Title
Identifying and Filling Gaps in the Conspiracy Theory Literature

Author
Mankoff, Bradley Stephen

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Identifying and Filling Gaps in the Conspiracy Theory Literature

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Political Science

by

Bradley Stephen Mankoff

2022

ABSTRACT OF THE DISSERTATION

Identifying and Filling Gaps in the Conspiracy Theory Literature

by

Bradley Stephen Mankoff

Doctor of Philosophy in Political Science

University of California, Los Angeles, 2022

Professor Susanne Lohmann, Chair

This dissertation consists of three chapters. In the first chapter, I employ Luhmann's orders of observation to organize and identify blind spots that plague the academic literature that studies conspiracy theory and conspiracism. In so doing, I identify two blind spots that I address in the second and third chapters. In the second chapter, I take seriously epistemological research arguing that not all conspiracy theories are dangerous. I develop a framework by which one may approximate the danger associated with mass, serious consideration of particular conspiracy theories. I conclude that mass, serious consideration of most conspiracy theories is not as dangerous as ignoring those conspiracy theories. In the third chapter, I take seriously psychological research arguing that censorship tends to backfire. That research motivates my effort to test the viability allowing un-encumbered discussion of conspiracy theories on social media. I train a Random Forest classifier that estimates the probability that a tweet employs stigma against the conspiracy theory referenced in that tweet. I use the model's predictions to test

whether the stigma associated with conspiracy theory and conspiracism is increasing, whether stigmatizing events cause changes in stigma on Twitter, and whether changes in stigma on Twitter cause changes in retweets for high-stigma tweets relative to low-stigma tweets. I conclude that while non-censorship is unlikely to reduce the spread of conspiracy theories, censorship is also unlikely to reduce the spread of censored conspiracy theories.

The dissertation of Bradley Stephen Mankoff is approved.

Martin Orr

Davide Panagia

Zachary Steinert-Threlkeld

Susanne Lohmann, Committee Chair

University of California, Los Angeles

2022

TABLE OF CONTENTS

TABLES AND FIGURES

VITA

EDUCATION

2021 M.A., Political Science, University of California, Los Angeles

2018 B.A., Political Science, Washington University in St. Louis

2017 Certificate, Stanford Summer Institute in Political Psychology

PUBLICATIONS AND CONFERENCE PAPERS

Mankoff, Bradley. "Testing approaches to policing conspiracy theory discourse." Accepted for MPSA 2022.

Mankoff, Bradley. 2021. "Is increased ideological distance from one's congresspeople associated with increased self-identification as an Independent?" Journal of Management Policy and Practice, Vol. 22(2): 57-71. http://m.www.na-businesspress.com/JMPP/JMPP22-2/6_MankoffFinal.pdf.

Mankoff, Bradley. "Investigating an environmental predictor of independent-identity." Presented at MPSA 2021.

Mankoff, Bradley. "Can a change in the party in power alter the way that news affects your happiness?" Presented at MPSA 2019. Accepted for ISPP 2018.

RESEARCH AND PROFESSIONAL EXPERIENCE

2021	Adjunct Professor teaching Empirical Approaches to the Study of Political Behavior to 53 students at Loyola Marymount University.
2021	Teaching Fellow teaching Introduction to American Politics to 140 students at UCLA.
2021	Formatted tables in LaTeX for Professor Joel D. Aberbach's (UCLA) forthcoming chapter "A developing conservative ideology" for the *Oxford-Stanford Book on the Changing Character of the American Right*.
2021	Conducted preliminary data analysis and designed graphs for Professor Karen Orren's (UCLA) working paper "Doing Time con'd: Border Trouble and the Trump Administration."
2019-2021	Served as a Teaching Assistant for six quarters in the Department of Political Science at UCLA.
2019-2020	Served as a Co-Chairperson (2X) to assist faculty in planning and actualizing Admitted Students Preview Weekend for the Department of Political Science at UCLA.

2018-2020	Served as a Newsletter Co-Chairperson for the Early Career Committee of the International Society for Political Psychology.
2019	Created a comprehensive dataset of state policies for Professor Chris Tausanovitch (UCLA) to complement the *UC Nationscape* project.
2017	Coded past partisan affiliations for every member of the European Courts of Justice for Professor Matthew Gabel (WUSTL).
2017	Compared survey data with "true value" census data to analyze the representativeness of samples for Professor Matthew Gabel (WUSTL).
2016	Coded party-distribution of winning coalitions of EU countries for Professor Matthew Gabel (WUSTL).

Preface: Purpose and Roadmap of the Dissertation

Most scholars who study conspiracy theories belong in one of two camps—generalism or particularism. Generalists believe that we should dismiss certain conspiracy theories as a function of their claims and without regard for their evidence. Particularists believe that we should assess conspiracy theories as a function of their evidence (Buenting and Taylor 2010). One's beliefs about how a society ought to engage with conspiracy theories (ignoring them, ridiculing them, censoring them, etc.) follow from one's status as a generalist or particularist.

Philosophical works deliberate the merits of generalism and particularism (Generalist works: Popper 1994; Keeley 1999; Clarke 2002; Mandik 2007; Levy 2007; Stokes 2018; and Cassam 2018. Particularist works: Pigden 1995; Basham 2003; Coady 2003, 2006; Dentith 2014, 2018; and Hagen 2010, 2011, 2018, 2020, 2022). Where generalists do not explicitly concede particularism, their arguments fail to rebut their particularist foils. Nonetheless, work in psychology and political science continues to build on generalism toward conclusions about how a society ought to treat conspiracy theories (Fenster 2008; Swami et al. 2011; Darwin, Neave, and Holmes 2011; Wood, Douglas, and Sutton 2012; Jolley and Douglas 2014; Swami et al. 2014; Uscinski and Parent 2014; Melley 2016; van Proojien 2016; DiGrazia 2017; Imhoff and Lamberty 2017; Enders, Smallpage, and Lupton 2018; Lantian et al. 2021). The three chapter in this dissertation take a step back from these discussions to identify (Chapter One) and address (Chapters Two and Three) blind-spots that allow generalism to thrive *in practice* while it is dead *in theory*.

In Chapter One, I investigate the nature of elite-generalism—generalism among scholars, journalists, politicians, and social media corporations. I organize professional treatments of conspiracy theory to elucidate the causes and consequences of elite-generalism. First, I find that

elite-generalists do not attempt to justify their generalism except with the occasional recital of decades-old arguments (Popper 1994; Hofstadter 1965; Groh 1987; Jameson 1988) that fail to meet the rigors of logic and evidence. Second, I find that generalism is self-perpetuating. Generalism informs what a scholar chooses to measure. Measurement choices inform that scholar's conclusions. Those conclusions become the assumptions of subsequent scholars. As this cycle continues, echo-chambers form and intellectual progress halts (Orr and Dentith 2018).

In Chapter Two, I show that elite-generalists fear conspiracy theories more than they fear conspiracies. Elite-generalists assume that conspiracies tend to be thwarted by law enforcement and investigative journalism with such frequency that popular belief in conspiracy theories is not only useless by also harmful. On the other hand, elite-particularists refuse to concede that there may exist a set of conspiracy theories for which mass belief in those conspiracy theories is more harmful that the alleged conspiracy itself. In Chapter Two I bring these opposing perspectives into conversation with one another. Acknowledging the harm caused by mass, serious consideration of some conspiracy theories, I find that most conspiracies are much more dangerous than corresponding conspiracy theories—even when assuming a low probability that those conspiracies are real. Accordingly, I argue that governments and social media corporations act against the public interest when they steer discourse away from conspiracy theories.

In Chapter Three, I turn to the study of the stigma associated with belief in conspiracy theories. I use machine learning and natural language processing to develop a model that predicts the probability that a tweet about a conspiracy theory embodies the stigma associated with that conspiracy theory. Then, I use the model's estimates to answer questions in the conspiracy theory literature. I show that the stigma has increased over the last decade. I show that events that make a conspiracy theory look dangerous and/or unlikely increase the stigma associated that

conspiracy theory. And I show that when some event stigmatizes a conspiracy theory, tweets with higher probabilities of embodying stigma tend to receive more likes than tweets with lower probabilities. I conclude that organic practices of Twitter-users may accomplish some of the goals of censorship. Because studies show that censorship tends to backfire (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018), I advise Twitter and other social media corporations to cease censorship until they can prove that it reduces belief in the censored content.

Chapter One: Classifying Conspiracy-Theory-Endeavors

I. Introduction

"Conspiracy theory" means different things to different people. Accordingly, actors professionally engaged with conspiracy theory (esp. scholars, journalists, politicians, and social media corporations) often develop blind spots that prevent them from seeing relationships between their work and related work. This chapter seeks to illuminate and resolve these blind spots.

A. Organization of This Chapter

After this sub-section, I define key terms (the reader's own definitions are sufficient until then) and advocate for my definitions. Then, I introduce Luhmann's organizational schema (1993) and show how it applies to conspiracy-theory-endeavors.

The second section, "Blind Spots Among Conspiracy-Theory-Endeavors," is divided into two sub-sections, one for each blind spot. I focus on two manifestations of each blind spot. Regarding the first blind spot, social scientists studying conspiracy theory tend to ignore work that invalidates their assumptions. First, these social scientists tend to ignore particularist philosophers who argue that we should not dismiss conspiracy theories. Second, these social scientists tend to ignore historians and journalists who demonstrate the plausibility of individual conspiracy theories.

Regarding the second blind spot, actors professionally engaged with conspiracy theory tend to ignore some relevant work from psychology. First, social media corporations that censor some conspiracy theories tend to ignore the boomerang effect whereby attempts to restrict access to information tend to increase interest in that information. Second, scholars who study the stigma associated with conspiracy theory tend to ignore the study of stigma outside the context

of conspiracy theory. For the sake of readability, I discuss the blind spots before I classify the conspiracy-theory-endeavors even though it is the classification that reveals the blind spots.

Thus, in the third section, "Classifying Conspiracy-Theory-Endeavors," I utilize my adaptation of Luhmann's schema to classify conspiracy-theory-endeavors. In the fourth section, "Conclusion", I summarize and propose solutions to the problems I identify in preceding sections.

B. Definitions of Key Terms

Table 1

Term	My Definition	Flawed Alternative Definition
conspiracy	a multi-person plan or action employing morally suspect means and/or toward morally suspect ends	a *true* claim that a conspiracy is the best explanation for an outcome
conspiracy theory	a claim that a conspiracy is the best explanation for an outcome	a *false* claim that a conspiracy is the best explanation for an outcome
non-conspiratorial theory	a claim that some explanation other than a conspiracy best explains an outcome	
conspiracism	one's tendency to prefer a conspiracy theory to plausible, non-conspiratorial theories	one's tendency to prefer a *false* conspiracy theory to plausible, non-conspiratorial theories
conspiracist	one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories	
conspiracy theorist		one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories
generalism	the belief that we should dismiss certain conspiracy theories as a function of their claims and without regard for their evidence	
particularism	the belief that we should assess conspiracy theories based on the evidence supporting their claims	

Table 1 shows my definitions of key terms. I define "conspiracy" as a multi-person plan or action employing morally suspect means and/or toward morally suspect ends. I define a "conspiracy theory" as a claim that a conspiracy is the best explanation for an outcome. Many plans and actions not colloquially referred to as conspiracies nonetheless qualify under my definition. For example, my definition of conspiracies includes 1) marital infidelity, 2) political

corruption, 3) insider trading, and 4) the official theory of 9/11 (examples borrowed from Pigden 1995, 2006 and Dentith 2014). On the other hand, I define a "non-conspiratorial theory" as a claim that some explanation other than a conspiracy best explains an outcome. Examples of common non-conspiratorial theories include coincidences, lone actors, and institutional forces.

Under my definitions, confirmation that a conspiracy theory is accurate (confessions, DNA, fingerprints, etc.) does not cause the conspiracy theory to lose its status as such. Thus, we all believe many conspiracy theories and we are all conspiracy theorists. To allow for variance, I define "conspiracism" as one's tendency to prefer a conspiracy theory to plausible, non-conspiratorial theories, and I define a "conspiracist" as one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories. Next, I provide two reasons that my definitions are superior to alternatives.

First, there is an intimate relationship between a scholar's preferred definitions of terms and that scholar's perspective on conspiracy theories. Perspectives and definitions together cluster into two camps—generalism and particularism. Generalists believe that we should dismiss certain conspiracy theories as a function of their claims and without regard for their evidence. Particularists believe that we should assess conspiracy theories as a function of their evidence (Buenting and Taylor 2010). While particularists tend to prefer definitions like mine that allow for some conspiracy theories to be true (Pigden 1995, 2006; Dentith 2014), generalists tend to prefer definitions that by assumption render all conspiracy theories false (Pipes 1997; Hall 2006). These differences in definitions manifest as differences in research questions and methods.

Imagine a generalist scholar who defines a conspiracy theory as "a false claim that a conspiracy is the best explanation for an outcome." When this scholar uses a survey to measure

conspiracism, the scholar does not ask about belief in everyday conspiracies like marital infidelity or Watergate. Instead, they exclude these conspiracy theories from their surveys in favor of fringe conspiracy theories. This constitutes selection bias and shifts the subject of the research from conspiracists to those who believe fringe conspiracy theories. Thus, unbeknownst to them, these scholars study two phenomena at once (fringe-ness and conspiracism). Next, the scholar uses linear regression to show that, all else being equal, fringe conspiracists are less educated, more violent, etc. Finally, they conclude by ascribing this finding to *all conspiracists*.

Second, there is an intimate relationship between a scholar's identity (age, race, gender, nationality, sexual orientation, religion, etc.) and that scholar's attitudes toward some conspiracy theories. Consider the case of the 2018 chemical attack in Douma, Syria. American, French, and British officials claim that the Syrian military attacked opposition occupied Douma as part of the Syrian Civil War (Stewart and Perry 2018), but Russian officials claim that the United Kingdom (UK) staged the attack to frame the Syrian Regime (TASS 2018). If we use a generalist definition of conspiracy theory under which all conspiracy theories are false, Russian scholars are likely to disagree with American, French, and British scholars about whether the official Russian claim qualifies as a conspiracy theory. A Russian generalist would not include the alleged false flag attack on a conspiracy-theory-survey, but an American generalist would. In addition to nationality, time also confounds use of generalist definitions.

Consider the COVID-19 Lab Leak Hypothesis. In 2020, a *PolitiFact* fact-check labeled the COVID-19 Lab Leak Hypotheses "Pants on Fire"—"a conspiracy theory that has been debunked since the beginning of the coronavirus pandemic" (Funke 2020). Eight months later, we find appended to the original fact-check an Editor's Note stating, "…we are removing this fact-check from our database pending a more thorough review. Currently, we consider the claim

to be unsupported by evidence and in dispute (Editor's Note 2021). A generalist scholar would have included the Lab Leak Hypothesis on her conspiracy-theory-survey in 2020, but the same scholar would have omitted it in 2021.

Thus, when scholars use a generalist definition of conspiracy theory, they subject their research to both geographically and temporally sensitive selection bias that threatens the external validity of the research. Because we want our work to be cross-cultural and to stand the test of time, we must reject generalist definitions of conspiracy theory. Now that I have introduced and shown the value of my definitions, I introduce Luhmann's schema. Then, I apply the schema to conspiracy-theory-endeavors.

C. Defining Luhmann's Orders of Observation

Luhmann (1993) defines "orders of observation" in terms of distinctions and the blind spots that the distinctions leave in their wakes. Below, I explore this concept in the context of a simplified analysis of the toothpaste industry (summarized in Table 2).

Table 2

Order	Actor	Distinction	Blind-Spot
0	Consumer		
1	Manufacturer	Profitable vs. Unprofitable	Pollution
2	Regulatory Agency	Compliant vs. Noncompliant	Jobs
3	Economist	Positive vs. Negative Consequences	Odor
4	Legislator	Constituent-Likes vs. -Dislikes	Profits

Order 0 actors perform actions thoughtlessly. At some point during our day, we each act on Order 0. Imagine that we are all thoughtless Order 0 actors when we buy toothpaste. When we run out of toothpaste, we go to the store and buy the first tube we see. While we are thoughtless in this endeavor, a toothpaste manufacturer thoughtfully observes these choices in the aggregate.

First order actors observe the behavior of everyday citizens acting on Order 0. Consider a toothpaste manufacturer that prioritizes profit and ignores other considerations. The manufacturer observes the toothpaste consumption of everyday citizens. It distinguishes between profitable and unprofitable toothpaste ventures. In pursuing a profitable opportunity, the manufacturer becomes blind to environmental regulations. The manufacturer illegally dumps toxic waste into a local river.

Second order actors observe the behavior of first order actors. Consider a regulatory agency that enforces environmental regulations. The agency observes the behavior of manufacturers with respect to environmental regulations. It distinguishes between lawful and unlawful corporate behaviors. Pursuing enforcement of regulations, the agency becomes blind to the prudence of enforcing its mandate. In pursuit of compliance with regulations, the agency prosecutes the polluting manufacturer. In so doing, the agency is blind to the fact that the manufacturer provides many local jobs and is mobile enough to re-locate to a jurisdiction with fewer environmental regulations.

Third order actors observe the behavior of second order actors. Consider an economist who studies environmental regulations. The economist observes the behavior of the regulatory agency and studies the fiscal consequences of that behavior for society at large. She distinguishes between positive and negative fiscal consequences of enforcement. Pursuing sound analysis, the economist publishes a paper showing that this instance of enforcement has negative fiscal consequences. In so doing, the economist is blind to the accumulation of odor caused by the manufacturer's pollution.

Finally, fourth order actors observe the behavior of third order actors. Consider a legislator who prioritizes the will of constituents. The legislator observes her constituents'

growing desire to eradicate the odor. In distinguishing between what her constituents desire and what they do not desire, the legislator is blind to the manufacturer's profits.

Now that the reader understands Luhmann's orders of observation, I show how one may classify conspiracy-theory-endeavors into these orders. That will conclude Section I.

D. Orders of Observation Among Conspiracy-Theory-Endeavors

In this sub-section, I develop Luhmann's orders of observation in the context of conspiracy-theory-endeavors. After introducing each order in context, I provide two examples of endeavors in that order. Table 3 summarizes one example for each order (except Order 0).

Table 3

Order	Actor	Distinction	Blind-Spot
0	Anybody		
1	FBI Agent	Would my boss be interested in hearing about this potential conspiracy?	Philosophical merits of the legal definition of conspiracy
2	Journalist	Should I call this claim a "conspiracy theory"?	Influence of nationality on labeling decisions
3	Facebook	What handling of claims labeled "conspiracy theory" is best for our users?	Philosophical merits of second order labels
4	Scholar	What are the consequences of third order recommendations and/or actions?	The consequences of witholding the third order actions

In the context of the conspiracy theory literature, I define Order 0 as the locus of everyday activity. In all of our actions, we either conspire or we do not. Thus, conspiratorial plans and conspiracies that actualize those plans populate Order 0. These actions do not manifest as Luhmannian distinctions because the actors do not distinguish between conspiring and not conspiring. Rather, the actor acts toward her private objectives (going to the grocery store, rigging an election, etc.) and may in the process happen to conspire or not conspire.

One acts on Order 1 when one investigates and/or articulates Order 0 activity. First order observers—federal agents, journalists, or others—observe Order 0 activity and draw a distinction

between activity that is worthy and activity that is un-worthy of their attention as Order 1 actors. In drawing this distinction, they become blind to the merits of their threshold for worthiness.

Consider a Federal Bureau of Investigation (FBI) agent who chooses whether to report a potential conspiracy to her superiors. In doing so, she makes a distinction between what does and does not qualify as something her superior would want her to investigate, but she is blind to the philosophical merits of her superior's desires. Consider a Youtuber who declares "It's a conspiracy!" to her audience on YouTube. In doing so, she makes a distinction between what her viewers would or would not find worthy of their attention, but she is blind to the philosophical merits of their preferences.

One acts on Order 2 when one decides whether to label Order 1 activity as a conspiracy theory. Second order observers—journalists, politicians, scholars, or others—observe first order investigations or articulations of conspiracy and draw a distinction between claims they define as conspiracy theories and other claims. Order 2 activity is blind to criteria it uses to distinguish conspiracy theories from other claims.

