies, and of the "data-driven" economy in general. Their flexibility in visualizing complex patterns in data, and capabilities for application to massive data sets and high-volume analysis, drive a sharp intensification of the performative aspect of data-analysis that is concerned with illusions of universalism and control.

In my dream of a realignment of statistical practice with a socially-inclusive worldview, I dismiss universalism's dazzling illusions in favor of connection and translation of knowledge. The latter is a reflection of Haraway's assertion of feminists' need of an "earth-wide network of connections, including the ability partially to translate knowledges among very different—and power-differentiated—communities" (pp. 579-580). As a stance for statistical methodology, this need is unusual for being in line with objectives that are usually taken on by qualitative methods, where techniques such as narrative and storytelling can be effective at creating connection by activating the reader's sensitivity to themes that are revealed during the course of research. Consider, *e.g.*, the context of auto-ethnography, wherein Ellis (2004, p. 195) offers that the use of storytelling "brings 'felt' news from one world to another." When aligned with a socially-inclusive worldview, these are techniques that statistical methodology stands to become part of, especially within a context of mixed methods research. The performative aspect of data-analysis would bypass chest-thumping displays in favor of "dialogical performance," a stance that Conquergood (1985, p. 9) describes as having an aim to "bring the self and other together" in a dialogue that "resists conclusions" and is committed to continued dialogue. When aligned in this way, the uniquely Bayesian operation of updating prior to posterior knowledge would be less a technical calculation than a story of transformation.

4 Epistemological grounding

Although Section 3 lays out my vision for realignment of statistical practice with a socially-inclusive worldview, I worry that my specific proposal for the use of Bayesian methods leaves unresolved issues that I expect some philosophically-minded readers, and some of my Bayesian colleagues, would be curious to have addressed. Given my expressed dissatisfaction with traditional Bayesian epistemology as grounding for socially-inclusive practice, it would be fair to wonder what sort of epistemology I *would* find satisfactory. Similarly, even if I have persuaded my reader that elicitation is a social process, it would be fair to wonder how I came to settle upon accessibility and thoroughness as guiding criteria for pondering questions of how community elicitation is to be carried out effectively. Attention to these wonderings takes this chapter to the deeper places among philosophical tensions between qualitative and quantitative stances; yet, these are crucial places for socially-inclusive research to visit, in order that the practitioner may produce well-grounded arguments. I myself travel to these places as much out of inspiration by the courage of scholars such as Haraway, Harding, and de Sousa Santos in calling out exclusionary knowledge systems; I seek to replace such systems *at their roots* with a set of responsible guiding principles.

4.1 The meaning of probability in community elicitation

Because Bayesian methods are rooted in the concept of probability, epistemologically grounding these methods is a matter of offering a meaning for probability that is

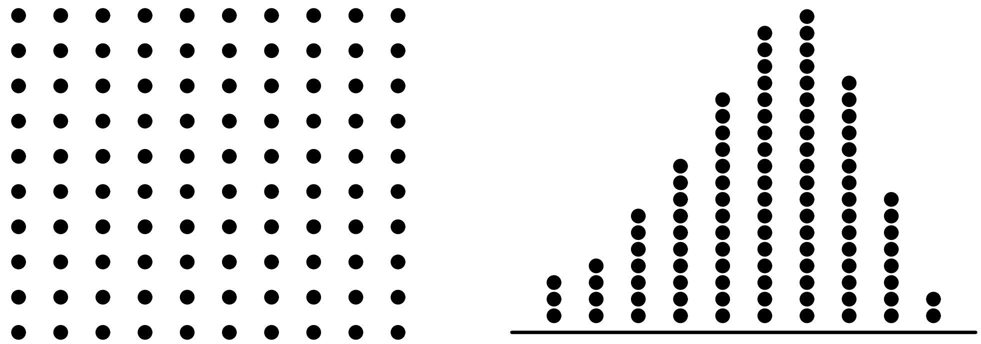

Figure 2: *The roles of "100" (left) and of probability (right).*

consistent with a community-based conception of elicitation. Epstein (2009, p. 11) identifies three interpretations of probability, all touched on in this chapter, that have been important to its mathematical development: sampling theory, frequentist theory, and traditional Bayesian epistemology. None of these quite capture the manner in which I envision probability would be understood in community elicitation, where it would serve entirely as a device for description, and would be free of allusions to idealized processes or entities that would ascribe its meaning to some equally idealized picture of reality. To make sense of probability, I turn to language-based social theories of knowledge, which are a common source of philosophical grounding outside of the scientific worldview. Specifically, I find it helpful to draw on the later work of Ludwig Wittgenstein, as interpreted in Bloor (1983), for its attention to theories of *mathematical* knowledge.

