Part I: Incentives, Bias, and Scientific Inquiry

Casey Petroff

6/17/2019

I. Introduction

How can bias affect beliefs about factual truths? This project examines beliefs about the transmission of cholera in the 19th century using a dataset of articles, books, and other writings from scientists and medical practitioners. Despite several major scientific breakthroughs, it took the Western medical research community until the end of the 19th century to converge on the correct understanding that cholera is contracted through direct or indirect contact with infected individuals (rather than having environmental origins). Public health historians attribute this slow shift in beliefs to political biases: if cholera is contagious, then it may be necessary for policymakers to impose a quarantine, which could have costly economic consequences, especially to cities whose local economies depended on trade or transportation. This incentivizes researchers dependent on patronage from political interest groups to ignore evidence of cholera's contagious nature and continue to propagate "anti-contagionist" ideas in the face of contrary evidence. I propose to empirically test this theory with a large database of contemporary scientific writing, drawing on theories from social science on motivated reasoning, biased belief formation, and the spread of misinformation.

II. Dependent variable

The dependent variable for this project is the *stated beliefs* of researchers about the origins of disease. These are contained in a database of medical journal articles, books, letters, and other documents that have been digitized and collated by the Internet Archive. The Medical Heritage Library contains approximately 127,000 items from between 1800 and 1900, roughly 30,000 of which mention cholera. Many – although not all – of these documents contain their authors' thoughts on the origins of cholera (vis-à-vis its symptoms or suggested treatment). Some articles specifically reference perceived negative externalities for economic activity under different public health policies. While these articles cannot reveal their authors' true, private beliefs, they do capture their publicly stated beliefs, and information about how they

arrived at their public conclusions – for instance, what evidence they give, and what other authors' work they cite – as well as information about the authors themselves – their place of practice, their education, their professional associations – that paint a picture of their political environments, social networks, and private incentives.

II. Review of literature

I refer to two main bodies of literature for this project: a substantive literature on the history of disease control and scientific progress, and a theoretical literature from social psychology on the nature and origins of biased reasoning. There is ample scholarly precedent to support a theory that scientific views of public health were not altogether free of material and political incentives: Ackerknecht (2009) summarizes the history of the debate over contagionism as

"never a discussion on contagion alone, but always on contagion and quarantines. Quarantines meant, to a rapidly growing classs of merchants and industrialists, a source of losses, a limitation to expansion, a weapon of bureaucratic control that it was no longer willing to tolerate, and this class was quite naturally with its press and deputies, its material, moral, and political resources behind those who showed that the scientific foundations of quarantine were naught, and who anyhow were usually sons of this class." (p. 9; the emphasis is given in the text).

Case studies of individual cities – Paris (Kudlick 1996), London (Durey 1979), and Hamburg (Evans 1987) – lend evidence to Ackerknecht's claim. In Paris, anti-contagionism was a "doctrine virtually tailor-made for the Parisian bourgeosie" because of its compatibility with "laissez-faire economics" (Kudlick 1996, p. 79); in London in 1832, some Whigs believed cholera to be a hoax intended to distract from political reform by way of quarantines and commercial disruption (Durey 1979, p. 188-189). Sixty years later, a cholera epidemic in Hamburg killed between eight and ten thousand people under the administration of Max von Pettenkoffer, a public health advisor and convinced anti-contagionist with close ties to Hamburg's merchant elite.

This project is also theoretically informed by literature from social psychology on how bias affects the updating of beliefs. Kunda (1990), a foundational work in this field, proposes

a theory of motivated reasoning: that "motivation may affect reasoning through reliance on a biased set of cognitive processes – that is, strategies for accessing, constructing, and evaluating beliefs" to "yield the desired conclusion" (p. 1). Within the constraints of plausibility, people are inclined to bias their reasoning process to produce preferred outcomes. The proposed project will evaluate explanations proposed by public health scholars for the slow adoption of modern germ theory as an explanation for cholera using the toolbox from social psychology for understanding why even rational scientific enquiry might be subject to bias.