Consider a Russian journalist who ruminates about how to refer the claim that the UK staged the Douma Chemical attack. She decides to refer to it as a "hypothesis" rather than a "conspiracy theory." In doing so, she makes a distinction between what should and should not be labeled a "conspiracy theory," but she is blind to the influence of national identity on her decision. Consider a scholar who in 2021 omits the Lab Leak Hypothesis from her conspiracy-theory-survey. In doing so, she makes a distinction between what should and should not be labeled a "conspiracy theory," but she is blind to the influence of dynamic authoritative opinion on her decision.

Table 4

Sub-Order	Description
2a	Labeling for convenience
2b	Labeling to measure conspiracism and its correlates
2c	Labeling to measure the individual-level structural correlates with conspiracism
2d	Labeling to explain the society-level appeal of conspiracism

Such a range of actions populate Order 2 that, for ease of communication, I articulate five sub-orders within this order (see Table 4). All actions in an order share a distinction and a blind spot in Luhmannian fashion, but all actions in a sub-order also share a research agenda. The sub-orders are: 2a) labeling for convenience; 2b) labeling to measure conspiracism and its correlates; 2c) labeling to measure the individual-level structural correlates with conspiracism; and 2d) labeling to explain the society-level appeal of conspiracism.

One acts on Order 3 when one recommends and/or takes actions as function of second order labels. Third order observers—scholars, fact-checkers, newspapers, governments, social media corporations, and others—observe second order labels and draw a distinction between justified and unjustified interaction with whatever is labeled "conspiracy theory." To generate general rules, Order 3 activity is blind to the nuances of most preceding orders. For example, when making a recommendation about what one should do with conspiracy theories, one abstracts away differences in the quality of the reasoning used to yield second order labels.

Consider a social media corporation that argues that it is wise to censor that which Order 2 labels "conspiracy theory." In so doing, the corporation makes a distinction between wise and un-wise responses to encountering the label, but it is blind to the biases that steer generation of the label. Consider a philosopher who argues that social media corporations should disregard a claim's label when they decide what to censor. In so doing, she makes a distinction between wise

and un-wise responses to encountering a claim labeled "conspiracy theory" on Order 2, but she is blind to the influence of factors other than wisdom (such as liability, profit, or risk aversion) that may motivate corporate censorship.

I articulate two sub-orders within this order. On sub-order 3a, philosophers debate generalism and particularism. On sub-order 3b, social media corporations attempt to intervene into conspiracy-theory-discourse (I call these actors "interventionists"). Also on sub-order 3b, scholars debate the merits of interventionism.

Lastly, one acts on Order 4 when one studies the consequences of third order actions. Fourth order observers—almost always scholars—observe third order recommendations and/or actions and draw a distinction 1) the consequences of recommendations and/or actions and 2) outcomes that are not the consequence of those recommendations and/or actions. Fourth order activity is blind to the consequences of *withholding* those third order actions.

Consider a scholar who argues that censorship of a conspiracy theory causes increased interest in that conspiracy theory. In so doing, she makes a distinction between 1) censorship and 2) all that does not cause increased interest in the conspiracy theory. But by making this distinction, she becomes blind to the possibility that interest in that conspiracy theory could have increased *even more* if the conspiracy theory were not censored. Consider a scholar who argues that government-sponsored educational programs that teach students to ignore conspiracy theories may cause society to mistakenly ignore a future conspiracy, thereby causing harm. In making this argument, she makes a distinction between 1) what the program causes and 2) all that the program does not cause. But in so doing, she becomes blind to the possibility that, in the absence of the program, belief in non-existent conspiracies may cause *even more* harm (consider how the Holocaust was motivated to some extent by belief in a false conspiracy theory).

With these orders in mind, I conclude Section I and turn to Section II. In Section II, I discuss blind spots among conspiracy-theory-endeavors that, I argue, my orders resolve. I provide at least one example of each failure.

II. Blind Spots Among Conspiracy-Theory-Endeavors

In this section, I explore two blind spots that plague conspiracy-theory-endeavors. I explore two manifestations of each blind spot and provide at least one example of each manifestation. I argue that classifying work according to Luhmannian orders of observation helps to reveal and thereby resolve these blind spots.

Regarding the first blind spot, second order social scientists studying conspiracy theory tend to ignore work that invalidates their assumptions. First, these social scientists also tend to ignore first order history and journalism that demonstrates the plausibility of individual conspiracy theories. Second, these social scientists tend to ignore third order philosophy that argues that we should not dismiss conspiracy theories.

Regarding the second blind spot, third and fourth order actors professionally engaged with conspiracy theory tend to ignore some relevant fourth order work from psychology. First, social media corporations that censor some conspiracy theories tend to ignore the boomerang effect whereby attempts to restrict access to information tend to increase interest in that information (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018). Second, scholars who study the stigma associated with conspiracy theory tend to ignore the study of stigma outside the context of conspiracy theory.

Note that Luhmann warns against undue demands for actors to attend to the blind spots that cause problems like those I review below. There are, after all, only so many hours in a day

and so many subjects to which one may attend. But the two blind spots I present below are too glaring and important to withhold. I explore both of these blind spots below.

A. First Blind Spot

Here, I focus on blindness by second order social scientists to the fallibility of their assumptions. This blindness trickles into mass consciousness as false assumptions manifest as measurements, conclusions, and lay tropes amplified by the mainstream news industry and social media corporations. I divide this blind spot into two manifestations.

First, second order social scientists also tend to ignore first order history and journalism that demonstrates the plausibility of individual conspiracy theories. Second, these social scientists tend to ignore third order philosophy that argues that we should not dismiss conspiracy theories. I use examples to explore both manifestations of the first blind spot.

i. First manifestation of the first blind spot

Here I explore the first manifestation of the first blind spot that plagues conspiracy-theory-endeavors. Examples of this manifestation feature scholars constructing surveys that they use to measure conspiracism and its correlates. I show how these scholars' choices to include and exclude certain conspiracy theories from their measures demonstrate insufficient knowledge of first order history and journalism. Regarding selection of conspiracy theories, these scholars ignore meaningful differences in the epistemic statuses of conspiracy theories. Regarding selection of potential correlates, these scholars demonstrate an anti-conspiracy-theory bias justified only by ignorance of the historical prevalence of conspiracy. These scholars begin with the generalist assumption that all conspiracy theories are false and, therefore, people higher in conspiracism must exhibit pathologies. They choose which correlates to measure as a function of this assumption. Three examples demonstrate my concerns.

Example 1: Goertzel (1994) begins with a fact that he designates as a puzzle worthy of inquiry: over half of Americans agree with the conspiracy theory claiming that "President Kennedy [(JFK)] was killed by an organized conspiracy" (732)? In designating this fact as a puzzle, Goertzel assumes that the conspiracy theory is false. To resolve this puzzle, Goertzel administers a telephone survey to 348 respondents in southwestern New Jersey. He measures belief in ten conspiracy theories (including the above claim about JFK) as well as several demographic and attitudinal variables. To determine which claims appear on his survey, Goertzel observes first order articulations of conspiracy and makes a second order distinction between conspiracy theories (some of which he includes on the survey) and other claims (all of which he excludes from the survey). I reproduce four more of his conspiracy theories in Table 5.

Table 5

Conspiracy Theory	% Who Believe it is True
JFK: President Kennedy was killed by an organized conspiracy	69
AIDS: The government deliberately spread the AIDS virus in the black community	10
UFOs: The Air Force is hiding evidence that the United States has been visited by flying saucers	41
MLK: The FBI was involved in the assassination of Martin Luther King	42
October Surprise: Ronald Reagan and George [H.W.] Bush conspired with the Iranians so that the American hostages would not be released until after the 1980 elections	55

I argue that substantive differences between these conspiracy theories invalidate the decision to treat them as examples of the same concept. I also argue that the unrelatedness of these conspiracy theories invalidates the assumption that the same traits motivate belief in each conspiracy theory. I explore the substantive differences between these conspiracy theories below.

Consider the assassination of JFK. While the first order Warren Commission concluded that Lee Harvey Oswald was JFK's lone assassin (1964), the first order House Select Committee on Assassinations (HSCA) concluded that a second gunman was likely involved (1979). Goertzel

does not attempt to adjudicate this ambiguity. In fact, he does not mention the findings of the HSCA. If he reviewed those findings and other first order scholarship about the JFK assassination before selecting cases for his survey, he would have noted the epistemic incongruity between the JFK conspiracy theory and his other conspiracy theories.

Similarly, while the AIDS-claim lacks evidential support and alleges state crimes against humanity unparalleled since the Holocaust, the subsequent three claims (UFOs, MLK, and October Surprise) allege activities that exhibit some evidential support.

Regarding MLK, the Church Committee (US Senate, Book II 1976) found that the FBI spied on and harassed MLK, and a Memphis jury in a civil suit decided that a there was a conspiracy to assassinate him (Yellin 1999). This does not prove the MLK conspiracy theory in Table 4, but it does distinguish it from the AIDS conspiracy theory. Regarding UFOs, the last several years features various disclosures that confirm that the US Department of Defense takes UFOs seriously (Andrew 2019; Duster 2021). Again, this does not prove the UFOs conspiracy theory in Table 4, but it does distinguish it from the AIDS conspiracy theory. In both of these cases, there was good reason to believe these conspiracy theories even before these relatively strong pieces of evidence emerged. There is and always has been little evidence to support the AIDS conspiracy theory.

Accordingly, while only 10% of respondents thought the AIDS conspiracy theory was true, 42% (MLK), 41% (UFOs), and 55% (October Surprise) thought the other three conspiracy theories were true. 69% thought the JFK conspiracy theory was true. Goertzel should not attribute belief in these conspiracy theories to traits of the believers when he does not measure variance in how likely each conspiracy theory is to be true nor how much the believers know

about each conspiracy theory. Thus, we see how uncritical inheritance assignment of Order 2 labels affects what a scholar measures and, as a result, what that scholar concludes.

Among other correlations, Goertzel find higher conspiracism correlates with lower interpersonal trust. Goertzel does not consider that it is rational for a person with high conspiracism to have low interpersonal trust. Goertzel's choice to emphasize the interpersonal-trust-result and ignore the rationality-result reveals Goertzel's anti-conspiracism bias.

The above flaws are characteristic of measures that rely on belief in individual conspiracy theories for the construction of a conspiracy-belief-scale. Seeking to avoid flaws like these, other second order research develops conspiracism-scales that measure belief in general claims about the prevalence of conspiratorial activity rather than belief in particular conspiracy theories. I address two of the most prominent of these scales in Example 2 and Example 3 below.

Example 2: Brotherton, French, and Pickering (2013) develop a conspiracism-scale using general claims about conspiratorial activity rather than claims that put forward specific conspiracy theories. Dozens of papers use their scale, inheriting their second order labeling choices. Beginning with 75 statements related to conspiratorial activity, the authors reduced this set to fifteen items using exploratory factor analysis. The epistemically agnostic mathematical method yielded a set of claims with varying epistemic statuses.

Consider the claim, "The spread of certain viruses and/or diseases is the result of the deliberate, concealed efforts of some organization." This claim lacks evidential support. Juxtapose this claim against the claim "Experiments involving new drugs or technologies are routinely carried out on the public without their knowledge or consent." To classify the latter claim in the same category as the former, one must ignore troves of first order research. Government-investigations, confessions, and first-hand testimony reveal dozens of experiments

carried out on the public without their knowledge or consent. The most prominent among these experiments are the Tuskegee Syphilis experiment and MKUltra (Wikipedia: "Unethical"). Thus, we see that from a historically informed perspective, the viruses/diseases claim is fundamentally different from the experiments claim. Treating the claims as manifestations of the same phenomenon compromises inferences related to that phenomenon.

The flaws of the scale do not end there. Consider the claim, "The government permits or perpetrates acts of terrorism on its own soil, disguising its involvement." Operation Northwoods, a proposed government program, recommended perpetrating acts of terrorism on US soil, disguising US involvement, and blaming the Cubans as an impetus to depose Fidel Castro (Wikipedia: "Operation Northwoods"). Even though Secretary of Defense Robert McNamara quashed the proposed program, the fact that such a program was proposed in the first place affects the epistemic status of the above claim. While there is little evidence of government officials proposing the intentional spread of disease (the attempt by the 18th century British soldiers to spread smallpox in defense of Fort Pitt excepted), such a record exists with respect to the "terrorism" claim above.

Thus, while endorsement of some items on this scale indicates anti-social mistrust or gullibility, endorsement of other items on this scale indicates commendable knowledge and skeptical application of that knowledge in a contemporary context. This incongruity compromises the utility of the scale and compromises the validity of findings that rely on the scale.

Example 3: Uscinski and Parent (2014) try another approach to measuring conspiracism. They base their conspiracism scale based on three questions to which subjects respond on a five-point scale ranging from "strongly agree" to "strongly disagree". I focus on the first two: 1)

"Much of our lives are being controlled by plots hatched in secret places." And 2) "Even though we live in a democracy, a few people will always run things anyway." The problems with this scale are similar to those with Goertzel's survey and the Brotherton-French-Pickering scale.

Consider the first claim: "Much of our lives are being controlled by plots hatched in secret places." Leading American financiers framed the Federal Reserve System in a secret meeting at the Jekyll Island Club in 1910 (Wikipedia: "Jekyll Island"). General Motors conspired to replace electric streetcars with its own gas-powered buses from 1938-1950 (Wikipedia: "General Motors"). The Manhattan Project was organized in secret. And ExxonMobil's efforts to delay acknowledgement of its industry's role in climate change took place in secret (Wikipedia: "ExxonMobil"). The modern phenomena of international finance, public transportation, mutually assured destruction, and climate change control much of our lives and stem from plots hatched in secret. Knowledge of these historical incidents ought to inform one's assessment of the claim that "Much of our lives are being controlled by plots hatched in secret places," yet agreement with this claim is classified as evidence of conspiracism rather than evidence of fundamental historical knowledge. Thus, inclusion of this survey-item in a conspiracism-scale reveals meaningful ignorance of first order articulations of conspiracy.

Next, consider the second claim: "Even though we live in a democracy, a few people will always run things anyway." Few scholars of American politics would dispute the empirical validity of this statement. A few people *will* always run things. Such is the theoretical basis of representative democracy wherein voters delegate the authority to govern to a few representatives. Further, institutions including the Federal Open Market Committee, the Supreme Court, and boards of directors at large corporations each 1) are composed of a small number of individuals, 2) "run things," and 3) are barely accountable to the demos. Thus, again we find

Uscinski and Parent eager to construe fundamental knowledge as evidence that one is a conspiracist.

Before turning to the second failure of communication, I demonstrate how choices of measurement and inference reveal the anti-conspiracism bias of these researchers. First, consider an example of measurement-choice. Rather than measure knowledge of historical conspiracies, the most obvious correlate with conspiracism, Uscinski and Parent measure, among other things, agreement with the statement "violence is sometimes an acceptable way to express disagreement with the government." To measure a non-obvious, negatively connoted correlate without measuring an obvious, positively connoted correlate demonstrates bias in measurement choices.

One of Uscinski's and Parent's inferences also demonstrates their bias. They show that those classified as "high" in conspiracism agree with the violence-statement more often than those classified as medium or low in conspiracism. From this result they infer that conspiracism is related to violence. They omit, however, that they also find that those low in conspiracism agree with the statement more often that those with medium conspiracism. Just as the former inter-group comparison supports their conclusion, the latter comparison contradicts it. An unbiased inference from these conclusions would reject the relationship between conspiracism and violence as unsettled by this analysis. Nevertheless, many scholars cite this alleged relationship with violence (Lamberty and Leiser 2019; Jolley and Paterson 2020; Rottweiler and Gill 2020; Douglas 2021), the news media is fascinated by the alleged relationship (Morrison 2019; Beckett 2020; Edsall 2021; Morris 2021), and social media corporations may repress discussion of conspiracy theories in part on the basis of this alleged relationship and the uncritical amplification thereof. The above scenario exemplifies how bias against conspiracy theories affects not only what scholars measure but also how they infer from their measurements.

Overall, scholars making second order claims should evaluate first order research before they compose their conspiracism-scales. I argue that this gap in their knowledge follows from their presumption against conspiracy theories—a presumption they perpetuate in their own work. Because they assume that conspiracy theories and general conspiratorial claims are false, they see no need to investigate them before including them in their measures. Thus, second order labelers will not see the need to review first order research until they first read third order philosophy that rebuts their generalist presumptions against conspiracy theories. Below, I review the failure by second order labelers to review third order philosophy that proves the merits of particularism over generalism.

ii. Second manifestation of the first blind spot

In the second manifestation of the first blind spot that plagues conspiracy-theory-endeavors, endeavors across orders ignore particularism's victory over generalism taking place on Order 3a. To demonstrate this manifestation, I explore one popular brand of generalism. This brand of generalism argues that we can dismiss conspiracy theories when they allege that large (in breadth of accusations and the number of people accused) conspiracies are the best explanation for important outcomes. To justify this position, they assume that conspiracies of this type are always thwarted before they actualize their plans because these conspiracies require a large number of people to keep a big secret. Three empirical contexts undermine this assumption.

First, the Holocaust was a conspiracy by Hitler's regime. It was orchestrated by multiple people and employed morally suspect means (murder) toward morally suspect ends (ethnic cleansing). Even though many knew about mass exterminations as early as 1942, the conspiracy's goals meaningfully came to pass (Plesch 2017). Second, the Manhattan Project

pursued morally suspect means (secretive spending of taxpayer dollars) toward morally suspect ends (indiscriminate murder). This conspiracy featured a team of approximately 129,000 workers, yet its existence was not revealed until the first bombs were dropped in Japan (Jones 1985). Third, the realities of Watergate and the Iran-Contra affair were both revealed fortuitously (tape on the door in the Watergate building), not by leaks stemming from fundamental impossibilities of large conspiracies (deHaven-Smith 2013). Thus, we observe three flaws with this presumption against conspiracy theories. First, even partially revealed conspiracies are not necessarily thwarted before they meaningfully actualize their plans. Second, some massive conspiracies are not revealed until they actualize their plans. And third, while some conspiracies are revealed and thwarted before they actualize their plans, revelation is not necessarily caused by fundamental impossibilities of conspiracy. When a conspiracy is revealed and thwarted at the eleventh hour and by serendipity, this calls into question how many large conspiracies we fail to reveal and thwart.

Different brands of generalism specify different sets of claims that, the generalists argue, justify dismissal of conspiracy theories that make those claims. Each brand of generalism fails. Many of the progenitors of these brands have even conceded as much, yet second order social scientists and others continue to reference these brands to support their own generalist dismissal of conspiracy theories. Thus, I argue that generalists across orders of observation should review third order research that contradicts their generalist assumptions. Next, I turn to the second blind spot and two manifestations thereof.

B. Second Blind Spot

Here, I focus on blindness by Order 3b interventionists and Order 4 scholars of conspiracy theory stigma to attend to psychological research related to their endeavors. First,

Order 3b social media interventionists tend to ignore the boomerang effect whereby attempts to restrict access to information tend to increase interest in that information (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018). Second, scholars who study the stigma associated with conspiracy theory tend to ignore the study of stigma outside the context of conspiracy theory.

i. First manifestation of the second blind spot

In the first manifestation of the second blind spot that plagues conspiracy-theory-endeavors, Order 3b interventionists ignore Order 4 efforts by psychologists to demonstrate that censorship tends to backfire. Each act of interventionism ignores decades of research demonstrating the presence of a boomerang effect (sometimes referred to as a Streisand Effect) whereby restricting freedom of access to information tends to cause increased interest the restricted information (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018). I provide three brief examples wherein an Order 3b effort seems unaware that its actions are self-defeating.

In 2019, YouTube declared that it will "begin reducing recommendations of borderline content and content that could misinform users in harmful ways—such as videos…making blatantly false claims about historic events like 9/11" (YouTube Team 2019). In the weeks preceding the 2020 US Presidential Election, Facebook and Twitter to varying extents censored a *New York Post* article about Hunter and Joe Biden conspiring toward mutual profit (Paul 2020; Manskar 2020; Devine 2021). Twitter blocked *The Post*'s account for 17 days (Sargent 2020). Nowhere in explanations of these moves do the interventionists acknowledge the likelihood that their actions increased interest in that which they censored.

Each of these massive corporations has the resources necessary to hire a psychologist to explain the boomerang effect or, even better, a philosopher who can explain commonsense rebuttals to generalist dismissal of conspiracy theories. Because these corporations seek to help their users assess the validity of complex information, they should have a strong interest in heeding this modest call to apply the insights of the best research that relates to their practices and goals. Next, I explore the second manifestation of the second blind spot that plagues conspiracy-theory-endeavors.

ii. Second manifestation of the second blind spot

In a second manifestation of the second blind spot that plagues conspiracy-theory-endeavors, fourth order scholars of the conspiracy-theory-stigma fail to sufficiently engage with fourth order psychological research that studies stigma in other contexts.