Bloor (1983) describes the basic building blocks of Wittgenstein's ideas about language and meaning, which include the concepts of "forms of life" and "language games." The former is, roughly, the way that a community lives, and the latter is, roughly, the ways in which a community interacts and communicates. These aspects of a community allow its members to make sense of their shared world. In discussing mathematical truth, Bloor writes of Wittgenstein's rejection of the Platonist view that "mathematical results are discoveries about a special realm of objects that exist prior to our knowledge of them" (pp. 83-84). Wittgenstein instead stresses the importance of *usage* in meaning. Reality, of an "*entirely* different sort" than Platonist reality, is that certain techniques are useful to a community; "it's an ethnological fact," writes Wittgenstein, "it's something to do with the way we live" (as cited in Bloor, p. 94).

On the meaning of numbers, Wittgenstein offers a variation on demonstrations put forward by the nineteenth-century philosopher J.S. Mill that involves manipulating groups of pebbles. In Wittgenstein's version he stacks one-hundred marbles in rows of ten, as illustrated in the left panel of Figure 2, and offers that the arrangement demonstrates "the *role* which '100' plays in our calculating system" (as cited in Bloor, p. 94). Bloor writes that "the emergence of the mathematical out of the physical occurs when the empirical manipulations are put to a certain use; when they become taken up in a certain language-game; when they become part of a certain technique; and when they become subject to certain conventions and norms" (p. 94). In other words, the stack of marble-rows is not a representation, but a tool in the language that has been developed for discussing calculation; the use of that tool within language is what gives the concept of one-hundred meaning.

Probability may be understood in precisely this way: its meaning—that is, the meaning of the probability histograms displayed in Figure 1 and their mathematical generalizations that define probability theory—is entirely determined by how it is used within community elicitation, as a device of the language that is used to describe uncertainty. In contemplating this idea, I like to imagine an alternative history where Wittgenstein explains it to my statistics class: he arranges one-hundred marbles into a histogram, as in the right panel of Figure 2, and offers that they demonstrate "the *role* which probability plays in statistical practice."

4.2 Evaluating community elicitation

As for evaluating the effectiveness of community elicitation, a major motivation for my choice of accessibility and thoroughness as evaluative criteria is my personal experience of analyzing data: when I become absorbed in statistical practice, and my curiosity is aroused, it *feels* correct to say that these criteria guide me through the production of an analysis. Nevertheless, I also came to this choice through a process of whittling down a collection of other possibilities. In this section, I examine several alternative criteria, and offer my thoughts on how I came to find them unsatisfying. Note that, in what follows, I expand my usage of the term community elicitation to cover not just elicitation of a prior distribution, but of knowledge in general.

My dismissal of universal generalizability, renders concepts such as "truth" or "states of nature" irrelevant as building blocks for evaluative criteria. A seemingly related concept, one that is social in nature, is "agreement," which is the basis of one version of the stable-estimation argument I describe in Section 2.4. However, I am persuaded by an argument in Goldman (1999, sec. 3.1.A) whose conclusion is that agreement is insufficient for evaluating the processes of inquiry. His reasoning is that it is easy to come up with examples of agreement situations that are clearly ineffective, such as consensus achieved by totalitarian practices. My view is that evaluation would instead take into account *tensions* between agreement and disagreement, which give rise to variation in usable knowledge systems across a diversity of communities. A task of epistemology is then to describe the "truths" about which a community agrees, and the "dimensions" along which distinct communities disagree, a task that accessibility and thoroughness are suitable to evaluate.

My early thinking on this topic focused instead on criteria based on trust and their overlap with pragmatism, by which I mean the view that inquiry is valued for the utility of its outcomes. At one point I contemplated the suitability of giving preference to practices that foster public trust in community elicitation itself, such as practices that encourage transparency. However, I found this criterion problematic for its close connection to democracy, and the nuanced role of public trust within that political system. For instance, the arguments in Lenard (2008, p. 313) make clear that "alongside trust we need to adopt a kind of vigilance, an attitude that is motivated by mistrust, in order to maintain a healthy democracy." Yet, by building in such nuance to a criterion based on trust, it seems to me that the result would be instrumental to a pragmatic preference for practices that enable one's community to thrive. Why not just accept a pragmatist criterion?