IV. Proposed explanations

As outlined in Section II, my main explanatory outcome is stated belief about how cholera is transmitted, measured at the level of individual written works on the subject. My working hypotheses, informed by historical case study and in-depth examination of a small subset of the dataset ($n = 200$), and subject to updating as I begin my analyses, are as follows:

a. Researchers' opinions will be subject to bias against the theory of contagionism if the public policies implied by contagionism would be harmful to their audience.
b. Opinions on the cause of cholera will eventually converge to the correct one – that cholera is primarily waterborne (or foodborne) and is thus contagious – over time. This will be driven by new scientific evidence so incontrovertible that it cannot be overcome by plausible reasoning, however motivated. However, researchers with an incentive to be biased will be slower to acknowledge this new evidence.

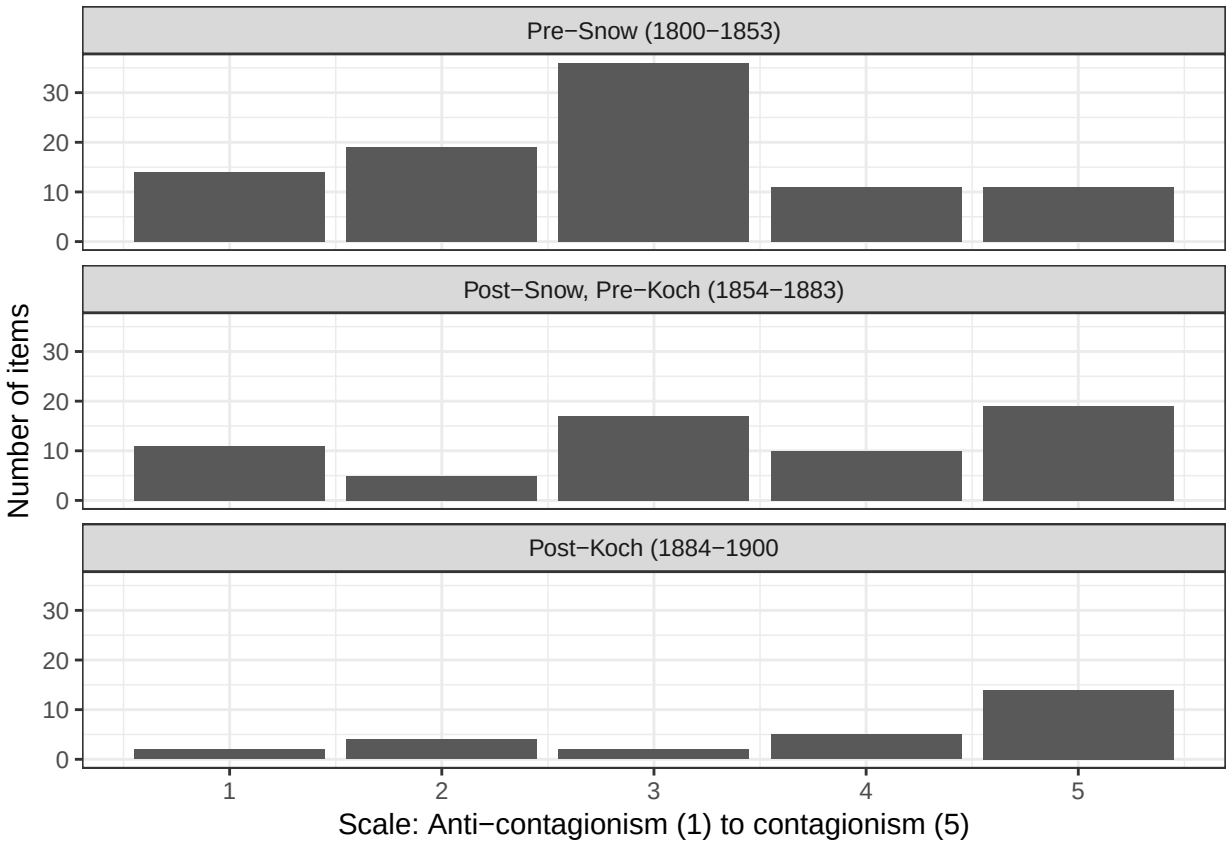
Borrowing the language of Kunda, I assume that all researchers have an "accuracy" motive – it is at least somewhat important to understand the true cause of cholera, in order to stop the spread of disease. However, some, depending on where they practice and to what social networks they belong, are also affected by a "directional" motive that introduces bias – they receive some payoff from coming to a particular conclusion (per my hypothesis, an anti-contagionist conclusion that is convenient for trade).