In clinical psychology, several papers measure and test hypotheses related to the stigma associated with mental illness, obesity, AIDS, and other stigmatized ailments (Robinson et al. 2019; Li et al. 2020; Powers, Koliska, and Guha 2019 and Kaufmann 2021). This work features quantitative content analyses of social media posts that would be ideal for use in the context of conspiracy theories while most of the conspiracy-theory-stigma research eschews quantitative research in favor of case studies and theoretical lamentations that duplicate work by Husting and Orr (2007) and Bratich (2008). As the psychological research and the social media data on which it relies are relatively new, this blind spot will likely resolve over the next few years. Chapter Three in this dissertation begins that resolution.

Fourth order research that utilizes quantitative methods fails to meet stigma where it lives—in the context of popular conspiracy theories and in the public discourse formed by newspapers, television news, and social media. Instead, these experiments study the stigma

associated with broad statements about conspiratorial activity (Wood 2015) and ask subjects to visualize potentially stigmatized acts (Lantian et al. 2018). Scholars should address the goals of fourth order research in the conspiracy theory literature with the methods of quantitative clinical psychology. Greater collaboration between those who study stigma inside and outside the context of conspiracy theory would help scholars in both fields by giving psychologists an additional domain in which to study stigma and by giving scholars of conspiracy theory novel methods to test their hypotheses.

III. Classifying Conspiracy-Theory-Endeavors

A. Order 0

I define Order 0 as the locus of everyday activity. Different Order 0 actions evidence of conspiratorial activity that varies in quantity and quality. When you go to the store to buy toothpaste, that Order 0 activity produces no evidence of conspiratorial activity. Real conspiracies—especially those that first order actors will eventual deem worthy of their attention—tend to produce evidence. For example, Watergate produced the Oval Office audio recordings of President Nixon. Conspiracies undertaken by the government sometimes produce declassified articulations of their conspiratorial activity. On the other hand, some Order 0 evidence of conspiratorial activity fails to meet critical muster. For example, consider the hoax *Protocols of the Elders of Zion* or supposed evidence of conspiracy provided by governments with extensive records of intentional deception (such as Russian and Syrian claims about a British false flag attack in Douma).

While Order 0 activity is fundamental and interesting, there is no obligation for a social scientist or philosopher to uncover or analyze primary source evidence of conspiracies. Rather, it is appropriate for them to defer to first order historians and journalists who professionally

investigate and synthesize Order 0 activity. For example, scholars beyond Order 1 are justified in deferring to Robert Woodward and Carl Bernstein's coverage of Watergate without listening to the Oval Office tape recordings.

B. Order 1

I define Order 1 as the locus of investigations and articulations of conspiratorial activity. First order actors investigate Order 0 data and articulate their findings. They argue that conspiratorial activity has or has not transpired. Federal agents, investigative journalists, historians, and others are paid to act on Order 1. Note that some while rank-and-file conspiracists act on Order 1, others prefer to act on Order 2, consuming rather than producing first order articulations.

Order 1 activity takes many forms. When an FBI agent investigates a crime, the agent will often consider the possibility of conspiracy—that more than one person was involved in the criminal activity. When investigating gang-activity or reporting that activity to her superiors, the agent investigates or articulates a conspiracy. When First Lady Hillary Clinton alleged a "vast right-wing conspiracy" against her husband, she acted on Order 1. When Bernstein reported on the Central Intelligence Agency's (CIA) influence within major news publications, he acts on Order 1 (Bernstein 1977). When Olmsted published a book exploring the history of conspiracy in the United States (US), she acts on Order 1 (Olmsted 2009).

Different first order actors use different standards in deciding whether to articulate a conspiracy. When a fact-checker like *Politifact* and a rank-and-file conspiracist consider the same Order 0 evidence, the fact-checker is less likely to articulate a conspiracy. Legacy newspapers fall somewhere in between. While they tend to defer to fact-checkers and reason from the same anti-conspiracy-theory-biases as fact-checkers, one of their investigative

journalists sometimes finds a scoop that the fact-checkers lack. *San Jose Mercury News* allowed its journalist Gary Webb to publish the "Dark Alliance" series about the CIA, but rival newspapers continued to ignore his findings (Bratich 2008).

As I transition toward discussion of second order actors, consider the following. According to common sense, differences in exposure to first order sources drive differences in conspiracism. If you read "Dark Alliance," you should be more likely to believe conspiracy theories that implicate the CIA than if you did not read "Dark Alliance." Nevertheless, I show in the next sub-section that second order actors ignore this common sense. They do not measure exposure to first order sources. Instead, they measure personality traits that may be caused by un-measured exposure.

C. Order 2

I define Order 2 as the locus of labeling. Here, a range of actors choose whether to apply the pejorative label "conspiracy theory" to first order articulations of conspiracy. Because this Order features hundreds of exemplars pursuing a variety of goals, I divide this Order into five sub-Orders: a) labeling for convenience; b) labeling to conspiracism and its correlates; c) labeling to measure the individual-level structural correlates with conspiracism; and d) labeling to explain the society-level appeal of conspiracy theories (see Table 4).

i. Labeling for convenience (2a)

Here I discuss Order 2a labeling for convenience. Consider the following example wherein a journalist labels claims "conspiracy theories":

> For centuries, antisemitic conspiracy theories have blamed Jews for all manner of atrocities. A number of current [conspiracy theories] allege Jews control banking, Hollywood and the news media.
>
> A 2021 report by the anti-hate group HOPE Not Hate, among others, concludes the COVID-19 pandemic has amplified interest in the "New World Order" conspiracy

> theory that a secret cabal of Jewish elites is plotting to take over the world. The report adds [that] social media has amplified those beliefs. (Schott 2022)

The claims Schott reviews are conspiracy theories under any definition, and they are always labeled as such by legacy news outlets. "Conspiracy theories" is simply how journalists refer to claims that Jews are to blame for all manner of atrocities. Labeling of this kind is thus trivial from the perspective of this chapter, but it nonetheless belongs on Order 2. Most second order labeling in society takes place on Order 2a, but most second order labeling in academia takes place on other sub-orders.

ii. Labeling to measure conspiracism and its correlates (2b)

Next, I turn to the Order 2b actions that must label claims "conspiracy theory" or not to select which claims they'll use to measure conspiracism and its correlates. I begin by focusing on attempts to identify the demographic correlates of conspiracism, then I turn to attempts to identify other correlates.

Race, partisanship, income and other demographics are highly predictive of belief in particular conspiracy theories (especially those that invoke a demographic trait) and conspiracism in general. For example, Black people are more likely than White people to believe in AIDS-as-genocide conspiracy theories (Goertzel 1994; Fiske 1999; Bratich 2008), and Republicans are more likely than Democrats to believe that a conspiracy conceals the fact that President Barack Obama is a Muslim and/or was not born in the US (Miller, Saunders, and Farhart 2016; Enders, Smallpage, and Lupton 2018).

Order 2b research consistently finds that education is inversely correlated with conspiracism. Those low in education tend to be more conspiracist than those high in education. It is unclear whether this correlation is due to 1) the influence of education on critical thinking (van Prooijen 2016) or 2) an anti-conspiracy-theory culture at universities. Different Order 2b

research projects measure different combinations of independent and dependent variables, thus inhibiting the execution of a meta-analysis. This leaves us with an unclear snapshot of the demographics associated with conspiracism and belief in specific conspiracy theories. Above, I focus on attempts to identify the demographic correlates of conspiracism. Below, I turn to attempts to identify other correlates of conspiracism.

Most attempts to identify non-demographic correlates of conspiracism rely on the assumption that there is something wrong with conspiracists. Thus, following from that assumption, this research measures pathologies and tests whether they correlate with conspiracism. Pathologizing efforts build on each other, so I review those efforts roughly in the order in which they were published.

Popper (1994) argues that conspiracists are prone to premodern forms of thinking rooted in the Cartesian assumption that outcomes must follow from motives and agency. That is, 1) the worse the outcome, the more malicious the agents must be and 2) the grander the outcome, the more powerful the agents must be. Popper attributes premodern thinking to ignorance about the social sciences which, Popper claims, demonstrate the myriad ways in which intentions fail to manifest desired outcomes. Popper implies connections between low educational attainment, poor reasoning skills, and conspiracism.

Influenced by Popper, Hofstadter (1965) argues that conspiracists are predisposed toward the paranoid style of politics. While a clinical paranoiac argues that there is a conspiracy targeting him alone, one practicing the paranoid *style* argues that there is a conspiracy targeting all Americans, all Christians, all White people, or some other group that usually includes the conspiracist. Hofstadter argues that while conspiracies happen, they don't matter for three related reasons. First, conspiracies tend to be revealed before they can actualize their goals. Second,

when conspiracies actualize their goals, those goals tend to fail to manifest intended downstream outcomes. Third, when the goals do manifest intended downstream outcomes, the impact tends to be small. Hofstadter argues that practitioners of the paranoid style of politics tend to over-estimate the impact of conspiracy theories. Hofstadter implies connections between ignorance, paranoia, anti-social mistrust, and conspiracism.

Popper and Hofstadter argue that the conspiracy theory of society—the claim that history is best explained as a series of conspiracies—and the paranoid style of politics together fomented and enabled totalitarianism in Europe before and during World War II. Popper and Hofstadter offer no scientific evidence supporting their claim. Nonetheless, subsequent second order research builds on Popper's and Hofstadter's conclusion that something is wrong with conspiracists. Now I pivot from qualitative to quantitative research that seeks the pathologies of conspiracists. I begin with three explorations of belief in the conspiracy theory of society.

First, Goertzel (1994) correlates his flawed conspiracy-belief measure with interpersonal mistrust and pessimism about society. He also shows that belief in one of his ten conspiracy theories was associated with belief in at least one more of his conspiracy theories. He infers from this result that conspiracists operate with a monological belief system resembling Popper's conspiracy theory of society.

Second, Swami et al. (2011) replicate Goertzel's findings. They mimic Goertzel's inference from this finding, arguing that conspiracists exhibit a monological belief system. They also find that stronger belief in conspiracy theories about the 7/7 London Bombing is associated with higher political cynicism, greater support for democratic principles, more negative attitudes to authority, lower self-esteem, and lower agreeableness.

Third, Wood, Douglas, and Sutton (2012) show that the more subjects believe that Princess Diana faked her own death, the more they believed that she was murdered. They replicate this finding by showing that the more subjects believe that Osama Bin Laden was already dead when US special forces raided his compound in Pakistan, the more they believe he is still alive. They infer that conspiracists hold contradictory beliefs

These three papers share two flaws. First, they forget that for a rational person, belief in conspiracy theory X *should* increase the prior probability one associates with conspiracy theory Y. Second, Basham (2017) and Hagen (2018) show how study-design, not belief in contradictory conspiracy theories, drives results for Wood, Douglas, and Sutton (2012). Wood, Douglas, and Sutton ask respondents to rate their level of agreement (on a seven-point scale) with various conspiracy theories, some of which contradict each other. Average agreement with the claim that Princess Diana faked her own death was less than two-out-of-seven. When the researchers encounter respondents who express two-out-of-seven agreement (read: *disagreement*) with the "Diana faked her own death" and "Diana was murdered" claims, they construe this as evidence that conspiracists sometimes believe contradictory conspiracy theories. They did not ask respondents whether they believe both claims simultaneously, and they do not indicate whether any respondents expressed strong agreement with both claims separately. As of November 5, 2022, the Wood, Douglas, and Sutton (2012) paper has 683 citations. Basham's (2017) paper has 12 citations, and Hagen's (2018) rebuttal has 24 citations. I conclude discussion of Order 2b with review of papers that seek other pathologies in conspiracists.

Clarke (2002) proposes that conspiracists more often than others commit the fundamental attribution error by overestimating the influence of dispositional factors (such as malice held by conspirators) at the expense of situational factors (such as market forces). Swami et al. (2014)

and Lantian et al. (2021) show that belief in conspiracy theories is associated with inferior critical thinking. Pennycook et al. (2015) show that those who are more receptive to "bullshit statements" (this is the scholarly term) exhibit more conspiratorial ideation. Cassam (2015) argues that conspiracists are more gullible than others. Imhoff and Lamberty (2017) show that conspiracism is correlated with higher need-for-uniqueness which, in turn, "motivat[es] the endorsement of such seemingly irrational beliefs" (Abstract). Darwin, Neave, and Holmes (2011) show that conspiracism correlates with paranoid ideation and schizotypy.

Given these findings, readers are left wondering how these cognitively impaired, paranoid dupes function in society. The answer is simple. All the above studies are flawed beyond usability. Their ignorance about first and third order research and/or their anti-conspiracy-theory bias causes them to develop research designs pre-disposed to find that many pathologies correlate with conspiracism. Their selection of statements they use to measure conspiracism and their selection of correlates to measure alongside conspiracism both manifest their ignorance and/or bias. This inhibits the ability to make credible inferences from their results.

Even when they find that a positive trait correlates with conspiracism, they find a way to twist that result. When Swami et al. 2011 find that support for democratic principles (SDP) correlates with conspiracism, the authors *post hoc* negatively connote SDP. They state that "it is possible to explain this set of findings as a function of greater anomie among conspiracy theorists" (450). Swami et al. 2011 has 527 citations, and only one English-language paper rebuts the inference related to SDP (Hagen 2018).

Thus, we observe a stark juxtaposition. In a world in which the historical record indicates that important conspiracies thrive, unscientific pathologizing of conspiracism earns hundreds of

citations while rebuttals to that work are scant and un-recognized. Next, I turn from research focused on pathologies to research focused on structural drivers of conspiracism.

iii. Labeling to measure the individual-level structural correlates with conspiracism (2c)

Above I demonstrate the pathologizing of conspiracism on Order 2b. Below, I turn to Order 2c claims that seek to explain individual-level conspiracism in terms of structural factors (factors induced by society) such as feelings of powerlessness and economic insecurity. Research embracing claims of this type rejects pathologization by arguing that pathologies may be downstream from structural factors. That is, claims on this sub-order argue that while conspiracists may indeed be inferior critical thinkers, that outcome may follow from 1) economic insecurity and/or 2) mistrust of government learned from generations of marginalization. Research on this sub-order thus seeks to excuse conspiracists for rationally held false beliefs.

Echoing Popper (1994), Groh (1987) and Jameson (1988) argue that conspiracism is caused by the yearning to find agentic, malevolent explanations for one's plight. Goertzel (1994) shows that belief in conspiracies is associated with lack of interpersonal trust and insecurity about employment. DiGrazia (2017) replicates the insecurity-finding.

Sunstein and Vermeule (2009) argue that conspiracism follows from a crippled epistemology whereby individuals arrive at false conclusions due to belief in false information. This research builds on arguments that crippled epistemology explains cult fanaticism (Hardin 2002) and religious fundamentalism (Baurmann 2007). Collectively, these three papers argue that *people rationally* become cult fanatics, religious fundamentalists, and conspiracists in response to belief in falsehoods. Despite the absence of data and causal inference techniques in these papers, the 2008-version of Sunstein's and Vermeule's paper has 916 citations. As I will review in my discussion of Order 3b endeavors, Sunstein and Vermeule use their conclusions to

motivate third order "cognitive infiltration of the groups that produce conspiracy theories" (Sunstein and Vermeule 2009:218).

Continuing the structuralist theme, Uscinski and Parent (2014) use the American context to show that conspiracism ebbs and flows with electoral tides. Politicians of a political party are more likely to be the target of conspiracy theories when their party has more political power. DiGrazia (2017) replicates this finding. Similarly, Enders, Smallpage, and Lupton (2018) show that political orientation is a strong predictor of which conspiracy theories people endorse. Beyond partisan effects, Uscinski and Parent (2014) show that greater external threats to the US are correlate with higher proportions of American conspiracy theories that implicate threatening external actors.

Jolley and Douglas (2014) show that feelings of political powerlessness mediate the causal relationship between exposure to information supporting a conspiracy theory and diminished intention to participate in politics. Van Proojien (2016) shows that feelings of powerlessness mediate the inverse relationship between education and conspiracism. Thus, if one feels powerless in politics, 1) exposure to a conspiracy theory reduces one's intention to participate in politics and 2) lower education is associated with higher conspiracism.

While Order 2c research continues to utilize meaningfully flawed measurement practices, this research at least features tight connections between theories and hypotheses. The Order 2b research lacks that connection. Next, I turn from research that focuses on individual-conspiracism to research that focuses on society-wide conspiracism.

iv. Labeling to explain the society-level appeal of conspiracy theories (2d)

I conclude my discussion of Order 2 claims with Order 2d claims that frame conspiracy theory as a societal rather than individual-phenomenon. This research does not study individual-

differences in conspiracism. This research claims to take seriously popular concern with conspiracy theory, however it still regards conspiracy theory as an epiphenomenon while not taking seriously the substance of conspiracy theories (Goshorn 2000).

Robins and Post (1997) attribute conspiracism in society to paranoia among both the creators and observers of conspiracy theories. This sounds like Order 2b research, however the focus on conspiracism *in society* makes it Order 2d research. Robins and Post do not measure individual differences in conspiracism nor paranoid. Pipes (1997) argues that conspiracy theories tend to emerge from the socially disaffected and culturally suspicious people within a society, which he identifies as Black Americans and the far-right. Dean (1998) argues that conspiracism stems from informational chaos caused by rapid technological development. Amid a proliferation of sources in the Information Age, individuals develop alternatives to and mistrust of official claims. Mark Fenster (2008) argues that society-level conspiracism follows from the prevalence of populist attitudes in that society. Lastly, Timothy Melley (2016) argues that popular realizations related to the dishonesty of the public relations industry and propaganda cause conspiracism in society.

v. Summarizing criticisms of Order 2 research

In almost all cases, second order research does not employ causal inference to test hypotheses. Instead, where researchers use data, they employ small samples, a-theoretical hypotheses, and multi-variate correlational regressions whose results rely on biased selection of variables. These researchers could improve their work by pursuing larger samples, letting theory motivate their hypotheses, and employing research designs that enable causal inference.

In a final methodological criticism, note that findings expressed in terms of regression coefficients relating independent to dependent variables mask variation among conspiracists. A

positive and statistically significant coefficient relating a paranoia scale to a conspiracism scale could indicate that 1) *every* respondent high in conspiracism are *a bit more* paranoid than the average respondent low in conspiracism or 2) *some* respondents high in conspiracism are *much more* paranoid than the average respondent low in conspiracism. While the former finding would pathologize conspiracism in general, the latter finding would merely show that paranoid people are more likely to believe conspiracy theories than others—a finding with no bearing on non-paranoid conspiracists nor conspiracism in general. Inferences from regression-results in second order research demonstrate insensitivity to this nuance.

Considering the bevy of methodological flaws on Order 2, those reviewing the literature on conspiracy theory should be wary of reading second order research before reading the Order 3a debate between generalism and particularism. Doing so could engender the epistemically unjustified biases that plague Order 2 research.

On the other hand, reliance on flawed methods is not sufficient to dismiss second order findings. The flawed methods of second order yield a narrow set of conclusions. Thus, where hypotheses follow from popular social science theories and scholars are willing to refine their measurement choices, this research is worthy of replication. If scholars can replicate their findings while satisfying critics, resulting research will deserve the attention that flawed research earns today. Next, I turn to third order conspiracy-theory-endeavors.

D. Order 3

I define Order 3 as the locus of recommendations issued or actions taken as a function of second order labels. This order divides into two sub-orders. Order 3a deliberates the merits of generalism and particularism, and Order 3b deliberates the merits of intervention into conspiracy theory discourse.

i. Epistemology (3a)

Order 3a research deliberates the merits of generalism and particularism. Below, I discuss Order 3a arguments. Dentith and Hagen provide longer reviews (Dentith 2014, 2018; Hagen 2010, 2011, 2018, 2020, 2022). I begin with three examples of non-scholarly generalism.

First, consider the case of President George W. Bush's speech to the United Nations in November 2001: "Let us never tolerate outrageous conspiracy theories concerning the attacks of September the 11th; malicious lies that attempt to shift the blame away from the terrorists themselves, away from the guilty" (Bush 2001). Rather than argue against conspiracy theories concerning 9/11, Bush advocates against conspiracy theories with fallacious begging of the question. By referring to these conspiracy theories as lies and the accused hijackers as guilty terrorists, he asserts as fact something that conspiracists refute.