If one were to take a pragmatist's view, questions arise as to how to take into account the multiplicity of communities, and the possibility of individual membership in multiple communities. Those who study social aggregation, which includes a sizable number of traditional Bayesians, take on similar questions, and examine the potential for resolution using such rational strategies as voting or compromising; but, these typically lead to logical paradoxes: For example, Garthwaite, Kadane, and O'Hagan (2005) conclude that voting often leads to intransitivity (*e.g.*, $A > B > C > A$), and non-trivial compromises tend to violate the "Pareto Principle," which is the requirement that "if each member of the group prefers A to B, then the compromise cannot prefer B" (p. 697). It seems that rational aggregation strategies suffer from their own insistence on rationality.

A promising alternative to aggregation is to look through the lens of argumentation, expanded to cover the processes of argumentation across constituent communities. De Sousa Santos (2018, p. 40) comments on how this plays out: To start, he is blunt in connecting trust to pragmatism, writing that "trust is always referred to in relation to the objectives to be reached using the knowledge that is trusted." He then defines two categories of objectives for knowledge, "knowledge as regulation" which is concerned with moving from chaos to order, and "knowledge as emancipation," which is concerned with moving from exclusion to solidarity. Using this language, he identifies two problems with the landscape it describes: The first is that under colonialist and imperialist agendas, knowledge as emancipation is violently suppressed. The second is that the primacy of knowledge as regulation has tended to be "pushed to the utmost," to the point of cannibalizing knowledge as emancipation by reconceptualizing solidarity as chaos and exclusion as order.

Drawing on these ideas to account for pragmatism in argumentation across communities, I find accessibility and thoroughness to be useful for evaluating knowledge conceived of either as regulation or as emancipation; but, it is significant that their use within the latter is motivated directly rather than, in large part, through the baggage of the scientific worldview. Ultimately, I find that accessibility and thoroughness have the double use of evaluating both description itself, as standalone concepts in their sense from Section 2.2, and of evaluating the pragmatic use of the *power* of description, most directly in struggles for achieving social inclusion against violent and cannibalizing pressures of exclusion.

5 Conclusions and Future Directions

As I write the conclusion to this chapter, I recall the path I describe in Section 1.1, and its accompanying oppressive feeling that led me to move from domain to domain in an attempt to escape the scientific worldview, but never succeeding. Having now visualized a dream of realignment, I now feel relief, liberation, and hopefulness at having potentially opened a viable way to operate within my discipline as an integrated self. I feel released from a sense that practicing statistics necessarily puts me under the weight of a stifling machine. In what follows, I offer a few comments on a new dream: the way forward.

As Meixner and Hathcoat (2018) discuss, ongoing debate around mixed methods research struggles with contradictions between the foundations of qualitative and traditional quantitative methodology. They write of how "philosophical agnosticism," a potential consequence of a researcher's inattention to their own "mental model" toward inquiry, can "endanger various facets of the research design, most especially the interpretation phase" (p. 11). Summarizing the current state of debate, they write that some scholars maintain that quantitative and qualitative research are inherently incompatible and cannot be integrated; others assert they can be integrated even while respecting distinctions between their differing philosophical underpinnings; and, still others allow integration under cover of a pragmatist position, though one that has tended to be unhelpful for informing substantial decisions during the course of research. Hathcoat and Meixner (2017) offer further discussion of these issues, and contribute a "conditional incompatibility thesis" to assist in resolving them. The realignment of statistical practice that I have described in this chapter bypasses this struggle entirely by replacing traditional quantitative methodology with a version whose foundations are radically altered for compatibility with qualitative perspectives.

On the other hand, "radical" is a relative term: some may find my imaginings insufficiently radical, still carrying an imprint of the scientific worldview. De Sousa Santos (2018, p. 5) would perhaps characterize my realignment as an "internal criticism." Certainly, however, the way forward is not closed off to the possibility of alternative realignments.