Crucially, cholera was a new disease in 19th-century Europe; it originated in India, where it had long been endemic, and spread westward in a series of pandemics over the course of

decades. At the beginning of the 1800s, most European physicians and natural scientists would have had weak priors about its underlying causes. This environment would provide the cover needed to sustain motivated, directional beliefs at first, until subsequent exposure to cholera, and experience observing it, would make sustaining those beliefs impossible. However, researchers receiving a positive payoff for proclaiming a directional conclusion – that cholera is not contagious, and thus cannot pose any threat to trade – will take longer to update to the correct posterior conclusion.

V. Operationalization

As I detail above, the outcome variable of interest is the stated view of researchers whose writings on cholera have been cataloged and digitized in the Medical Heritage Library. I have read a subsample of 200 articles and have classified them on a scale of 1 (extremely anti-contagious) to 5 (extremely contagious). Comfortingly, it is clear that over time, represented views converge on the right answer, and contagious views are more represented over time. The accompanying figure shows the distribution of viewpoints over time. Section VI (Methodology) discusses my plans to classify the remainder of the dataset.



Next, I turn to the question of explanatory variables. The public health literature indicates that pressure from trade and transportation interests is the primary factor that biased scientific views on disease in the 19th century. Accordingly, I would expect researchers located in cities that were more trade-dependent to be more inclined towards anti-contagionism. However, there is no precedent in the literature of which I am aware that provides granular 19th-century city-level data for trade dependency or market access outside of the United States, leaving the problem of sorting cities into trade- vs non-trade cities to me. This will require me to create my own classification system pending the organization of non-standard city names in the dataset.

Other than geographic origin, there is other information of substantive interest to the question of how researchers eventually converged on the correct conclusion about the origins of cholera. Frequently, the metadata supplied by contributing organizations includes information about authors' professional affiliation, for instance, universities or scholarly societies. In some cases, this may give information about proximity to political elites (for instance,

membership in the Royal Society) and enable me to build a network of scholars who likely would have been in close personal contact with one another. Extracting citations of others' work from their text will further enable me to study social networks – for instance, whether researchers more frequently cite others whose conclusions mirrored or differed from their own, and how key turning points in the epidemiology of cholera – for instance, Snow's 1854 natural experiments in waterborne transmission, or Koch's identification of *Vibrio cholerae* in 1884 – were acknowledged and received by different scholars.

VI. Methodology

I have downloaded the comprehensive contents of the Medical Heritage Library from 1800-1900 (as text files) and their accompanying metadata via the Internet Archive API and am now beginning the analysis. The first task will be to classify articles according to their sentiment towards contagionist theory. Because of the large amount of available data, I plan to draw on automated text analysis for this part of the project. Because I feel confident in my ability to classify articles myself, I can create a training set of correctly classified documents and then trial different ways of automated classification for the remainder. One common approach is that of Gentzkow and Shapiro (2010), which analyzes U.S. newspaper articles to determine political slant. Congressional speeches (which, classified by political affiliation of the speaker, make up the training set) are analyzed to determine the two- and three-word phrases most likely to be used by Democrats but not Republicans, and by Republicans but not Democrats. These phrases are then used to identify the likely political slant of newspapers. Another approach is given by King, Lam, and Robertson (2017). In this approach, an algorithm suggests appropriate classifying key words or phrases to a human user, whose acceptance or rejection of the key words is used to suggest a new set of possible classifying words and phrases. Should an automated approach prove unable to satisfactorily classify articles, I will consider alternate approaches, for instance, extracting sentences that are most likely to reveal information about the scientific persuasion of the author and classifying them (or having others classify them) by hand. While the majority of the articles in the dataset are in English, a substantial number are not; German, French, Spanish are particularly well-represented, and other languages (i.e. Italian and Hebrew) are

also present. So far, translation software has proven effective at supplying a comprehensible English translation that is good enough for classifying the document.