Next, consider advocacy against conspiracy theories by the US Department of Defense. In 2005, that department's International Information Programs sub-division created a website which reads: "Does the story claim that vast, powerful, evil forces are secretly manipulating events? If so, this fits the profile of a conspiracy theory. Conspiracy theories are rarely true, even though they have great appeal and are often widely believed. In reality, events usually have much less exciting explanations." First, the website describes a particular type of conspiracy theory that is more likely to be false than other conspiracy theories. Then, the website argues that *all* conspiracy theories—not just those that meet the aforementioned criteria—are rarely true. This argument demonstrates the composition/division fallacy by extrapolating an argument related to a part (some conspiracy theories) to the whole (all conspiracy theories). See Bratich (2008) for deeper analysis.

Next, consider an example from Rebecca Solnit of *The New York Times*:

> While much has been said about the moral and political stance of people who support right-wing conspiracy theories, their gullibility is itself alarming.[… A] swath of the conservative population is available for the next delusion and the one after that. And on Jan. 6, 2021, we saw that a lot of them were willing to act on those beliefs. (Solnit 2022)

Solnit associates right-wing conspiracy theories with gullibility and delusion. She implicitly advocates for generalist dismissal of right-wing conspiracy theories, a dismissal with no greater epistemic support that broader generalism. Rather than engaging with conspiracy theories at the level of evidence, she infers falsehood from a claim's status as a conspiracy theory. She infers from that falsehood that believers of those false conspiracy theories must be gullible. Ultimately, Solnit instructs readers to ignore right-wing conspiracy theories because they are right-wing conspiracy theories, not because their evidence fails to prove their claims. While her epistemology is crude, it nonetheless places her article on Order 3a.

Above I reviewed three examples of non-scholarly advocacy for generalism. Below I turn to the scholarly debate between generalists and particularists. First, claims in Popper (1995), Hofstadter (1965), Robins and Post (1997), and Pipes (1997) advocate for generalism. This research disregards epistemology, instead offering circular arguments. First, they assume that conspiracists believe false things. Second, they infer from this assumption that conspiracy theories are false. Finally, they conclude that conspiracy theorists believe false things. Thus, these researchers assume that which they seek to prove—the logical fallacy *petitio principii*. In the process, they tend to avoid engagement with actual conspiracy theories, preferring instead to engage with straw-manned articulations of particular conspiracy theories and/or false generalizations about conspiracy theories in general.

Rebutting Popper and introducing the idea of particularism to the conspiracy theory literature, Pigden (1995) points out that no opponents of conspiracy theory *explicitly* deny that

conspiracies happen. Yet, he shows, they *implicitly* deny that platitude in their advocacy for generalism. This realization sparked the twenty-five-year discourse that I review below.

Conceding Pigden's criticism of generalism, Keeley (1999) attempts to salvage generalism by restricting *prima facie* dismissal of conspiracy theories to "Unwarranted Conspiracy Theories" (UCTs). Keeley defines UCTs as conspiracy theories that contradict official sources, allege nefarious intentions, tie together seemingly unrelated events, allege secrecy among conspirators, and rely for their persuasive force on their ability to explain errant data. Keeley concedes that Watergate and the Iran-Contra affair are both UCTs. Thus, Keeley argues that it is unwarranted to believe some true claims. To get there, Keeley argues on consequentialist grounds that mass belief in UCTs reduces public trust. Without public trust, Keeley argues, the foundations of democracy and science erode. Keeley has since adapted his position in the direction of particularism.

Next, to refine Keeley's (1999) argument, Clarke (2002) seeks criteria that allow one to dismiss some conspiracy theories without dismissing confirmed conspiracies like Watergate and the Iran-Contra affair. First, he rebuts Keeley's concern with public trust. Clarke argues that believing Watergate should cause one to mistrust Richard Nixon, not all of government. Shifting from a consequentialist to epistemological criterion for qualified generalism, Clarke offers Lakatos's notion of a degenerate research paradigm (Lakatos 1970). He argues that we should dismiss conspiracy theories that constitute a degenerate research paradigm, which Lakatos defines as a paradigm that experiences successive emergence of information that contradicts or fails to corroborate the paradigm's claims. Thus, if a conspiracy theory predicts outcomes that fail to transpire, the conspiracy theory becomes degenerate, and we should dismiss it without regard for its evidence.

Clarke's proposal fails for two main reasons. First, as pointed out by Keeley (1999) and emphasized by Basham (2003), while degeneration of a research paradigm is valuable in the hard sciences where evidence cannot fabricate itself, conspirators may produce evidence that obstructs revelation of their conspiratorial activity (a cover-up). Basham goes further, arguing that even the most degenerate conspiracy theories—those that lack evidence yet allege sprawling, influential malevolence—cannot be dismissed on epistemic grounds because a conspiracy that large would be capable of a successful cover-up. Basham does not argue that conspiracy theories of this type are non-falsifiable. Rather, he argues that falsifying them requires scrutiny of their evidence rather than merely identifying their claims as *prima facie* unlikely to be true.

Second, Clarke engages in special pleading. Cosmological science degenerates each decade (Ekeberg 2019), yet Clarke would not ask that we dismiss all work in those disciplines. Black holes, dark matter, dark energy, and adaptations to the estimated rate of expansion of the universe were each discovered long after cosmological scientists believed that their theories were correct. As each successive change to the facts manifested, scientists simply tweaked their old theories rather than casting them aside. Regardless, knowledge of and concession to consensual theories in the cosmological sciences are marks of intelligence in modern society. Degeneration is more damning in the hard sciences than outside of them, so if we tolerate extensive degeneration in the cosmological sciences, we should tolerate it in the realm of conspiracy theories.

Next, Coady (2003) argues on consequentialist grounds that there are tangible costs to adopting Keeley's or Clarke's position. Conspiracies happen, after all, and conspiracists help us prevent some of those conspiracies. Thus, even if conspiracy theories exhibit some epistemic

imperfections, consequentialism compels us to take them seriously. Coady's 2006 volume aggregates several of the essays referenced above as well as other relevant essays (Coady 2006).

Next, Mandik (2007) argues that conspiracy theories that conflict with "shit happens" (again, this is the scholarly term) explanations are *prima facie* unlikely. First, Mandik argues against a strawman, as no scholarship argues that conspiracy theories are likely. Rather, some philosophers make the modest claim that one should evaluate the evidence supporting conspiracy theories before dismissing them. Second, review of history indicates that conspiracy theories contradicting "shit happens" explanations are not as unlikely as Mandik thinks they are. Consider that most confirmed conspiracy theories were at one point rivaled by "shit happens" explanations offered by conspirators and/or their lawyers. Third, Mandik ignores consequentialist concern for stopping real conspiracies. Mandik does not propose how unlikely a conspiracy theory must be to warrant ignoring it. Nor does Mandik offer criteria by which we may assign varying likelihoods to conspiracy theories. I undergo a related effort in Chapter Two of this dissertation.

Next, Levy (2007) argues that it is unwarranted to believe conspiracy theories, but he offers no novel arguments. Stokes (2018) argues that false conspiracy theories make harmful accusations that reduce public trust. His distinction between true and false conspiracy theories concedes particularism on epistemic grounds. To justify generalism, Stokes argues on consequentialist grounds. Stokes emphasizes the harm caused by false accusations, but this harm pales in comparison to the harm caused by conspiracies like the Holocaust or the push to misrepresent evidence about WMDs in Iraq (Basham and Dentith 2018). If Stokes deferred to his theory, he would have us remain silent about rumors of genocide or a war fought on false pretenses in order to protect the feelings and reputations of potential conspirators. In practice,

Stokes would never advise this. Thus, Stokes ought to articulate more specific arguments about the likelihood and severity of conspiracy that justifies articulating it.

In addition to arguing on consequentialist groups that conspiracy theories cause harm via their accusations, Stokes argues that conspiracy theories cause harm via reduction in public trust. He argues that this erodes the foundations of democracy and science. Coady (2006) and MacIntyre (2007) argue that this perverts the norm of public trust beyond recognition. It is a virtue of citizenship in a free society to exhibit skepticism toward government.

Next, Harris (2019) concedes defenses of particularism. He shifts his focus from problems with the epistemology implicit in conspiracy *theories* to problems with the epistemology used by *conspiracists*. Harris claims that conspiracists tend to commit five epistemic errors more frequently than others. He does not provide empirical evidence to support this claim. Hagen (2020) takes Harris's concern seriously, but we need not do so. Errors committed by some conspiracists (but also committed by some non-conspiracists) have no bearing on the epistemic status of conspiracy theories nor on the reasoning abilities of conspiracists who do not commit those errors.

Lastly, Cassam (2018) returns our focus to conspiracy theories, arguing that it is important to determine the epistemic vices of conspiracy theories because of the harms caused by mass belief in those theories. This begs the question about how often conspiracy theories are true. Cassam's assumption that they are almost never true lacks evidence and serves as the foundation for his arguments. Nonetheless, in pursuit of these epistemic vices, Cassam argues that conspiracy theories are speculative, contrarian, esoteric, amateurish, and premodern. The latter four traits are not epistemic virtues or vices. In the case of each of those traits, it is a logical fallacy to claim that one can infer the likelihood that a claim is true from the claim's status with

respect to those traits. The first claim—that conspiracy theories are speculative—is an epistemic vice but requires one evaluating a conspiracy theory to consider its evidence to determine whether the conspiracy theory is speculative. Thus, Cassam mistakenly advocates for particularism. Hagen (2022) provides a more detailed rebuttal to Cassam. Among other rebuttals, Hagen points out that many explanatory rivals to conspiracy theories, including official theories, are speculative. Thus, Cassam engages in special pleading, asking us to dismiss some but not all speculative theories.

I conclude with a summarizing remark by Pigden (2018): "What is wrong with false, crazy or dangerous conspiracy theories is that they are *false, crazy or dangerous*, NOT that they are *conspiracy theories*, since many conspiracy theories are neither dangerous, crazy, nor false" (204).

Having now reviewed what scholarly opponents of conspiracy theory regard as the best arguments against particularism, one can see why research that proceeds from generalist assumptions (especially Order 2b research and Order 3b interventionism) should be taken with a grain of salt until it reconciles itself with particularism. Next, I collect Order 3b interventions that attempt to actualize generalism in the public consciousness.

ii. Interventionism (3b)

In this sub-section, I collect and scrutinize intervention into conspiracy theory discourse. First, I discuss Sunstein's and Vermeule's famous advocacy for "cognitive infiltration of the groups that produce conspiracy theories" (Sunstein and Vermeule 2009:218). Second, I discuss tests of strategies that could reduce belief in conspiracy theories. Third, I discuss fact-checkers. Fourth, I discuss the French Ministry of Education program that receives the bulk of scholarly ire against interventionism. And fifth, I scrutinize social media interventionism.

First, Sunstein and Vermeule unironically conspire to infiltrate extremist groups to reduce belief in conspiracies. Hagen (2010, 2011) critiques this proposal as epistemically and consequentially unwise and, light of the boomerang effect, counterproductive.

Second, I review tests of strategies that may reduce belief in conspiracy theories. First, Orosz et al. (2016) show that rational and ridiculing arguments reduce belief in conspiracy theories. Jolley and Douglas (2017) show that anti-conspiracy-theory arguments increased intentions to vaccinate a fictional child, but this effect only held when arguments were presented prior to conspiracy theories.

From the perspective of a generalist, Orosz et al.'s (2016) and Jolley's and Douglas's (2017) findings could justify intervention against belief in conspiracy theories before the mass public has as chance to consider the evidence for those theories. Setting aside the lack of epistemic justification for such a practice, the reader should note that whenever a conspiracy is true, interventionism would act on the behalf of conspirators. No interventionist considers this quandary. I grapple with the quandary in Chapters 2 and 3 of this dissertation.

Third, I turn to the case of fact-checkers. The consistency with which fact-checkers label conspiracy theories false indicates that fact-checkers are generalists. Uscinski and Butler (2013) articulate three flaws associated with contemporary fact-checking. First, fact-checkers have no clear criteria for selecting cases. Second, they sometimes fact-check unverifiable claims. Third, they have no clear criteria for distinguishing fact from fiction.

Amazeen (2015) responded by arguing from the fallacy of consensus that we can trust fact-checkers whenever they agree with each other. Uscinski rebuts this argument, pointing out that fact-checkers come from similar ideological backgrounds, thus presenting a confounding variable. In certain cases, fact-checkers may agree with each other not because they are *correct*

but because they share a set of biases predisposing them to be *incorrect* (Uscinski 2015). We cannot distinguish these claims to explain cases in which fact-checkers agree with each other without reviewing the evidence ourselves.

Consider a 2020 *PolitiFact* fact-check that labeled the COVID-19 Lab Leak Hypotheses "Pants on Fire." The original fact-check reads "In a Sept. 15 interview, the most-watched program on cable network television aired a conspiracy theory that has been debunked since the beginning of the coronavirus pandemic" (Funke 2020). Months later, we find an addendum from the editor:

> When this fact-check was first published in September 2020, *PolitiFact*'s sources included researchers who asserted the SARS-CoV-2 virus could not have been manipulated. That assertion is now more widely disputed. For that reason, we are removing this fact-check from our database pending a more thorough review. Currently, we consider the claim to be unsupported by evidence and in dispute. (Editor's Note 2021)

Note how the addendum attempts to normalize the consensus fallacy. The falsehood of the Lab Leak Hypothesis was always disputed, hence the existence of the hypothesis and demand for a fact-check. All that changed in the interval between the original fact-check and the addendum was the breadth of disputation. Thus, *PolitiFact* implies that when the consensus supporting its original conclusion dissolved, the epistemic status of the original claim changed. In reality, the original claim was always uncertain, not "Pants on Fire."

It appears that *PolitiFact* begins its fact-checks by deferring to consensus. Then, *Politifact* works backwards to justify the consensual position. Thus, *PolitiFact* provides no value and, when it is wrong, harms its readers by giving them undue confidence in the fact-checker's conclusions. Ironically, this is the same reasoning that leads Popper (1994), Hofstadter (1965), and others to claim that conspiracy theories are harmful. Note that Facebook outsources its fact-checking to *PolitiFact* and similar fact-checkers (Jin 2020). When Facebook censors as a

function of a fact-checker's conclusions and the fact-check proves incorrect, Facebook multiplies the harms caused by the fact-checkers.

Fourth, I turn to at attempt by the French government to use education to intervene into conspiracy theory discourse. In 2015, the French Ministry of Education created a program designed in part to convince students to adopt generalism. According to Order 3a research, generalism is not justified. When an educational program teaches unjustified dogma, that system engages in indoctrination, not education. Dentith's 2018 volume collects similar condemnation of the French program.

In more direct interventionism, YouTube in 2019 declared that it will "begin reducing recommendations of borderline content and content that could misinform users in harmful ways-such as videos…making blatantly false claims about historic events like 9/11" (YouTube Team 2019) To some, the official theory about 9/11 is blatantly false and conspiracy theories more accurately describe historic events. Who, one might ask, decides whether claims about historic events are blatantly false?

Like Facebook, YouTube defers to fact-checkers to determine which claims it steers users away from. To teach its users the truth, YouTube provides links to Wikipedia, an open-source blog with many pages that articulate dubious positions (BBC Trending 2018; Media Lens 2018). The open-source nature of Wikipedia bakes in the fallacy of consensus. Wikipedia is useful as a repository of historical facts, but it is not useful for resolving disputed claims.

While deference consensus tends to cultivate a well-informed public, it sometimes backfires for two reasons. First, the consensus of experts is often incorrect. Such is the normal course of scientific development. While interventionists updated their fact-checks as they emerge from academia and journalism, their interventionism may slow scientific progress by 1)

inhibiting the spread of counter-consensual information and 2) entrenching in the next generation of scientists and historians the epistemically unsound practice of deference to consensus. Second, interventionists ignore scholarship demonstrating the presence of a boomerang effect whereby actions meant to restrict freedom of information cause increased interest in that which was restricted (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018).

Above I review examples of interventionism and introduce the reader to associated rebuttals. Below, I discuss fourth order claims that regard the conspiracy-theory-stigma—not conspiracists or conspiracy theories—as the subject of their research.

E. Order 4

Fourth order claims regard as their subject the stigma associated with conspiracists and their conspiracy theories. Unlike endeavors on other orders, fourth order research does not depend on what one thinks about conspiracy theories nor conspiracy theorists. I discuss fourth order research below. First, I discuss un-evidenced claims about the stigma. Second, I discuss efforts to identify and measure the stigma using evidence. Third, I discuss claims about the weaponization of the labels "conspiracy theory" and "conspiracy theorist" and weaponization of the stigma associated with those labels. Fourth, I discuss claims about the causes and consequences of the stigma.

First, I discuss casual mentions of the stigma. Some scholars take the presence of the stigma as a given (Coady 2003; Dentith 2019). Building on this assumption, some scholars make claims about whether stigma is increasing or decreasing. Barkun (2015) and Basham (2018) claim that the stigma is decreasing. These four works mention the stigma but do not attempt to measure stigma.

Second, I discuss efforts to identify and measure the stigma. Some scholars use the labels "conspiracy theory" and "conspiracy theorist" in experiments to discover the conditions under which the stigma manifests. Wood and Douglas show that people tend to be unwilling to apply the label "conspiracy theory" to their own claims (Wood and Douglas 2013, 2015). This, they argue, constitutes evidence that the label, if not the concept itself, is stigmatized. Wood (2015) attempts to test the stigma associated with the "conspiracy theory" label using an experiment that varies whether a claim is referred to as a conspiracy theory. He finds that the presence of the label is not associated with expressed belief in the claim, however each of his studies features a common flaw. Neither study utilizes a concrete, real conspiracy theory, thus compromising the external validity of Wood's results.

Broadening focus from the labels to the stigma itself, Thalmann (2019) uses qualitative evidence to show that the stigma increased from the late-1960s until the mid-1990s. Klein et al. (2015) show that "…believers [in conspiracy theories] were perceived [by subjects]…as more gullible, crazy, easily influenced, stupid, naïve, manipulative, dishonest, assertive, and selfish than nonbelievers" (Abstract). Lantian et al. (2018) show that in France there is a stigma associated with conspiracy theories about the 2015 *Charlie Hebdo* shooting. They ask subjects to visualize advocating for or against those conspiracy theories in front of others. Those in the "advocate for" condition exhibit greater fear of social exclusion than those in the "advocate against" condition. Fear of social exclusion captures the essence of stigma.

Third, I discuss weaponization of the labels and/or stigma. Fiske (1999) and Bratich (2002) use qualitative evidence to show that elites condemn AIDS-as-genocide conspiracy theories to police debate and to maintain the Left's discursive authority to articulate the problems faced by Black Americans. Husting and Orr (2007) use qualitative evidence to show that actors

use the labels to "go meta." These actors weaponize the stigma to avoid engaging with the evidence supporting a conspiracist's claims. Bratich (2008) argues that just as the Christian Right orchestrates moral panics to blame the Left for society's problems, elites—especially on the Left—orchestrate conspiracy panics to blame conspiracists for polarization, echo-chambers, and misinformation.

Alexander Wendt and Raymond Duvall (2008) focus on what they call the "UFO taboo." They argue that by implying that extra-terrestrials exist, conspiracy theories about UFOs challenge human sovereignty. Thus, some scientists and governments employ the UFO taboo to defend against that challenge. One can extrapolate their case beyond the context of UFOs. One challenges the sovereignty of government when one accuses elements of government in a conspiracy theory. Because many conspiracy theories implicate many elements of government, politicians and bureaucrats are incentivized to stigmatize conspiracy theories. We observe this practice in the cases of Bush's 2001 speech at the UN, the 2005 Department of State website, and the 2015 French educational program.

DeHaven-Smith (2013) develops Wendt's and Duvall's (2008) argument by showing that the "conspiracy theory" label is applied with the greatest consistency to conspiracy theories that implicate the government. He argues that elites' tendency to avoid application of the label to official conspiratorial activity shows that the stigma and label are to some extent weapons the state uses against other states, its citizens, and citizens of other states. Harambaum and Aupers (2015) argue that the label performs "boundary work" that distinguishes the mainstream from the fringe. Barkun (2015) agrees, adding that the label is also used to distinguish authoritative from non-authoritative claims.