One troublesome practical issue that does not appear to have a quick resolution is a lack of materials for learning about Bayesian methodology within the framework of this chapter. Nearly all textbooks on the topic of Bayesian methods are written at the graduate level (*e.g.*, Robert, 2001; Gelman *et al.*, 2013), for students with specialized mathematical training, although a handful of undergraduate Bayesian textbooks have been published (*e.g.*, Berry, 1996). Moreover, the available learning materials heavily emphasize the use of default prior distributions, akin to that of Scenario 4 in Figure 1, and leave little to no room for covering elicitation in any systematic way, apart from its role in the conceptual development of Bayesianism. On elicitation, the books by O'Hagan *et al.* (2006) and Kadane (2011, 2016) are important resources, but are not configured for crossing the qualitative-quantitative dyad. What this situation suggests is work to be done on the way forward, that of creating educational bridges that would enable a flowering of the methodological bridges that I have envisioned here.

References

Allman, D. (2013). The sociology of social inclusion. *SAGE Open*, 1-16. doi: 10.1177/2158244012471957

Berger, J. O. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, *1*(3), 385-402. doi: 10.1214/06-BA115

Berry, D. A. (1996). *Statistics: A Bayesian perspective*. Belmont, CA: Duxbury Press.

Bloor, D. (1983). *Wittgenstein: A social theory of knowledge*. New York: Columbia University Press.

Cappo, D., & Verity, F. (2014). Social inclusion and integrative practices. *Social Inclusion*, *2*(1), 24-33.

Conquergood, D. (1985). Performing as a moral act: Ethical dimensions of the ethnography of performance. *Literature in Performance*, *5*(2), 1-13. doi: 10.1080/10462938509391578

Daponte, B. O., Kadane, J. B., & Wolfson, L. J. (1997). Bayesian demography: Projecting the Iraqi Kurdish population, 1977-1990. *Journal of the American Statistical Association*, *92*(440), 1256-1267.

de Sousa Santos, B. (2018). *The end of the cognitive empire: The coming of age of epistemologies of the south* (Kindle ed.). Durham, NC: Duke University Press. Retrieved from `Amazon.com`

Diaconis, P., & Freedman, D. (1986). On the consistency of Bayes estimates. *The Annals of Statistics*, *14*(1), 1-26. doi: 10.1214/aos/1176349830

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*(3), 193-242. doi: 10.1037/h0044139

Ellingson, L. L., & Ellis, C. (2008). Autoethnography as constructionist project. In J. A. Holstein & J. F. Gubrium (Eds.), *Handbook of constructionist research* (p. 445-466). New York: Guilford Press.

Ellis, C. (2004). *The ethnographic I: A methodological novel about autoethnography.* Walnut Creek, CA: AltaMira Press.

Epstein, R. A. (2009). *The theory of gambling and statistical logic* (2nd ed.). Burlington, MA: Academic Press. doi: 10.1016/C2009-0-20160-7

Garthwaite, P. H., Kadane, J. B., & O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, *100*(470), 680-700.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). New York: Chapman and Hall/CRC. doi: 10.1201/b16018

Goldman, A. I. (1999). *Knowledge in a social world.* Oxford: Clarendon Press.

Goldstein, M. (2006). Subjective Bayesian analysis: Principles and practice. *Bayesian Analysis*, *1*(3), 403-420. doi: 10.1214/06-BA116

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575-599. doi: 10.2307/3178066

Harding, S. (2008). *Sciences from below: Feminisms, postcolonialities, and modernities* (Kindle ed.). Durham, NC: Duke University Press. Retrieved from `Amazon.com`

Hathcoat, J. D., & Meixner, C. (2017). Pragmatism, factor analysis, and the conditional incompatibility thesis in mixed methods research. *Journal of Mixed Methods Research*, *11*(4), 433-449.

Howson, C., & Urbach, P. (1989). *Scientific reasoning: The Bayesian approach.* La Salle, Ill: Open Court.

Kadane, J. B. (2011). *Principles of uncertainty.* New York: Chapman and Hall/CRC.

Kadane, J. B. (2016). *Pragmatics of uncertainty.* New York: Chapman and Hall/CRC.

Kimmerer, R. W. (2013). *Braiding sweetgrass: Indigenous wisdom, scientific knowledge and the teachings of plants.* Minneapolis, MN: Milkweed Editions.

Koikkalainen, P. (2011). Social inclusion. In M. Bevir (Ed.), *The SAGE handbook of governance.* London: SAGE Publications.