Another task is the extraction of key metadata about sources and their authors. Much of this work has already been done by contributing organizations, however, it will require some hand-cleaning on my part. For all but the largest cities, place of publication has proven to be a reasonable indicator of an author's professional location, however, some large cities (like London) published books and articles for authors further afield. Luckily, the texts in my subsample analysis include sufficient information about the author to allow for hand-coding.

In preliminary presentations of my research plans, my audience has suggested that I supplement this dataset with other data that could potentially shed light on authors' private vs. public statements and thus separate a mechanism of intentional self-censorship (of politically unpopular research) from one of motivated reasoning (in the style of Kunda 1990, where beliefs themselves are sincere but biased). While I presently do not count on being able to find a comparably large dataset of private beliefs (i.e. private correspondence), I will consider whether it would make sense to do a case study analysis if the right evidence is available. Additionally, I am considering a suggestion that I create a complementary dataset of views on cholera from non-expert sources, for instance, from newspapers, to understand the parallel evolution of beliefs between experts and non-experts.

VII. Implications

The question of how bias can enter the knowledge production process – even under the supposedly rigorous conditions of scientific inquiry – has implications beyond this case study of 19th-century public health. In particular, I see it as a useful contribution to the literature on misinformation and false perceptions. Flynn, Nyhan, and Reifler (2017) uses the motivated reasoning framework of Kunda (1990) and others to frame the impact of political bias on factual beliefs and provides a comprehensive list of relevant studies. Kahan et al. (2016), for instance, finds that subjects' stated beliefs about the risks posed by the Zika virus can be altered when they are exposed to framings that attributed Zika to global warming (which causes subjects with right-wing political views to downplay the risk of Zika, relatively) versus immigration (which causes subject with left-wing views to downplay the risk). The case

of cholera in the 19th century is especially interesting because the corpus of data I plan to analyze documents the stated views of experts whose views should, ideally, be unbiased, and could provide valuable insight into *how* people with false beliefs justify them. On a more positive note, this project can also contribute to our understanding of how ideas – including good ones – propagate and become mainstream, and what kind of evidence is required to make a correct idea palatable in the face of opposition.

References:

Ackerknecht, Erwin H. (2009) "Anticontagionism between 1821 and 1867." International Journal of Epidemiology 38, p. 7-21.

Durey, Michael. (1979) The Return of the Plague: British Society and the Cholera 1831-2. Dublin: Gill and Macmillan Humanities Press.

Evans, Richard. (1987) Death in Hamburg: Society and Politics in the Cholera Years 1830-1910. Oxford: Clarendon Press.

Flynn, D.J., Brendan Nyhan, and Jason Reifler. (2017) "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics." Political Psychology 38:S1, p. 127-150.

Gentzkow, Matthew and Jesse M. Shapiro. (2010) "What Drives Media Slant? Evidence from U.S. Daily Newspapers." Econometrica 78:1, p. 35-71.

Kahan, Dan M., Kathleen Hall Jamieson, Asheley Landrum, and Kenneth Winneg. (2016) "Culturally antagonistic memes and the Zika virus: an experimental test." Journal of Risk Research 20:1, p. 1-40.

King, Gary, Patrick Lam, and Margaret E. Roberts. (2017) "Computer-Assisted Keyword and Document Set Discovery from Unstructured Text." American Journal of Political Science 00:00, p. 1-17.

Kudlick, Catherine J. (1996) Cholera in Post-Revolutionary Paris: A Cultural History. Berkeley: University of California Press.

Kunda, Ziva. (1990) "The Case for Motivated Reasoning." Psychological Bulletin 108:3,

p. 480-498. Available at

Medical Heritage Library. (2019) Christine Wagner, creator. Internet Archive. https://archive.org/details/medicalheritagelibrary.