Fourth, some scholars address the causes and consequences of stigma. Regarding causes, Wendt and Duvall identify four causes of the stigma that, they argue, inhibit ufologists' efforts to convince others to take UFOs seriously. First, institutional authorities claim that UFOs are not a national security threat. Second, pop culture fictionalizes the UFO phenomenon and refers to ufology as pseudo-science. Third, official secrecy withholds the best evidence about the existence of UFOs from public discourse. And fourth, elites pathologize and dismiss ufology to cement their elite-status. At the time of this writing, it seems that it is no longer important for elites to dismiss ufology.

Turning from causes to consequences, DeHaven-Smith (2013) argues that the stigma discourages Americans from doubting and investigating the motives and actions of their elected leaders. This subverts the design of the founders and inhibits the intended functioning of representative democracy. Dean (2000) argues that the commonality of the "conspiracy theory" label confounds concerns about absurd conspiracy theories with pressing political concerns about corporate corruption, campaign finance, and covert governmental actions. She argues that this confounding protects everyday conspirators. Keith Goshorn (2000) agrees, arguing that the categorization of dissent as paranoia disincentivizes dissent in all its forms. Coady (2018) also agrees, arguing that reluctance to challenge authoritative explanations enables conspirators to provide weaker evidence of their innocence than we would otherwise require. This, in turn, allows more conspirators to escape detection. Raikka and Basham (2019) argue that a "conspiracy theory phobia" undermines the quality of social science research studying conspiracy theories.

Finally, Orr and Husting (2019) focus on the consequences of stigma when viewed from the perspective of marginalized groups. They analyze the legacy press, showing that elites

weaponize the "conspiracy theory" label and associated stigma to exclude racial and ethnic minorities from the "…community of reasonable participants in democratic discourse" (83). They argue that "going meta" to criticize the mode of expression rather than the substance of claims protects conspirators. Further, they argue that going meta especially harms members of marginalized communities by ignoring critiques of inequality embedded within some conspiracy theories.

Ultimately, fourth order research seems to jump the gun, addressing the uses, causes, and consequences of stigma before agreeing on how to identify and measure stigma where it lives—in political speeches, newspapers, television news, and on social media. This lack of measurement allows for disagreements about whether or not there is a stigma and whether that stigma is increasing or decreasing. Meanwhile, psychologists use advanced quantitative methods to study the presence of various stigmas on social media (Robinson et al. 2019; Li et al. 2020; Powers, Koliska, and Guha 2019 and Kaufmann 2021). Thus, researchers acting on Order 4 should collaborate with psychologists. They would bring psychologists an additional domain, and psychologists would bring forth order scholars of conspiracy theory more substantive expertise regarding stigma as well as the ability to measure stigma where it lives, not in a laboratory-setting.

IV. Conclusion

This chapter uses Luhmann's orders of observation to identify blind spots that plague the conspiracy theory literature. By classifying endeavors surrounding that literature according to Luhmannian orders of observation, I identify two main blind spots. First, social scientists attempting to measure conspiracism and identify its correlates demonstrate unsubstantiated bias against conspiracy theories and conspiracists. If they attended to third order epistemological

research that debunks their presumptions against conspiracism, they would be more willing to consider first order historical and journalistic research that demonstrates the plausibility of many conspiracies. Until they take conspiracy theories seriously, they will continue to make measurement choices that correspond more to their biases than epistemology and evidence.

Second, research in psychology yields insights that are fundamental to some endeavors in the conspiracy theory literature. Would-be censors ignore research that shows that attempts to restrict access to information tend to increase interest in that information. At the same time, scholars of the conspiracy theory stigma fail to incorporate insights and methods from psychology that would enhance their attempts to make claims about the conspiracy-theory-stigma.

Thus, from this project emerges two recommendations for professionals whose work relates to conspiracy theory. First, social scientists seeking to measure and study the correlates of conspiracism should carefully consider both the merits of their assumptions related to conspiracism and the plausibility of particular conspiracy theories. Second, social media interventionists and scholars of the conspiracy theory stigma should embrace the psychological literatures related to their subjects.

Chapter Two: Locating the Danger Associated with Conspiracy Theories

I. Introduction

Many scholars argue that mass, serious consideration of conspiracy theories endangers society (Popper 1994; Hofstadter 1965; Gilbert 1996; Post and Robins 1997; Pipes 1997; Keeley 1999; Clarke 2002; Barkun 2015; Stokes 2018; Cassam 2019; Harris 2019). This sentiment trickles into public discourse via politicians, journalists, op-ed writers, and videos viewed, shared, and discussed by millions of people (Fiske 1999, Dean 1998, 2000; Bratich 2002, 2008; Husting and Orr 2007, Wendt and Duvall 2008; deHaven-Smith 2013; Harambam and Aupers 2015; Raikka and Basham 2019). Governments, news-providers, and social media corporations respond to this supposed danger by intervening into public discourse to reduce the spread of some conspiracy theories. I refer to this practice as "interventionism" and those who practice it as "interventionists."

Notably for scholars of public opinion, opinion-change, public discourse, and freedoms of expression, the most impactful interventionists—Alphabet inc. (including Google and YouTube), Meta (including Facebook and Instagram), and Twitter (including Periscope)—censor and throttle access to posts, users, and groups that advance politically relevant conspiracy theories. The interventionists direct most of their interventionism toward conspiracy theories related to COVID-19, QAnon, and elections. Interventionism demonstrates three theoretical flaws that have implications for political science as well as corporate and public policy.

First, interventionists ignore extensive psychological research on reactance theory and its relevant manifestation in the domain of censorship: the boomerang effect (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018). Second, interventionists fail to engage with extensive

epistemological research that shows that society is better off if a certain amount of conspiracy-theory-discourse transpires and that there is nothing about conspiracy theories that makes them more dangerous than other ideas (Pigden 1995; Goshorn 2000; Clarke 2002; Coady 2006; Dentith 2014; Raikka 2017; Dentith 2018). Third, interventionists fail to compare the supposed harms of conspiracy-theory-discourse to the potential harms caused by the corresponding conspiracy.

First, interventionists ignore extensive psychological research on reactance theory and its relevant manifestation in the domain of censorship: the boomerang effect (sometimes referred to as the Streisand effect). This research argues that attempts to restrict access to information tend to backfire, causing increased interest in and access to the restricted information. Predictably, increases in censorship on YouTube, Facebook, and Twitter mirror increases in the popularity of freer social media platforms including Gab, Rumble, WeMe, Telegram, and Parler. Each of these new platforms is experiencing an influx of users from legacy platforms (Ray 2021; Hart 2021). Thus, even setting aside the boomerang effect, censorship by YouTube, Facebook, and Twitter may not reduce harm so much as displace it.

In tandem, the potential displacement effect and the boomerang effect should give interventionists pause (Whitten 2015; Menczer 2016; Carey 2017; Meserole 2018; Yaraghi 2019; Bond 2020; Suciu 2021; Myers 2022). Interventionism absolves interventionists from some moral liability and criticism, but it may not be faithful to the harm-reduction goals that motivate it. I do not test the boomerang effect nor the displacement effect in this chapter. Rather, I introduce those two notions as context that draws our skepticism toward any social media policy that attempts to reduce harm by restricting access to information.

Second, interventionists fail to engage with epistemological research that proves three compounding truths. First, some conspiracy theories are true. Second, a claim's status as a conspiracy theory does not render the claim dangerous. Third, society is better off if a certain amount of conspiracy-theory-discourse transpires (Pigden 1995; Goshorn 2000; Clarke 2002; Coady 2006; Dentith 2014; Raikka 2017; Dentith 2018). Interventionists claim that interventionism constitutes deference to experts, but they are deferring to the wrong experts. For example, while it is reasonable to consult medical authorities when assessing the soundness of claims related to COVID-19, it is unreasonable to assume that throttling access to all claims that contradict those authorities serves the public good. Social media corporations ought to consult experts in epistemology and censorship, not medicine or election-integrity, to assess the soundness of interventionism. Using this lens, I demonstrate harmful manifestations of deference to the wrong experts in Section II.

Third, interventionists fail to compare the supposed harm of conspiracy-theory-discourse to the harm that may be caused by the corresponding conspiracies. For example, accusations implied by conspiracy theories harm the accused (Stokes 2018), but that harm rarely exceeds the harm alleged by the conspiracy theory. For example, mass, serious consideration of a conspiracy theory that accuses a marginalized group imposes extensive harms on that group. Discourse of that type contributed to the Holocaust and, more recently, anti-Asian hate crimes in the United States (US). But not all conspiracy theories feature such risks. That is, mass, serious consideration of some conspiracy theories (e.g., political corruption or covert crimes by intelligence agencies) imposes small harms on powerful, nameless, faceless actors in exchange for a chance to thwart associated conspiracies. After Section II, the rest of this chapter attempts

to isolate and make use of the variables most closely associated with the danger imposed by mass, serious consideration of conspiracy theories.

A. Organization of This Chapter

To conclude the Introduction, I define key terms (the reader's own definitions are sufficient until then). Then, in Section II I outline interventionism that motivates this chapter. In Section III, I turn to the epistemological literature to isolate the variables most closely associated with the danger imposed by mass, serious consideration of conspiracy theories. In Section IV, I apply these variables to censored conspiracy theories to see whether interventionism aligns with my conclusions. Ultimately, I show that only in rare cases is it wise to discourage conspiracy-theory-discourse.

B. Definitions of Key Terms

Table 1

Term	My Definition	Flawed Alternative Definition
conspiracy	a multi-person plan or action employing morally suspect means and/or toward morally suspect ends	a *true* claim that a conspiracy is the best explanation for an outcome
conspiracy theory	a claim that a conspiracy is the best explanation for an outcome	a *false* claim that a conspiracy is the best explanation for an outcome
non-conspiratorial theory	a claim that some explanation other than a conspiracy best explains an outcome	
conspiracism	one's tendency to prefer a conspiracy theory to plausible, non-conspiratorial theories	one's tendency to prefer a *false* conspiracy theory to plausible, non-conspiratorial theories
conspiracist	one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories	
conspiracy theorist		one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories

Table 1 shows my definitions of key terms. I define "conspiracy" as a multi-person plan or action employing morally suspect means and/or toward morally suspect ends. I define a "conspiracy theory" as a claim that a conspiracy is the best explanation for an outcome. Many

plans and actions not colloquially referred to as conspiracies nonetheless qualify under my definition. For example, my definition of conspiracies includes 1) marital infidelity, 2) political corruption, 3) insider trading, and 4) the official theory of 9/11 (examples borrowed from Pigden 1995, 2006 and Dentith 2014). On the other hand, I define a "non-conspiratorial theory" as a claim that some explanation other than a conspiracy best explains an outcome. Examples of common non-conspiratorial theories include coincidences, lone actors, and institutional forces.

Under my definitions, confirmation that a conspiracy theory is accurate (confessions, DNA, fingerprints, etc.) does not cause the conspiracy theory to lose its status as such. Thus, we all believe many conspiracy theories and we are all conspiracy theorists. To allow for variance, I define "conspiracism" as one's tendency to prefer a conspiracy theory to plausible, non-conspiratorial theories, and I define a "conspiracist" as one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories.

II. Social Media Interventionism

A. Social Media Policies

Most social media policies related to intervention are logical, prudent, and resemble traditional exceptions to protections of freedoms of expression. For example, major social media corporations remove posts that threaten violence or organize harmful activities. They ban users who repeatedly violate these policies. But if we scrutinize removal of posts on the grounds of misinformation, we begin to observe vague policies, subjective enforcement, and uncertain benefits. I introduce the relevant policies articulated by Facebook, YouTube, and Twitter below.

Facebook says that "We remove misinformation where it is likely to directly contribute to the risk of imminent physical harm. We also remove content that is likely to directly contribute

to interference with the functioning of political processes and certain highly deceptive manipulated media" (Meta Transparency Center: "Misinformation").

YouTube says that "Certain types of misleading or deceptive content with serious risk of egregious harm are not allowed…. This includes certain types of misinformation that can cause real-world harm, like promoting harmful remedies or treatments, certain types of technically manipulated content, or content interfering with democratic processes" (YouTube Help: "Misinformation").

Twitter says that "You may not use Twitter's services to share false or misleading information about COVID-19 which may lead to harm" (Twitter Help Center: "COVID-19") and that "We will label or remove false or misleading information intended to undermine public confidence in an election or other civic process" (Twitter Help Center: "Civic"). Twitter has a separate set of policies for "synthetic and manipulated media" that target the harms imposed by spammers and bots, especially in relation to political elections (Twitter Help Center: "Synthetic").

Each platform exhibits similar language and focuses on similar areas of misinformation: COVID-19-related information, elections-related information, and utter fabrications. Each platform uses similar standards to inform its practices: Facebook uses a "likely to directly contribute to[…]harm" standard; YouTube uses a "with serious risk of egregious harm" standard, and Twitter uses a "which may lead to harm" standard. Thus, it is appropriate to consider these platforms in tandem.

B. Interventionism in Practice

i. Interventionism Outside of Social Media

Social media corporations are not the first entities to attempt to moderate public discourse. There is an extensive history of governments intervening into public discourse. That history recalls attempts to steer discourse toward national interests defined by elites who—like social media executives—are unaccountable to the ballot box. In the US, this history includes the Committee on Public Information during World War I ("Complete Report"; Hedges 2010), the Writers' War Board during World War II (Howell 1997), and various programs exposed by the Church Committee investigations in 1975 ("Hearings" 1975; US Senate, Book III 1976; Bernstein 1977).

In the Internet Age, governmental interventionism takes new forms, but the goals are similar. In 2003, US Marines staged the toppling of Saddam Hussein's statue to make it appear as if Iraqis toppled it themselves in an expression of support for American occupation (Zucchino 2004; The Bryant Park Project 2008). From 2004-2005, various elements of the Bush regime acted to prevent journalist James Risen from reporting on Operation Stellar Wind, the NSA program of warrantless wiretapping of American citizens (Risen 2018). In 2005, the US Department of State warned citizens against taking conspiracy theories seriously (US Department of State 2005). During the War on Terror, the Department of Defense spent 500 million dollars to produce, disseminate, and track the viewership of fake Al Qaeda videos that depicted the group in a negative light (Garcia 2016). In 2009, Sunstein and Vermeule advocated for "cognitive infiltration of the groups that produce conspiracy theories"—a government-conspiracy to convince conspiracy theorists that there are no government-conspiracies (Sunstein and Vermeule 2009:218). Lastly, in 2015, the French Ministry of Education created a program seeking to help students distinguish "that deleterious and deadly suspicion that feeds

[conspiracism]" from "that fair distance that establishes an informed gaze" (National Education 2015).

Each of these programs assumes that the masses lack the ability to reason well, so elites must steer those masses toward appropriate conclusions (Goshorn 2000). This mindset seems to be active at Alphabet Inc., Meta, and Twitter. I explore manifestations of that mindset in the remainder of Section II.

ii. Search Engine Manipulation by Google

In the wake of Russian interference into the 2016 US Presidential Election and the fake-news-panic that ensued, Google changed its search algorithm "to surface more authoritative content" (Gomes 2017). In 2017, the World Socialist Website (WSWS) revealed that Google was throttling access to progressive, socialist, and/or anti-war websites. While Google searches for the phrase "Leon Trotsky" produced 5,893 results from WSWS in May of 2017, the same phrase yielded *zero* of those results in June. The same pattern emerged with 145 of the 150 search terms on which WSWS relies most for its traffic. Eleven similar websites simultaneously experienced severe reductions in traffic on Google (North 2017). In the years that followed, social media joined this controversial practice.

iii. Social Media Intervention Regarding Health Misinformation

Facebook censors "health misinformation likely to directly contribute to imminent harm to public health and safety," which includes claims that "Vaccines or their ingredients are deadly, toxic, poisonous, harmful, or dangerous" (Meta Transparency Center: "Misinformation"). To make judgments, Facebook defers to third-party fact-checkers such as *PolitiFact* (Jin 2020). These fact-checkers tend to be epistemically naïve. They have no clear selection criteria to determine which claims they check, they often fact-check unverifiable claims, and they have no

consistent methodology to guide their rulings (Uscinski and Butler 2013; Uscinski 2015). When these fact-checks differ from claims made in Facebook posts, Facebook flags those claims and reduces their distribution (Meta Transparency Center: "How" 2022). In 2020, Facebook claimed that it labeled about 40 million posts based on about 4,000 fact-checks. They showed that those who saw the labels chose to view the associated content only 5% of the time. Thus, labeling combined with search algorithm manipulation form a potent duo (Rosen 2020). Below, I problematize these policies by considering a case related to the COVID-19 Lab Leak Hypothesis.

Consider a 2020 *PolitiFact* fact-check that labeled the COVID-19 Lab Leak Hypotheses "Pants on Fire." The original fact-check reads "In a Sept. 15 interview, the most-watched program on cable network television aired a conspiracy theory that has been debunked since the beginning of the coronavirus pandemic" (Funke 2020). Months later, we find an addendum from the editor:

> When this fact-check was first published in September 2020, PolitiFact's sources included researchers who asserted the SARS-CoV-2 virus could not have been manipulated. That assertion is now more widely disputed. For that reason, we are removing this fact-check from our database pending a more thorough review. Currently, we consider the claim to be unsupported by evidence and in dispute. (Editor's Note 2021)

Hence, the epistemic status of the Lab Leak Hypothesis changed from "debunked" and "Pants on Fire" to "unsupported by evidence and in dispute." Scrutiny of the language in the Editor's Note reveals epistemically unsound reasoning at *PolitiFact*.

First, the addendum attempts to normalize the consensus fallacy. The mere existence of the Lab Leak Hypothesis proves that associated claims were always in dispute. The breadth of disputation, not its presence and absence, changed in the interval between the original fact-check and the addendum. Thus, *PolitiFact* implies that when the consensus supporting its original conclusion dissolved, the epistemic status of the original claim changed. In reality, the original claim was always uncertain, not "Pants on Fire."

Second, plenty of sources have always argued that the Lab Leak Hypothesis was plausible. Given *PolitiFact*'s decision to fact-check the Lab Leak Hypothesis in the first place, it is not plausible that they were unaware of these sources. This demonstrates that *PolitiFact* chooses its sources—those that "debunked [the Lab Leak Hypothesis] since the beginning of the coronavirus pandemic"—as a function of agreement with a predetermined conclusion. Further, *PolitiFact*'s claim that the matter "is now more widely disputed" concedes that the matter was always disputed. *PolitiFact* does not offer a disputation-threshold short of which claims earn a "Pants on Fire" rating. Once the governments of the US and UK began to take the hypothesis seriously, Facebook remained in lockstep with *PolitiFact* by ceasing its efforts to decrease discussion of the Lab Leak Hypotheses (Sayers 2021). While deference to large majorities and subject-area experts will usually produce sound conclusions, such deference is antithetical to the purpose of a fact-checker.

Next, regarding YouTube, the case of biologist Bret Weinstein's YouTube channel demonstrates censorship of medical misinformation. Weinstein scrutinized potential shortcomings of Moderna's and Pfizer's vaccines with Robert Malone, a doctor who contributed to the invention of the mRNA technology used in Moderna's and Pfizer's COVID-19 vaccines (Ferreira 2021). During the interview, Weinstein claimed that the spike proteins in the vaccines were "very dangerous," and Malone elaborated this claim without disagreeing (Weinstein 2021). In response to this interview, YouTube informed Weinstein that he must remove the interview or face permanent demonetization of his channel (Taibbi 2021).

All the while, many journalists and scientists publicly discussed the weaknesses of the vaccines including the blood clots associated with the Johnson & Johnson vaccine (Robins and Jewett 2021; LaFraniere 2022) and the waning effectiveness of vaccines over time (Mandavilli

2021; Zimmer 2021). The day after Weinstein's interview with Malone aired, the CDC announced that it would convene a meeting of its advisers to discuss rare instances of heart inflammation linked to Pfizer's and Moderna's mRNA vaccines for COVID-19 (Tin 2021). Additionally, multiple academic studies discuss the potential harms of spike proteins in particular (Badshah et al. 2021; Chouchana et al. 2021; Istampoulouoglou et al. 2021; Lim et al. 2021; Matta et al. 2021; Habedank et al. 2022). Thus, the interview constituted discussion of unsettled science, not dangerous contradiction of fact. Claims made in the interview were speculative and should not have qualified as misinformation. Censoring speculation inhibits scientific progress.

iv. Social Media Intervention Regarding Misinformation That Could Incite Violence

In 2019, an FBI local field office in Phoenix, Arizona circulated an intelligence bulletin about conspiracy theories. The bulletin argued that there is a relationship between violence and belief in certain conspiracy theories (FBI Phoenix Field Office 2019). *Yahoo News* obtained the document and circulated it publicly (Winter 2019). I reproduce an excerpt below:

> The FBI assesses anti-government, identity based, and fringe political conspiracy theories very likely [that is, with an 80-95% chance (pp. 10),] will emerge, spread, and evolve in the modern information marketplace over the near term, fostering anti-government sentiment, promoting racial and religious prejudice, increasing political tensions, and occasionally driving both groups and individuals to commit criminal or violent acts. (FBI Phoenix Field Office:5)

The document cites several academic articles and anecdotes that attempt to pathologize conspiracism while ignoring associated rebuttals (Pigden 1995; Goshorn 2000; Clarke 2002; Coady 2006; Dentith 2014; Raikka 2017; Dentith 2018). The document also ignores three rebuttals to claims it articulates or implies.