Lenard, P. T. (2008). Trust your compatriots, but count your change: The roles of trust, mistrust and distrust in democracy. *Political Studies*, *56*, 312-332. doi: 10.1111/j.1467-9248.2007.00693.x

Levin, S. (2017, December 22). As Google AI researcher accused of harassment, female data scientists speak of 'broken system'. *The Guardian*. Retrieved from `https://www.theguardian.com/technology/2017/dec/22/google-ai-researcher-sexual-harassment-female-data-scientists`

Liamputtong, P. (2020). Social inclusion, research and practices in health and social care: Setting the scene. In P. Liamputtong (Ed.), *Handbook of social inclusion. research and practices in health and social sciences* (chap. 1). Singapore: Springer.

Lindley, D. V. (1958). Professor Hogben's 'crisis'—a survey of the foundations of statistics. *Journal of the Royal Statistical Society, Series C*, *7*(3), 186-198.

Lum, K. (2017, December 13). Statistics, we have a problem. *[blog post]*. Retrieved from `https://medium.com/@kristianlum/statistics-we-have-a-problem-304638dc5de5`

Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, *13*, 14-19.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think.* Boston: Houghton Mifflin Harcourt.

Mayo, D. G. (2005). Evidence as passing severe tests: Highly probable versus highly probed hypotheses. In P. Achinstein (Ed.), *Scientific evidence: Philosophical theories & applications* (p. 95-128). Baltimore: The Johns Hopkins University Press.

Mayo, D. G. (2018). *Statistical inference as severe testing: How to get beyond the statistics wars.* Cambridge: Cambridge University Press. doi: 10.1017/9781107286184

Meixner, C., & Hathcoat, J. D. (2018). The nature of mixed methods research. In P. Liamputtong (Ed.), *Handbook of research methods in health social sciences.* Singapore: Springer.

Meixner, C., & Hathcoat, J. D. (2020). Social inclusion and mixed methods research. In P. Liamputtong (Ed.), *Handbook of social inclusion. research and practices in health and social sciences.* Singapore: Springer.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74.

Moore, S., & Mitchell, R. (2011). Theorising rights-based restorative justice: The Canadian context. *The International Journal of Children's Rights*, *19*, 81-105. doi: 10.1163/157181810X508483

Moss, J., & Howlin, P. (2009). Autism spectrum disorders in genetic syndromes: implications for diagnosis, intervention and understanding the wider autism spectrum disorder population. *Journal of Intellectual Disability Research*, *53*(10), 852-873. doi: 10.1111/j.1365-2788.2009.01197.x

Naerland, T., Bakke, K. A., Storvik, S., Warner, G., & Howlin, P. (2017). Age and gender-related differences in emotional and behavioural problems and autistic features in children and adolescents with down syndrome: a survey-based study of 674 individuals. *Journal of Intellectual Disability Research*, *61*(6), 594-603. doi: 10.1111/jir.12342

Nind, M. (2014). *What is inclusive research?* London: Bloomsbury.

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., … Rakow, T. (2006). *Uncertain judgements: Eliciting experts' probabilities.* Wiley. doi: 10.1002/0470033312

Polite, A. (2018). To the roots and wounds: The case for restorative justice in the juvenile correctional system. *2018 Mellon Mays Undergraduate Fellowship Journal*, 103-107. Retrieved from `https://uraf.harvard.edu/mmuf-journal`

Popper, K. R. (1962). *Conjectures and refutations: The growth of scientific knowledge.* New York: Basic Books.

Robert, C. P. (2001). *The Bayesian choice: From decision-theoretic foundations to computational implementation* (2nd ed.). New York: Springer.

Savage, L. J. (1954). *The foundations of statistics.* London: Wiley.

Simplican, S. C., Leader, G., Kosciulek, J., & Leahy, M. (2015). Defining social inclusion of people with intellectual and developmental disabilities: An ecological model of social networks and community participation. *Research in Developmental Disabilities*, *38*, 18-29. doi: 10.1016/j.ridd.2014.10.008

Spitzner, D. J. (2020). *Reporting weight-of-evidence as pool reduction.* (Preprint)

Startin, C. M., D'Souza, H., Ball, G., Hamburg, S., Hithersay, R., Hughes, K. M. O., ... Strydom, A. (2020). Health comorbidities and cognitive abilities across the lifespan in down syndrome. *Journal of Neurodevelopmental Disorders*, *12*(4). doi: 10.1186/s11689-019-9306-9

Tua, A., & Banerjee, S. (2019). Social inclusion model: An adolescent parents intervention case study for measuring social inclusion outcomes. *Child Welfare*, *97*(3), 109-131.

Warren, K. (2015). Feminist environmental philosophy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy.* Retrieved from `https://plato.stanford.edu/entries/feminism-environmental`