First, the document offers no reason that its assessment is limited to anti-government, fringe conspiracy theories rather than all ideas that are anti-government and fringe. Second, the

article does not justify its assumption that conspiracism causes violence rather than violent tendencies causing conspiracism. One's interpretation of the association between conspiracism and violence influences the set of appropriate responses. Third, the document does not compare the potential harm of violent acts by conspiracists with the potential harm of conspiracies. Despite these rebuttals, interventionists began censoring QAnon, a fringe, anti-government conspiracy theory, soon after the FBI document circulated.

In July 2020, Twitter removed more than 7,000 QAnon accounts "for breaking [Twitter's] rules against platform manipulation, spam or ban evasion" (Collins and Zadrozny 2020). This action constitutes necessary maintenance for any social media corporation. At the same time, however, Twitter announced that it will also "stop recommending accounts and content related to QAnon," an action that affected about 150,000 accounts (Collins and Zadrozny 2020).

Next, on January 8, 2021, Twitter banned President Donald Trump "due to the risk of further incitement of violence" following his implied support for the January 6 storming of the US Capitol (Twitter 2021). Twitter does not claim that Trump incited violence in the two tweets that caused his removal. Rather, the tweets "can be mobilized by different audiences, including to incite violence, as well as in the context of the pattern of behavior from this account in recent weeks" (Twitter 2021). One could correctly claim that almost any tweet "can be mobilized by different audiences," yet it is against Twitter's interest to ban almost all of its users. Thus, this new standard compels Twitter to arbitrarily enforce its policies.

A few days after banning Trump, Twitter banned over 70,000 accounts that "were engaged in sharing harmful QAnon-associated content at scale and were primarily dedicated to the propagation of this conspiracy theory across the service" (Twitter Safety 2021). This action

was accompanied by an update to Twitter's policies related to coordinated harmful activity. The updated policy codifies guilt-by-association by stating that "…we assess groups, movements, and campaigns and then take enforcement action on any accounts which we identify as associated with those entities" ("Coordinated Harmful Activity" 2021). That is, Twitter did not ban these accounts due to incitement of violence but rather due to the proximity of these accounts to offending accounts. One could correctly claim that almost any user is "associated with" an entity that propagates conspiracy theories related to QAnon. Thus, again this new standard compels Twitter to arbitrarily enforce its policies.

By updating this policy at the same time that they banned the accounts, Twitter appears to modify its policies *ad hoc* to align with its preferred practices rather than using reasonable policies to guide its practices. If Twitter were the Federal Congress or a US State, this practice would violate the constitutional prohibition of the passage of *ex post facto* laws that retroactively punish actions that were legal when originally undertaken (Wex: "Ex post facto").

Additionally, Twitter fails to apply its new policy to anti-corporate or anti-religious conspiracy theories even when they use violence. Regarding anti-corporate conspiracy theories, the FBI warns that eco-terrorism causes millions of dollars of damage (Brown 2019). Nonetheless, Twitter does not throttle environmentalists' articulations that corporations conspire to cover up environmental degradation. Regarding anti-religious conspiracy theories, Twitter does not throttle access to claims that groups of Muslims conspire to cause harm (e.g., the official theory of 9/11) despite the existence of Islamophobic hate crimes (Alfonesca 2021). Selective enforcement of its policies reveals bias against any claims for which Twitter does invoke its policies. Next, I turn from Twitter to Facebook and YouTube.

After the FBI document circulated but before January 6, 2021, YouTube emphasized its misinformation policies to throttle access to and ban QAnon-content (YouTube Team 2020). After January 6, 2021, YouTube intervened further, and Facebook also began to intervene against QAnon. Facebook identified QAnon as a "violence-inducing conspiracy network" and removed thousands of pages, groups, and profiles (Meta: "An Update"). YouTube shifted its emphasis from misinformation to policies related to hate and harassment, identifying QAnon-content as "content that targets an individual or group with conspiracy theories that have been used to justify real-world violence" (YouTube Team 2020). This echoes Twitter's guilt-by-association standard as well as its reliance on *post facto* policy-tweaks to justify *ad hoc* decisions. Like Twitter's policy, YouTube's policy allows for selective enforcement, as hundreds of popular ideas "have been used to justify real-world violence" yet few face censorship. In the absence of transparent adjudication of the relative harms imposed by, for example, QAnon, eco-terrorism, and Islamophobia, censorship of one without censorship of the others will remain dubious.

v. Social Media Intervention Regarding Electoral Misinformation

In the weeks preceding the 2020 US Presidential Election, the *New York Post* reported that:

> "Hunter Biden introduced his father, then-Vice President Joe Biden, to a top executive at a Ukrainian energy firm less than a year before the elder Biden pressured government officials in Ukraine into firing a prosecutor who was investigating the company, according to emails obtained by *The Post*." (Morris and Fonrogue 2020)

Facebook and Twitter responded by severely limiting the ability of the article to spread via their platforms (Paul 2020; Manskar 2020; Devine 2021). Twitter blocked *The Post*'s account for 17 days (Sargent 2020). This marked the first time that Twitter throttled access to a legacy news entity (Paul 2020). To support its actions, Twitter cited its rule against the distribution of

materials acquired via hacking (Manskar 2020), but the materials in *The Post* article were not obtained via hacking (*Post* Editorial Board 2021). Ultimately, then-CEO Jack Dorsey claimed at a Congressional Hearing that Twitter's intervention was a mistake (Manskar 2021).

Facebook, meanwhile, cited worries about the veracity of the information. But according to its Policy Communications Director Andy Stone, these worries were based on mere suspicion, not the results of a formal investigation (@andymstone 2020). Throttling access to posts before investigating their veracity follows policy at Facebook so long as they "have signals that a piece of content is false" (Rosen 2019).

Because the Hunter Biden story is legitimate, the interventionists caused the election interference they intended to thwart. Given the close margins of the election, it is possible that this interventionism altered the outcome. Thus, regardless of how effectively interventionism achieves its goals, the stakes are clear.

C. Concluding Section II

In conclusion, each of the above anecdotes demonstrates the difficulties associated with leadership of a major social media platform. While leaders of these platforms fixate on the harm they could cause by allowing the spread of misinformation on their platforms, they ignore the harms they could cause by intervening erroneously and/or inconsistently. Further, they open themselves to criticism when they adjust their policies to align with their decisions. Perhaps most importantly, none of these interventionist programs grapple with the boomerang effect whereby these attempts are likely to increase the very harms they seek to diminish (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018).

III. Dangers Associated with Mass, Serious Consideration of Conspiracy Theories

Since World War II, scholars have attempted to locate the danger associated with mass, serious consideration of conspiracy theories. Few of those scholars, however, frame their inquiries in terms of locating danger. Below, I collect and review scholarship that, intentionally or otherwise, claims to locate danger. These claims cluster into three groups: epistemic claims, consequentialist claims, and claims that are both epistemic and consequentialist.

Epistemic claims assume that mass, serious consideration of a false conspiracy theory is dangerous because belief in false claims harms the believer. They claim to locate conditions under which we may infer that a conspiracy theory is false. Consequentialist claims hypothesize that mass, serious consideration of a conspiracy theory is dangerous when belief in that conspiracy theory is likely to motivate the believer to cause harm.

Below, I review sixteen hypotheses that some scholars claim locate the epistemic and/or consequentialist dangers associated with mass, serious consideration of a false conspiracy theory. I also review relevant rebuttals to those hypotheses. I begin with purely epistemic hypotheses, then I turn to consequentialist hypotheses, and I conclude with hypotheses that invoke epistemic and consequentialist grounds. For each hypothesis, I first introduce its origins, and then I evaluate its soundness. I assess each hypothesis on empirical and logical grounds.

On empirical grounds, I discuss accurate conspiracy theories that the hypothesis would have us dismiss. On logical grounds, I set aside counterexamples to discuss whether one may reasonably dismiss some hypothetical or real conspiracy theories with reference only to the hypothesis in question. If, however, I show that the amount of research required to assess the conspiracy theory with reference to the hypothesis is roughly the same as the amount of research required to assess the conspiracy theory *without* regard for the hypothesis, then I reject that hypothesis for want of utility. If I find that a hypothesis is plausible, I explain how I will go

about evaluating a conspiracy theory with respect to that hypothesis. I summarize the results in Table 2.

A. Epistemic Hypotheses (H1-H10)

i. Keeley's Unwarranted Conspiracy Theories (H1-H5)

Keeley (1999) argues that while conspiracies sometimes happen, it is unwarranted to believe what he calls "unwarranted conspiracy theories" (UCTs). Keeley defines UCTs as conspiracy theories that conflict with authoritative accounts, allege nefarious intentions, tie together seemingly unrelated events, allege secrecy among conspirators, and rely for their persuasive force on their ability to explain errant data. Watergate and the Iran-Contra affair both qualified as UCTs when they were first articulated. Thus, as a caveat in his argument, Keeley argues on consequentialist grounds that believing in UCTs before they are proven erodes the public trust that is essential to the functioning of modern society. Keeley has since adapted his position away from all of the above arguments.

Hence: H1) Conspiracy theories are unlikely to be true in proportion to the extent to which they disagree with authoritative claims; H2) Conspiracy theories are unlikely to be true in proportion to the nefariousness of the intentions they allege; H3) Conspiracy theories are unlikely to be true in proportion to the extent to which they correlate seemingly unrelated events; H4) Conspiracy theories are unlikely to be true in proportion to the secrecy they allege among conspirators; and H5) Conspiracy theories are unlikely to be true in proportion to the extent to which they rely on errant data. I consider Keeley's argument about public trust when I review consequentialist hypotheses below in sub-section B of this section. I explore H1 through H5 below.

Table 2

Type	Name	Claim	Plausible
Epistemic	H1: Authoritativeness	Conspiracy theories are unlikely to be true in proportion to the extent to which they disagree with authoritative claims.	No
	H2: Nefariousness	Conspiracy theories are unlikely to be true in proportion to the nefariousness of the intentions they allege.	No
	H3: Correlation	Conspiracy theories are unlikely to be true in proportion to the extent to which they correlate seemingly unrelated events.	Yes
	H4: Secrecy	Conspiracy theories are unlikely to be true in proportion to the secrecy they allege among conspirators.	No
	H5: Errant Data	Conspiracy theories are unlikely to be true in proportion to the extent to which they rely on errant data.	No
	H6: Degeneration	Conspiracy theories are unlikely to be true in proportion to the degeneration of their research paradigm.	Yes
	H7:Defectibility	Conspiracy theories are unlikely to be true in proportion to their defectibility.	No
	H8: Speculation	Conspiracy theories are unlikely to be true in proportion to the extent to which they rely on speculation.	No
	H9: Esotericism	Conspiracy theories are unlikely to be true in proportion to the esotericism of their claims.	Yes
	H10: Amateurishness	Conspiracy theories are unlikely to be true in proportion to the amateurishness of their creators.	No
Consequentialist	H11: Persuasiveness	*False* conspiracy theories are dangerous in proportion to their persuasiveness.	Yes
	H12: Crowding Out	*False* conspiracy theories are dangerous in proportion to the extent to which they crowd out discussion of more tractable problems.	Yes
	H13: Medical Advice	*False* conspiracy theories are dangerous in proportion to the extent to which believing them could cause one to diverge from authoritatively advised medical behavior.	Yes
	H14: Public Trust	Conspiracy theories are dangerous in proportion to in proportion to the extent to which their mass uptake erodes public trust.	No
Both	H15: # of Conspirators	Conspiracy theories are dangerous in proportion to the number of alleged conspirators.	Yes
	H16: Marginalization	Conspiracy theories are dangerous in proportion to the extent to which they accuse or scapegoat a marginalized group.	Yes

I begin with H1: Conspiracy theories are unlikely to be true in proportion to the extent to which they disagree with authoritative claims (Keeley 1999). This hypothesis is implausible on empirical and logical grounds.

Empirically, authoritative claims, especially the claims of institutional rather than epistemic authorities (e.g., politicians rather than scientists), are often false. We observe as much in the cases of President Richard Nixon's denial of Watergate's allegations and the conspiracy by the regimes of President George W. Bush and Prime Minister Tony Blair to manufacture consent for the second Iraq War. These authorities intentionally misled the public.

Logically, H1 encounters two problems. First, H1 relies on the fallacy of authority, thus rendering it illogical. Second, no scholar would ask that we defer to the judgements of authorities when those judgments contradict each other. Doing so is nonsensical, and yet the judgments of authorities commonly contradict, such as when Democratic and Republican members of Congress disagree about the effects of a policy. Thus, I reject H1.

H2: Conspiracy theories are unlikely to be true in proportion to the nefariousness of the intentions they allege (Keeley 1999). Because conspirators tend to offer "greater good" rationales, I assume that nefariousness is in the eye of the beholder, not the eye of the alleged conspirator. Otherwise, very few conspiracy theories would allege nefarious intentions. With this caveat in mind, H2 is implausible on empirical and logical grounds.

Empirically, Basham (2003) points to conspiratorial attempted genocides in the historical record to argue that we assign a non-negligible prior probability to the possibility that malevolent conspiracies currently exist. Further, Bratich (2008) reminds us that many *non-conspiratorial* theories allege nefarious intentions, yet few scholars attempt to debunk those articulations as a function of the nefariousness they allege. For example, consider President George W. Bush's articulation of North Korea, Iran, and Iraq as part of "an axis of evil" (Bush 2002). Keeley and other opponents of conspiracy theories fail to justify the isolated demand that conspiracy theories avoid alleging nefarious intent.

Logically, there is nothing about nefarious activity that conflicts with likelihood. We encounter proven nefarious activity in the news every day, whether it is insider trading, price-fixing, political corruption, murder, etc. Thus, I reject H2.

H3: Conspiracy theories are unlikely to be true in proportion to the extent to which they correlate seemingly unrelated events (Keeley 1999). This hypothesis is implausible on empirical grounds but is plausible on logical grounds.

Empirically, many conspiracy theories that correlate several seemingly unrelated events are both accurate and important. For example, the Iran-Contra affair correlates the sale of arms to Iran, rescuing hostages in Lebanon, and funding of the Contras in Nicaragua, yet no historian argues that this conspiracy took place.

Logically, however, if the number of seemingly unrelated events crosses some threshold, the conspiracy theory under question becomes so specific with so many potential points of failure that it would be reasonable to dismiss it. This number must exceed both the number of events affected by criminal organizations such as drug cartels as well as the number of events invoked by a rival theory. Consider a superconspiracy theory alleging that we may attribute most of the course of world history to coordination within and between large institutions. Included among these conspiracy theories is the most robust articulation of the antisemitic conspiracy theory. This conspiracy theory correlates more events than 1) law enforcement correlates when they prosecute a drug cartel and 2) opponents of the conspiracy theory correlate when they explain world history without the antisemitic conspiracy theory. Thus, I fail to reject H3 on logical grounds.

Because I fail to reject H3, I must explain how I will evaluate a conspiracy theory with regard to this hypothesis. The extent to which a theory relies on the correlation of seemingly

unrelated events engages two distinct components. First, I must consider the number of correlated events, and second, I must consider the extent to which those events seem correlated. One should only discredit a conspiracy theory if it correlates dozens of events and/or it correlates at least a dozen events that appear *completely* unrelated. Thus, I could not dismiss an articulation of the Iran-Contra affair because it correlates only three events, and each event seems at least somewhat related to at least one of the other two events.

H4: Conspiracy theories are unlikely to be true in proportion to the secrecy they allege among conspirators (Keeley 1999). This hypothesis is implausible on empirical grounds but is plausible on logical grounds.

Empirically, many counterexamples demonstrate the use of secrecy in conspiracies. First, consider the Manhattan Project, which utilized morally suspect means (secretive spending of taxpayer dollars) toward morally suspect ends (indiscriminate murder). The Manhattan Project featured a team of approximately 129,000 workers, and its existence was not revealed until the first bombs were dropped in Japan (Jones 1985). Second, approximately 30,000 employees of the NSA participated in warrantless wiretapping of American citizens for years. Despite the large number of conspirators and high significance of the conspiracy, few leaks emerged, and initial leaks were stifled (Risen 2018). Further, while Edward Snowden's whistleblowing would seem to demonstrate the failure of conspiracies reliant on secrecy, his punishment demonstrates why H4 is flawed, as subsequent prospective whistleblowers are less likely to blow the whistle than Snowden was.

Logically, however, it is plausible that once the alleged number of conspirators reaches a certain threshold, the allegation of secrecy dooms the conspiracy theory (Grimes 2016). To confirm this, consider the plausibility of a conspiracy theory alleging that half of the world is

keeping a secret from the other half. Such a conspiracy theory will never be true because no institution can maintain secrecy among such a large group of conspirators.

Because I will later consider a hypothesis related specifically to the number of alleged conspirators, I fail to reject *that* hypothesis (H15) and reject *this* one (H4). It is not secrecy but the *scale* of secrecy that makes some conspiracy theories unlikely. I discuss H15 at greater length in sub-section B of this section.

H5: Conspiracy theories are unlikely to be true in proportion to the extent to which they rely on errant data (Keeley 1999). This hypothesis is implausible on empirical and logical grounds.

Empirically, in many cases the best argument in favor of an accurate conspiracy theory relies on errant data—observations that are not accounted for by the official theory. Errant data (e.g., tape preventing a door from locking in the Watergate building) was fundamental to the case against the Watergate conspirators. Additionally, the House Select Committee on Assassinations relied on errant data—acoustical evidence indicating the presence of two gunmen—to conclude that President John F. Kennedy (JFK) was likely the victim of a conspiracy (HSCA 1979).

Logically, H5 demonstrates two flaws. First, consider the claim by Keeley (1999) and Harris (2019) that a conspiracy theory's ability to explain errant data does not prove that the conspiracy theory is correct. Hagen (2020) rebuts that Keeley and Harris confound proof with evidence. They are correct that the ability to explain errant data does not *prove* that a conspiracy theory is preferable to a rival theory. Rather, the ability to explain errant data constitutes *some evidence in favor of* the conspiracy theory. Thus, contrary to H5, the more errant data a conspiracy theory explains, the *more* likely a conspiracy theory is to be true.

Second, Keeley and Harris engage in special pleading. The ability to explain errant data is considered a merit of novel scientific explanations competing against entrenched theories. Keeley and Harris articulate no reason to divert from that reasoning in the case of conspiracy theories. Thus, I reject H5.

ii. Degeneration (H6)

Next, to refine Keeley's (1999) argument, Clarke (2002) seeks criteria that allow him to dismiss some conspiracy theories without dismissing confirmed conspiracies like Watergate and the Iran-Contra affair. Toward this end, he focuses on whether the conspiracy theory constitutes a degenerate research paradigm (Lakatos 1970). A conspiracy theory becomes degenerate when successive revelations of relevant information contradict or fail to corroborate the claims of the conspiracy theory. Henceforth emerges H6.

H6: Conspiracy theories are unlikely to be true in proportion to the degeneration of their research paradigm. This hypothesis is implausible on empirical grounds but plausible on logical grounds.

Empirically, Basham (2003) articulates a rebuttal against this hypothesis by pointing to the existence of cover-ups. While degeneration of a research paradigm dooms theories in the hard sciences because data in the hard sciences cannot fabricate itself, the same logic does not apply to conspiracy theories. In a cover-up, authorities disseminate data meant to convince onlookers that there is no conspiracy. In reality, that data is part of the conspiracy. Empirically, conspirators have a long track record of making conspiracy theories appear degenerative when they are not.

Logically, however, there must be a threshold beyond which degeneration is a valid criterion for deeming a conspiracy theory false, at least where investigators actively seek

confirmatory evidence (Keeley 2003). Thus, I fail to reject H6. Below, I use the example of QAnon to articulate the manner in which I will assess conspiracy theories with respect to H6.

Consider the case of QAnon and associated conspiracy theories. According to journalist Mike Rothschild (2021), the online profile attributed to Q has made over ten clear and false predictions. In some of those cases, Q claimed *post facto* that he knew those predictions would not come true. Such obvious degeneration seems like a reasonable indicator of falsehood of the overall research paradigm. This does not mean that we can dismiss all conspiracy theories endorsed by Q. Doing so would constitute the association fallacy whereby we dismiss X due to association with Y. Rather, we may only dismiss Q's predictions on an individual basis. Further, we may only dismiss predictions that meet two criteria. First, the prediction must have had an opportunity to manifest. Second, the prediction must have failed to manifest.

iii. Defectibility (H7)

Next, Pigden (2018) proposes a variable to assess the likelihood of a conspiracy theory: defectibility. He states that "A conspiracy theory is defectible if the costs of defection are low and the rewards are high" (209). He argues that defectible conspiracy theories are unlikely to be true because conspirators are incentivized to reveal the conspiracy before it actualizes its goals. Henceforth emerges H7.

H7: Conspiracy theories are unlikely to be true in proportion to their defectibility. This hypothesis is implausible on empirical and logical grounds.

Empirically, bribery and threat are powerful forces used throughout history. The presence of one of those forces would reduce the defectibility of a conspiracy theory to zero. No historian disputes that bribery and threat exist.

Logically, one must estimate the probability of bribery and threat in order to investigate defectibility. To estimate the probability of bribery and threat, one must conduct about as much research as would be required to assess a conspiracy theory without regard for defectibility. Thus, H7 is unhelpful toward concluding that a conspiracy theory is false, so I reject H7.

iv. Cassam's "Conspiracy Theories" (H8-H10)

Next, Cassam (2019) defines "Conspiracy Theories" as conspiracy theories that are speculative, contrarian, esoteric, amateurish, and premodern. He argues that Conspiracy Theories, not mere conspiracy theories, are unlikely to be true. Contrarianism is similar to H1 (disagreement with authority) and premodernism to H2 (nefariousness), so I do not endow those claims with their own hypotheses. The other three claims, however, warrant their own hypotheses and further exploration. To Cassam, a conspiracy theory is speculative when it relies on conjecture rather than solid fact. When a conspiracy theory is esoteric, it proposes a conclusion that is far from obvious. And when a conspiracy theory is amateurish, its creators are not authorities. Henceforth emerge H8, H9, and H10.

H8: Conspiracy theories are unlikely to be true in proportion to the extent to which they rely on speculation (Cassam 2019). This hypothesis is implausible on empirical and logical grounds.

Empirically, many official theories rely on speculation, including the official theories about 9/11 and the assassination of JFK. Regarding 9/11, Osama bin Laden was assassinated before he could stand trial, and confessions related to 9/11 were elicited via torture and thus lack epistemic value. Regarding JFK, Lee Harvey Oswald was assassinated before he could stand trial. Cassam (2019) offers no evidence to support his assumption that conspiracy theories rely on speculation more than their official rivals.

Logically, to estimate the speculativeness of a conspiracy theory, one must conduct about as much research as would be required to assess a conspiracy theory without regard for speculativeness. Thus, H8 is unhelpful toward concluding that a conspiracy theory is false, so I reject H8.

H9: Conspiracy theories are unlikely to be true in proportion to the esotericism of their claims (Cassam 2019). This hypothesis is implausible on empirical grounds but plausible on logical grounds.

Empirically, the existence of esoterically bonded elites compromises H9. Hundreds of elites including former Presidents of the US, scientists, and entertainment icons meet every year at the Bohemian Grove club in Northern California. There, in the setting of a two-week sleepaway camp, they socialize, relax, and stage a ceremonial play called *The Cremation of Care* in front of a 30-foot concrete owl. The encampment is guarded by ex-military personnel, and no women are allowed at night (Wikipedia: "Bohemian Grove"). The mere proximity of elites of such stature to such esoteric behaviors provides a benchmark for esotericism short of which we should not dismiss a conspiracy theory. Other groups of esoterically bonded elites include pedophiles among the Catholic Church, the Jacob Frank's cult in mid-1700s Europe, the Thule Society of early-1900s Germany, and participation in Marina Abramovic's Spirit Cooking by the Clintons and Podestas.

Logically, H9 has one weakness and one strength. Regarding the weakness, one's evaluation of the esotericism of a claim follows from the prior probability one assigns to conspiracy theories. What appears esoteric to one may not seem esoteric to another. Thus, H9's subjectiveness inhibits its usefulness.

Regarding the strength of H9, it is reasonable to dismiss conspiracy theories once their claims surpass a threshold of esotericism that goes far beyond what we observe in the cases mentioned above. I present an example of such a claim below. With this caveat in mind, I fail to reject H9.

To assess conspiracy theories with respect to H9, I disregard any esotericism related to a conspiracy theory's status as such and instead focus on assessing the esotericism of the allegations of each conspiracy theory. Here, I provide one example each for a conspiracy theory that has no esotericism, one with moderate esotericism, and one with maximal esotericism. An allegation of insider trading demonstrates no esotericism. An allegation that many (but far from most) elites participate in occult rituals demonstrates moderate esotericism on par with that demonstrated by Bohemian Grove, the Catholic Church, the Frankists, the Thule Society, and Spirit Cooking. For an example that demonstrates maximal esotericism, consider the claim (made by conspiracist David Icke and others) that many elites are shape-shifting reptilian creatures from the Earth's bowels, another planet, or another dimension. This claim is *so esoteric* that it can be dismissed as a function of its esotericism alone. Few conspiracy theories reach this level of esotericism. Next, I turn to H10.

H10: Conspiracy theories are unlikely to be true in proportion to the amateurishness of their creators (Cassam 2019). This hypothesis is implausible on empirical and logical grounds.

Empirically, modern standards would regard seminal Renaissance scientists as amateurs, yet the work of those scientists constitutes the foundation of much modern science. Additionally, in modern society, endorsement of a conspiracy theory before it is confirmed renders one an amateur (Coady 2003; Dentith 2019; Thalmann 2019; Klein et al. 2015; Lantian et al. 2018; Fiske 1999; Bratich 2002, 2008; Husting and Orr 2007; Wendt and Duvall 2008; deHaven-Smith

2013; Barkun 2015). Thus, H10 would render all unproven conspiracy theories highly amateurish and therefore highly unlikely. We know that some unproven conspiracy theories are accurate, so H10 fails on empirical grounds. Lastly, consider the case of investigative journalism covering Watergate. Journalists Robert Woodward and Carl Bernstein were experts on journalism, but they were amateurs in criminology. Nonetheless, no historian would argue that we should dismiss their reporting.

Logically, we can dismiss this hypothesis because it relies on the fallacy of authority. The validity of a claim is not related to the authority of the claimant. Thus, I reject H10.

Following scrutiny of the above ten epistemic reasons that conspiracy theories may be false, I conclude that three are plausible: H3, H6, and H9. It is plausible that conspiracy theories are unlikely to be true in proportion to: H3) the extent to which they correlate seemingly unrelated events; H6) the degeneration of their research paradigm; and H9) in proportion to the esotericism of their claims.

Next, I turn to four hypotheses that rely only on consequentialist grounds. These hypotheses say nothing about whether a conspiracy theory is true. Instead, they focus on how dangerous mass, serious consideration of a conspiracy theory is *given* that it is false.

B. Consequentialist Hypotheses (H11-H14)

i. Persuasiveness (H11)

Popper (1994) argues that Leninism, Stalinism, and the Holocaust were each to some extent caused by the prevalence of the "conspiracy theory of society," a monological belief system wherein outcomes are best explained by reference to the motives of conspirators. He argues that the prevalence of this premodern form of thought served as the kindling necessary for

the uptake of Lenin's vulgar, deterministic Marxism, Stalin's political purges, and Hitler's antisemitic conspiracy theory.

Hofstadter (1965) builds on Popper's work, arguing that the Red Scare belonged on Popper's list of bad outcomes caused by the conspiracy theory of society. Hofstadter embraces the elitist perspective that the masses lack the ability to reason well, so elites should steer them toward appropriate conclusions (Goshorn 2000). Hofstadter argues that the so-called paranoid style of politics lures unsuspecting citizens to believe unsubstantiated allegations of conspiracy. Robins and Post (1997) agree, arguing that the paranoid style was present in every social disaster of the preceding century. Pipes (1997) develops this approach further, analogizing conspiracy theories to pornography in their ability to seduce adherents. Each of these scholars would likely concede that the persuasiveness of a conspiracy theory is *beneficial* if the conspiracy theory is true. In that case, its persuasiveness contributes to its erasure. Henceforth emerges H11.

H11: *False* conspiracy theories are dangerous in proportion to their persuasiveness. This hypothesis is un-controversially true. A false conspiracy theory that nobody believes causes no harm, but one that the public embraces would likely cause the relevant investigatory bodies to waste their resources. To assess the persuasiveness of conspiracy theories, I use two indicators of persuasiveness from the conspiracy theory literature. First, I consider the self-sealing nature of belief in the conspiracy theory. Second, I consider the conspiracy theory's explanatory power.

Regarding the self-sealing quality of conspiracy theories, Keeley (1999), Sunstein and Vermeule (2009), and Harris (2019) argue that conspiracy theories tend to be self-sealing more often than their non-conspiratorial theories. They argue that once one begins to believe a conspiracy theory, the believer sometimes construes evidence against the conspiracy theory as

evidence in its favor. I alluded to this phenomenon above when I mentioned that Q sometimes responds to failed predictions by claiming that it expected those failures.

Regarding the explanatory power of conspiracy theories, many scholars argue that conspiracy theories are persuasive to their believers because of their ability to explain a range of important phenomenon that, without the help of the conspiracy theory, seem unexplained (Popper 1994; Hofstadter 1965; Jameson 1987; Pipes 1997; Dean 1998; Melley 2016). Thus, given that a conspiracy theory is false, it is more dangerous if it is more self-sealing and/or it provides more explanatory power.

ii. Crowding Out (H12)

Next, Gilbert (1996) argues that AIDS-as-genocide conspiracy theories are harmful because they distract from more tractable problems facing the Black community. Because an accurate conspiracy theory would articulate tractable problems, Gilbert implies that he is talking about only *false* conspiracy theories. Henceforth emerges H12.

H12: *False* conspiracy theories are dangerous in proportion to the extent to which they crowd out discussion of more tractable problems.

This hypothesis is un-controversially true. Consider a false conspiracy theory that implies that a government should use resources to prosecute alleged conspirators. This conspiracy theory is more dangerous than one that implies no such political action. For example, cryptozoological and sports-related conspiracy theories cause little harm.

I assess conspiracy theories according to H12 by considering two elements of the theory: fundamentality and tractability. First, I consider the tractability of problems implied by the conspiracy theory. If those problems are intractable, then the conspiracy theory crowds out discussion of more tractable problems. Second, if the problems implied by the conspiracy theory

are tractable, I assess the fundamentality of those problems compared to the fundamentality of problems implied by non-conspiratorial assessments of the conspiracy theory's topic. If the problems implied by the conspiracy theory are less fundamental than the problems implied by non-conspiratorial assessments, then the conspiracy theory crowds out discussion of more tractable problems.

iii. Medical Advice (H13)

Here, I develop a hypothesis to capture the special significance of conspiracy theories that imply medical advice and thus may affect medical practice. Accurate conspiracy theories that imply medical advice that *improves* medical outcomes are not dangerous. Henceforth emerges H13.

H13: *False* conspiracy theories are dangerous in proportion to the extent to which believing them could cause one to diverge from authoritatively advised medical behavior. This hypothesis is un-controversially true. If, for example, a conspiracy theory falsely claims that a vaccine for a transmissible disease is harmful or ineffective, belief in this conspiracy theory is harmful for the believer and those who encounter the believer.

To assess a conspiracy theory with respect to H13, we must consider the extent of harm caused by alleged shortcomings of an authoritatively advised medical practice and compare it to the harm of whatever that medical practice seeks to assist.

iv. Public Trust (H14)

As referenced previously, Keeley (1999) argues that belief in a conspiracy theory causes the believer to mistrust people, groups, and institutions accused by a conspiracy theory. He argues that public trust is fundamental for the sound operation of society, so conspiracy theories that implicate that trust are dangerous. The accuracy of a conspiracy theory does not influence

the extent to which belief in that conspiracy theory reduces public trust, so in this case I do *not* restrict the hypothesis to false conspiracy theories alone. Henceforth emerges H14.

H14: Conspiracy theories are dangerous in proportion to the extent to which their mass uptake erodes public trust. The difficulty of measuring both public trust and the extent to which a conspiracy theory erodes public trusts inhibits my ability to evaluate H14 on empirical grounds. Regardless, this hypothesis is implausible on three logical grounds.

First, the impact of belief in a conspiracy theory on public trust is limited to mistrust of those implicated by the conspiracy theory (Clarke 2002). For example, one need not scrutinize all of life's fundamental truths when one considers a possible instance of political corruption. Second, public mistrust up to a certain point is rational (Dean 2000) and good for society (Clarke 2002). It holds officials accountable, promotes transparency, and occasionally thwarts a harmful conspiracy. Third, when people conspire, the rational response is to reduce our trust in them. There are harms associated with mistakenly continuing to trust conspirators, and mass, serious consideration of some conspiracy theories (including all accurate ones) is a reasonable means of preventing those harms. Thus, I reject H14.

In conclusion, I reject H14, but I find H11, H12, and H13 are plausible. Each of the three plausible consequentialist hypotheses exhibits two glaring problems: the condition of falsehood and the mass uptake requirement. First, determining falsehood requires one of two practices that inhibit the utility of the hypothesis. Either one must rely on epistemic hypotheses, most of which I dismiss outright or qualify beyond practical usability, or one must undertake the careful consideration of evidence that deference to hypotheses H1-H16 intended to render obsolete. Second, consequentialist hypotheses have no force without the presence of mass uptake of a conspiracy theory. Even the most reckless conspiracy theory has little potential to cause harm if

very few people believe it. Thus, under these three hypotheses, a conspiracy theory that is dangerous today may be harmless tomorrow. This dynamism skirts the intentions of the hypotheses. To avoid this flaw, in Section IV I discuss popular conspiracy theories in the context of their mass uptake at the time of this writing. Next, I introduce two hypotheses that rely on both epistemic and consequentialist grounds.

C. Epistemic and Consequentialist Hypotheses (H15-H16)

i. Number of Alleged Conspirators (H15)

Stokes (2018) builds on H14 (public trust) by appending a consequentialist critique of conspiracy theories. He argues that conspiracy theories make morally weighty accusations. When conspiracy theories are false, these accusations cause harm to the alleged conspirators. Thus, the more conspirators alleged by the conspiracy theory, the more harm the conspiracy theory does. Henceforth emerges H15.

H15: Conspiracy theories are dangerous in proportion to the number of alleged conspirators. On the one hand, H15 is a consequentialist hypothesis. Accusations harm and reduce the public's trust of the accused. On the other hand, H15 is an epistemic hypothesis. Conspiracies that involve more conspirators are more difficult to conceal. Both the consequentialist and epistemic interpretations of this hypothesis are implausible on empirical grounds, but only the consequentialist interpretation is implausible on logical grounds.

Consequentialist interpretation: Empirically, some accurate conspiracy theories (e.g., the Manhattan Project and the Holocaust) that allege a large number of conspirators caused orders of magnitude more harm than the harm caused by accusing those alleged conspirators. While making an accusation about the Holocaust would require accusing thousands of Germans of grave wrongdoing, that accusation would also allege that those thousands of Germans are killing

millions of people. Logically, the more alleged conspirators, the more harmful the alleged conspiracy tends to be. Thus, while the number of alleged conspirators is positively correlated with the harm associated with mass, serious consideration of that conspiracy theory, that number is *also* positively correlated with the harm associated with dismissing the conspiracy theory. Thus, Stokes's argument appears to be self-defeating, so I reject the consequentialist interpretation of H15.

Epistemic interpretation: Empirically, the Manhattan Project and Holocaust were both conspiracies that involved secrecy by a large number of people. Logically, however, it is possible for a conspiracy theory to allege so many conspirators that they could not possibly maintain the secrecy necessary to pull off the conspiracy. Thus, where secrecy is required (which is not always the case (Dentith and Orr 2018)), I fail to reject H15's epistemic angle.

ii. Marginalization (H16)

Here, I introduce a hypothesis to capture the special significance of conspiracy theories that accuse a marginalized group of engagement in a conspiracy. Henceforth emerges H16: Conspiracy theories are dangerous in proportion to the extent to which they accuse or scapegoat a marginalized group. This is a consequentialist hypothesis insofar as we want to avoid the harms caused by accusing a marginalized group of a conspiracy. It is also an epistemic hypothesis insofar as a marginalized group, by virtue of marginalization, seems less likely to execute a meaningful conspiracy. Both interpretations of this hypothesis are plausible on empirical and logical grounds.

Consequentialist interpretation: Empirically, historical articulations of classical antisemitic conspiracy theories led to the mass displacements and murders of Jews dozens of times across European history. While various groups of Black activists articulate conspiracy

theories that accuse White people, mass uptake of those conspiracy theories (among Black people) rarely manifests in negative consequences for White people. Logically, a bigger, better resourced group can fend off conspiratorial accusations more effectively than a smaller, less resourced group.

Epistemic interpretation: Empirically, marginalized groups face greater hurdles to successfully conspiring. Consider some of the covert and illegal activities constituting COINTELPRO. This FBI program spied on and harassed leaders of the Civil Rights movement, thereby inhibiting the ability of that movement to conspire successfully (Wikipedia: "COINTELPRO"). Logically, a marginalized group is, by virtue of its marginalization, less capable of conspiring than other groups because it lacks the resources and leisure time necessary to conspire effectively. Thus, I fail to reject H16.

D. Concluding Section III

In the above section, I review sixteen hypotheses claiming to locate the danger associated with conspiracy theories. In conclusion, I identify eight plausible hypotheses: H3, H6, H9, H11, H12, H13, H15, and H16. The first three hypotheses can help us determine whether a conspiracy theory is true, the next three can help determine how dangerous a conspiracy theory is *given* that it is false, and the last two serve both roles.

In the next section, I apply these hypotheses to a set of popular conspiracy theories on which interventionists tend to fixate. I estimate the extent to which mass, serious consideration of each conspiracy theory should be regarded as dangerous under my eight hypotheses. Additionally, I compare the harms of mass, serious consideration to the harms of censorship.

IV. Assessing the Dangers Associated with Conspiracy Theories

In this section, I apply the eight hypotheses emerging from the above section to popular conspiracy theories to determine an estimate of the danger associated with mass, serious consideration of each conspiracy theory. First, I describe the process I use to determine not only which set of conspiracy theories but also which articulations from that set I select for analysis. Then, I introduce those articulations. Second, I apply the eight plausible hypotheses to the set of articulations. Third, I compare the estimated harms associated with mass, serious consideration of each conspiracy theory with the estimated harms associated with intervention into the discourses surrounding each articulation. This exercise is not meant to constitute a conclusion in the debate related to interventionism. Rather, this exercise serves as a model for how interventionists should consider interventionism.

A. Selecting Sets of Conspiracy Theories and Articulations Thereof

To determine the sets of conspiracy theories that I analyze, I rely on two criteria: prominence and interventionism on the grounds of misinformation. First, I focus only on prominent sets. Second, I focus only on sets that face interventionism by large social media corporations under those corporations' misinformation policies. Thus, I restrict my focus to conspiracy theories related to 1) COVID-19, 2) QAnon, and 3) the 2020 US Presidential Election.

To ensure the value of this project for interventionists and policymakers, I focus on articulations that would be likely to trigger intervention but should not face intervention according to academic scholarship on the matter. This focus reveals societally relevant contrasts: if deference to relevant academic scholarship would require that social media corporations do not intervene against these cases, then social media corporations should either revise their policies to align with scholarship or cease claiming that their policies are informed by deference to experts.

Below, I review two articulations of each of three sets of conspiracy theories. Each of the six resulting articulations meets the criteria I develop above.

i. COVID-19 Vaccines (V)

Facebook claims that they remove posts containing harmful health misinformation, including misinformation about vaccines. This includes posts that claim that vaccines are dangerous (Meta Transparency Center: "Misinformation":II.d.). Facebook also claims that it removes posts claiming that the COVID-19 vaccines are not effective at reducing symptomatic disease (Meta Transparency Center: "Misinformation":II.g.). Twitter requires the deletion of tweets that make "False claims about COVID-19 that invoke a deliberate conspiracy by malicious and/or powerful forces" (Twitter Help Center: "COVID-19"). Facebook has a similar policy (Facebook Help Center: "COVID-19"). Thus, I introduce V1 and V2:

> V1: To maximize revenue, Pfizer and Moderna intentionally misrepresent the danger associated with their COVID-19 vaccines.
>
> V2: To maximize revenue, Pfizer and Moderna intentionally misrepresent how effectively their COVID-19 vaccines prevent symptomatic cases of COVID-19.

ii. QAnon (Q)

Facebook, YouTube, and Twitter each intervene broadly to reduce the spread of claims related to QAnon, especially those that accuse specific people or groups (YouTube Team 2020). QAnon-related censorship spans misinformation policies as well as policies related to harm and harassment. Below I introduce Q1 and Q2, two articulations that I expect would be removed from these platforms at least with reference to if not total reliance on misinformation policies:

> Q1: Bill Clinton, Hillary Clinton, John Podesta, and Tony Podesta conspire to engage in and cover-up their engagement in pedophilic activity.
>
> Q2: There is "[…]a conspiracy involving 'deep state actors' and global elites engaged in an international child sex trafficking ring" (Adapted from FBI Phoenix Field Office 2019).

iii. 2020 US Presidential Election (E)

Facebook, YouTube, and Twitter each intervene broadly to reduce the spread of claims that question the validity of elections considered free and fair by third-party fact-checkers (Meta Transparency Center: "Misinformation"; YouTube Help: "Misinformation"; Twitter Help Center: "Civic"). Thus, I introduce E1 and E2:

> E1: Those in charge of Dominion Voting Systems conspired to rig the 2020 US Presidential Election for Joe Biden.
>
> E2: The Democratic Party and volunteers at polling places conspired to rig the 2020 US Presidential Election for Joe Biden.

B. Applying Hypotheses to Articulations

Table 3

Type	Name	Claim	Plausible
Epistemic	H3: Correlation	Conspiracy theories are unlikely to be true in proportion to the extent to which they correlate seemingly unrelated events.	Yes
	H6: Degeneration	Conspiracy theories are unlikely to be true in proportion to the degeneration of their research paradigm.	Yes
	H9: Esotericism	Conspiracy theories are unlikely to be true in proportion to the esotericism of their claims.	Yes
Consequentialist	H11: Persuasiveness	*False* conspiracy theories are dangerous in proportion to their persuasiveness.	Yes
	H12: Crowding Out	*False* conspiracy theories are dangerous in proportion to the extent to which they crowd out discussion of more tractable problems.	Yes
	H13: Medical Advice	*False* conspiracy theories are dangerous in proportion to the extent to which believing them could cause one to diverge from authoritatively advised medical behavior.	Yes
Both	H15: # of Conspirators	Conspiracy theories are dangerous in proportion to the number of alleged conspirators.	Yes
	H16: Marginalization	Conspiracy theories are dangerous in proportion to the extent to which they accuse or scapegoat a marginalized group.	Yes

In this sub-section, I apply eight plausible hypotheses that claim to locate the danger associated with conspiracy theories (reproduced above in Table 3) to six conspiracy theories: V1, V2, Q1, Q2, E1, and E2. For each articulation, I begin by applying to that articulation the epistemic hypotheses that argue that conspiracy theories are more likely to be false when they contain certain traits. Then, I turn to consequentialist hypotheses that argue that false conspiracy

theories are dangerous when many people believe them. As I move through the articulations, I develop and apply a method that interventionists should use when they consider whether to apply their policies to conspiratorial articulations.

i. COVID-19 Vaccines (V)

V1 states that "To maximize revenue, Pfizer and Moderna intentionally misrepresent the danger associated with their COVID-19 vaccines."

Epistemic hypotheses: Regarding H3 (Correlation), V1 makes a narrow claim, so it does not correlate seemingly unrelated events. Thus, V1 is not dangerous under H3. Regarding H6 (Degeneration), V1's status is ambiguous. Official validation of the safety of the vaccines was not revoked following review of the alleged danger of spike proteins. Several studies, however, maintain that this danger is plausible, if not likely (Badshah et al. 2021; Chouchana et al. 2021; Istampoulouoglou et al. 2021; Lim et al. 2021; Matta et al. 2021; Habedank et al. 2022). Thus, V1 is not dangerous under H6. Regarding H9 (Esotericism), V1 is not esoteric. Corporations (including pharmaceutical ones) sometimes misrepresent the dangers of their products when their profits are at stake. Thus, V1 is not dangerous under H9. Regarding H15 (# of Conspirators), V1 does not necessitate more conspirators than the successfully executed Holocaust, Manhattan Project, or NSA-spying, so V1 is not dangerous under H15. Regarding H16 (Marginalization), V1 does not accuse or scapegoat a marginalized group, so V1 is not dangerous under H16.

Consequentialist hypotheses: Regarding H11 (Persuasiveness), V1 is persuasive because it is self-sealing. Once one believes a conspiracy theory that relies on an assertion of the conspirators' motives, no new evidence can overturn that belief because a conspirator with bad motives will never concede that they operate with bad motives. Thus, V1 is dangerous under H11. Regarding H12 (Crowding Out), V1 invokes tractable concerns about prevention of

COVID-19. Additionally, prevention is fundamental to society-level management of a disease. Thus, V1 is not dangerous under H12. Regarding H13 (Medical Advice), mass, serious consideration of V1 is likely to cause some people to avoid vaccination. Thus, conditional on V1's falseness and thus the vaccines' safety, V1 is dangerous under H13. Thus, V1 is dangerous under H11 and H13.

V2 states that "To maximize revenue, Pfizer and Moderna intentionally misrepresent how effectively their COVID-19 vaccines prevent symptomatic cases of COVID-19." My analysis of V2 is identical to my analysis of V1 except for one exception. V2 is even less degenerative than V1, as many studies show that the effectiveness of the mRNA vaccine wanes over time (Mandavilli 2021; Zimmer 2021), and Pfizer and Moderna do not dispute this information. Thus, V2 is dangerous under H11 and H13, and there is little doubt that mass, serious consideration of V1 and V2 is dangerous for society if those articulations are false.

Most analysis would end at this point, concluding that the above analysis justifies intervention into discourse related to conspiracy theories about COVID-19 vaccines. But this analysis leaves out an essential step. To assess the impact of conspiracy theories or interventionism, we must ask not only the danger of conspiracy theories but also the danger of interventionism.

First, recall that H11 and H13 are consequentialist hypotheses that only have force if the requisite conspiracy theory is false. Thus, if V1 and V2 are true, there is no good reason to claim that V1 and V2 are dangerous. If interventionists censored V1 and V2 in a world in which they were true, those interventionists would become responsible for harm caused by vaccines or harm caused by the manner in which faith in vaccines crowds out more fundamental disease prevention methods. Because the science related to V1 and V2 is in flux, social media

corporations should allow skepticism on their platforms, waiting to consider censorship until V1 and V2 are knowable.

Ultimately, V1 and V2 embody American traditions of skepticism of large corporations and skepticism of emergency government actions that remain in place after emergencies have waned. Interventionism compromises these traditions in pursuit of uncertain benefits.

ii. QAnon (Q)

Q1 states that "Bill Clinton, Hillary Clinton, John Podesta, and Tony Podesta conspire to engage in and cover-up their engagement in pedophilic activity."

Epistemic hypotheses: Regarding H3 (Correlation), Q1 makes a narrow claim, so it does not correlate seemingly unrelated events. Thus, Q1 is not dangerous under H3. Regarding H6 (Degeneration), Q1's status is ambiguous. Q1 is degenerative insofar as no confirmatory evidence has emerged, however because no formal investigation has taken place, this degeneration lacks epistemic value. Thus, Q1 is not dangerous under H6. Regarding H9 (Esotericism), Q1 is esoteric, but it is not more esoteric than some confirmed phenomena. As revelations related to the Catholic and Baptist churches and Bill Clinton's frequent use of Jeffrey Epstein's private plane (Bryant 2015) show, pedophilia is more common and more proximate to trusted institutions than we would like to think. Thus, Q1 is not dangerous under H9. Regarding H15 (# of Conspirators), Q1 alleges only four conspirators, so it is not dangerous under H15. Regarding H16 (Marginalization), Q1 does not accuse or scapegoat a marginalized group, so it is not dangerous under H16.

Consequentialist hypotheses: Regarding H11 (Persuasiveness), Q1 is persuasive. Its mass uptake despite the lack of evidence hints that it is self-sealing. Thus, Q1 is dangerous under H11. Regarding H12 (Crowding Out), Q1 is tractable only insofar as it is plausible that a Republican-

controlled House of Representatives would investigate the Clintons and Podestas in the absence of probable cause. More importantly, Q1 is not fundamental to the problems of sex trafficking or pedophilia. Thus, Q1 crowds out discussion of more fundamental, tractable matters, so Q1 is dangerous under H12. Regarding H13 (Medical Practice), Q1 does not invoke medical behavior, therefore it is not dangerous under H13. I conclude that Q1 is dangerous under H11 and H12.

Social media corporations assume that claims like Q1 are false thereby triggering consequentialist hypotheses H11 and H12. But given the prevalence of pedophilia in society, this assumption seems unjustified. Even without evidence, Q1 is plausible. Thus, once we disregard H11 and H12 for their reliance on Q1's falsehood, we are left with no good reasons to assert that Q1 is dangerous. Thus, justification for intervention into discourse related to articulations like Q1 should rely on a case-specific comparison of the harms of belief to the harms of interventionism.

Q2 states that "There is '[…]a conspiracy involving deep state actors and global elites engaged in an international child sex trafficking ring.'" This articulation extends Q1. Relative to Q1, Q2 features a lower likelihood that it is true, greater danger if it is true, and increases the stakes if it is true.

Epistemic hypotheses: Regarding H9 (Esotericism), Q2 is less likely to be true than Q1 insofar as Q2 is more esoteric than Q1. Still, Q2's esotericism does not surpass that of pervasive pedophilia within the Catholic Church. Regarding H15 (# of Conspirators), Q2 accuses more people than Q1. Thus, relative to Q1, it would be more difficult for Q2's alleged conspirators to maintain secrecy. Because Q2 does not specify a very large number of 'deep state actors,' it is not so dangerous under H15 that H15 constitutes a distinct reason to claim that Q2 is dangerous.

I conclude that because of H9 and H15, Q2 is more likely to be false and thereby more dangerous than Q1.

Consequentialist hypotheses: For two reasons, Q2 is more dangerous than Q1. First, regarding H11 (Persuasiveness), Q2's increased breadth increases its explanatory power and thereby increases its persuasiveness. Second, regarding H15 (# of Conspirators), the greater number of alleged conspirators in Q2 relative to Q1 means that more alleged conspirators are falsely accused. This concern is mitigated, however, by the anonymity of the alleged conspirators in Q2. Thus, where Q1 and Q2 are both false, Q2 is more dangerous because of H11.

Application of these hypotheses to Q1 and Q2 does not reveal all meaningful incongruities between the two articulations. While Q2 is less likely to be true and more dangerous if it is false, Q2 also causes greater harm in a world in which it is true. Thus, I argue that increased danger of the articulation and increased danger of ignoring the articulation roughly offset. When I adjudicate benefits and costs of interventionism below, I focus on Q1.

The FBI's Phoenix Office attributes several harms to belief in claims like Q1. These harms include a thwarted bombing attempt, blocking traffic, harassment of alleged conspirators, and the famous armed threat to Comet Ping Pong in 2016 (FBI Phoenix Field Office 2019). Assuming that it is accurate to attribute those harms to belief in Q1, let mass, serious consideration of Q1 cause negative 1 util. To consider the harm associated with Q1 if it is true, we must estimate at least one of 1) the harm caused if Q1 is true and 2) the probability that Q1 is true. If we assume that Q1 is associated with negative 1,000 utils if it is true, then the assignment of any probability greater than 1 in 1000 renders interventionism a harmful action.

I offer no method for assigning utils, instead directing the reader toward moral philosophy to assist them in that endeavor. I do, however, offer a method for assigning

probabilities. The assigner should avoid reliance on epistemically unsound heuristics and assess the articulation as a function of its evidence. While this process is laborious and subjective, it is nonetheless possible and preferable to deference to fallacies and fact-checkers. Additionally, utilization of a method such as this one and transparent publication of the results of applying the method to particular cases could relieve some of the criticism faced by interventionists today.

Returning to the case of Q1 and Q2, I argue that in the absence of serious investigation of claims like Q1 and Q2 by the justice system or social media corporations, broad censorship of those claims is unjustified because the assumption of zero probability is unjustified. Potential harms if Q1 and Q2 are accurate, meanwhile, are so massive that even slight probabilities should give interventionists pause.

iii. 2020 US Presidential Election (E)

E1 states that "Those in charge of Dominion Voting Systems conspired to rig the 2020 US Presidential Election for Joe Biden."

Epistemic hypotheses: Regarding H3 (Correlation), E1 makes a narrow claim, so it does not correlate seemingly unrelated events. Thus, E1 is not dangerous under H3. Regarding H3 (Degeneration), the results of several legal cases severely degenerate E1. Thus, E1 is dangerous under H6. Regarding H9 (Esotericism), there is nothing esoteric about E1, as decisive election fraud occurs regularly outside the US. Thus, E1 is not dangerous under H9. Regarding H15 (# of Conspirators), E1 does not necessitate many conspirators, so it is not dangerous under H15. Regarding H16 (Marginalization), E1 does not accuse or scapegoat a marginalized group, so it is not dangerous under H16.

Consequentialist hypotheses: Regarding H11 (Persuasiveness), E1 has high explanatory power insofar as it explains the outcome of a major election. Empirically, E1 is self-sealing

insofar as many of its adherents maintain their belief despite the prevalence of failed legal challenges to the election. Thus, E1 is persuasive, so it is dangerous under H11. Regarding H12 (Crowding Out), election fraud is both tractable and fundamental to elections, so E1 is not dangerous under H12. Regarding H13 (Medical Advice), E1 does not invoke medical behavior, so it is not dangerous under H13.

E2 states that "The Democratic Party and volunteers at polling places conspired to rig the 2020 US Presidential Election for Joe Biden." I evaluate this articulation similarly, however this articulation is also dangerous under H15 insofar as it accuses many conspirators. Regarding the epistemic interpretation of H15, the number of conspirators E2 alleges (between 10,000 and 100,000, I estimate) is not sufficient to render E2 unlikely. But regarding the consequentialist interpretation of H15, a false accusation against 10,000 to 100,000 alleged conspirators causes great harm and thereby renders E2 dangerous under H15. Below, I compare the danger of E1 and E2 to the danger of interventionism against E1 and E2.

Censorship of claims like E1 and E2 may be about as dangerous as mass, serious consideration of those claims. On the one hand, mass, serious consideration is likely to cause temporary instability and may cause events resembling those we observed on January 6, 2021. On the other hand, prevention of mass, serious consideration is likely to inhibit the discovery of a stolen election at some point in the future.

Imagine that 1% of American presidential elections are stolen, 0% of stolen elections are discovered when we censor associated allegations, and 100% of stolen elections are discovered when we do not censor those allegations. Further, imagine that successfully stolen elections cause negative 100 utils while thwarted attempts cause 0 utils. Lastly, imagine that non-censorship provides 90% chance of negative 10 utils as instability ensues and resources are

directed toward investigation. Using these placeholder estimates of utility, we find that censorship causes an expected negative 10 utils (.01 * -1000 = -10) and non-censorship also causes an expected change in utility of negative 10 utils (.9 * -10 = -9). Using these placeholders for likelihoods and harms, censorship is more harmful than non-censorship (-10 < -9).

Again, the above probabilities and utils do not rely on evidence. I selected those probabilities and utils in order to demonstrate merely that it is at least plausible that interventionists should permit free discourse regarding election-integrity. I reproduce results from this section in Table 4 below.

Table 4

Set	Articulation	Grounds for Danger
COVID-19 Vaccines (V)	V1: To maximize revenue, Pfizer and Moderna intentionally misrepresent the danger associated with their COVID-19 vaccines.	H11, H13
	V2: To maximize revenue, Pfizer and Moderna intentionally misrepresent how effectively their COVID-19 vaccines prevent symptomatic cases of COVID-19.	H11, H13
QAnon (Q)	Q1: Bill Clinton, Hillary Clinton, John Podesta, and Tony Podesta conspire to engage in and cover-up their engagement in pedophilic activity.	H11, H12
	Q2: There is "[…]a conspiracy involving 'deep state actors' and global elites engaged in an international child sex trafficking ring."	H11, H12
2020 Presidential Election (E)	E1: Those in charge of Dominion Voting Systems conspired to rig the 2020 US Presidential Election for Joe Biden.	H6
	E2: The Democratic Party and volunteers at polling places conspired to rig the 2020 US Presidential Election for Joe Biden.	H6, H15

V. Conclusion

In reviewing articulations of conspiracy theories related to COVID-19, QAnon, and the 2020 US Presidential election, I show that social media interventionists need not rely on fallacies and fact-checkers when they consider interventionism. Instead, they could conduct transparent cost-benefit analyses. If they did so, they would find that under a certain range of estimated likelihoods and harms, interventionism causes more harm than it prevents.

Additionally, throughout the analysis undertaken in this chapter I have set aside consideration of the boomerang effect—the phenomenon whereby attempts to restrict access to information tend to backfire (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018). Interventionists pride themselves on deference to the judgements of subject-area experts in infectious disease, law enforcement, and the justice system when they decide which conspiracy theories are false, yet they ignore experts in psychology who demonstrate the futility of censoring false claims.

Ultimately, this chapter serves as a call for more scrutiny of interventionism. This scrutiny should take place across social media corporations, academia, and journalism. I hope that increased scrutiny will lead to the development of methodologies for assessing the costs and benefits of interventionism that meet four criteria. First, methodologies should consider the boomerang effect. Second, they should take conspiracy theories seriously, acknowledging that very few conspiracy theories are impossible and small probabilities matter. Third, they should assess not only the harms of mass, serious consideration of conspiracy theories but also the harms of censoring conspiracy theories if they are true. Fourth, the above steps should take place in plain view and with fidelity to clear policies that users accept when they sign up for social media platforms. Any methodology that meets these criteria would be superior to the status quo.

Chapter 3: Censorship, Stigma, and Conspiracy Theory Discourse on Twitter

I. Introduction

A. Definitions of Key Terms

In this chapter, I use definitions of some key terms that differ from colloquial definitions or those found in a dictionary. To avoid confusion, I begin by defining those terms.

Table 1

Term	My Definition	Flawed Alternative Definition
conspiracy	a multi-person plan or action employing morally suspect means and/or toward morally suspect ends	a *true* claim that a conspiracy is the best explanation for an outcome
conspiracy theory	a claim that a conspiracy is the best explanation for an outcome	a *false* claim that a conspiracy is the best explanation for an outcome
conspiracism	one's tendency to prefer a conspiracy theory to plausible, non-conspiratorial theories	one's tendency to prefer a *false* conspiracy theory to plausible, non-conspiratorial theories
conspiracist	one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories	
conspiracy theorist		one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories
stigma	a mark of disgrace upon a claim that causes a person making that claim to fear social exclusion	
conspiracy-theory-stigma	the stigma associated with public endorsement of a conspiracy theory	

Table 1 shows my definitions of key terms. I define "conspiracy" as a multi-person plan or action employing morally suspect means and/or toward morally suspect ends. I define a "conspiracy theory" as a claim that a conspiracy is the best explanation for an outcome. Many plans and actions not colloquially referred to as conspiracies nonetheless qualify under my definition. For example, my definition of conspiracies includes 1) marital infidelity, 2) political corruption, 3) insider trading, and 4) the official theory of 9/11 (examples borrowed from Pigden 1995, 2006 and Dentith 2014). On the other hand, I define a "non-conspiratorial theory" as a

claim that some explanation other than a conspiracy best explains an outcome. Examples of common non-conspiratorial theories include coincidences, lone actors, and institutional forces.

Under my definitions, confirmation that a conspiracy theory is accurate (confessions, DNA, fingerprints, etc.) does not cause the conspiracy theory to lose its status as such. Thus, we all believe many conspiracy theories, and we are all conspiracy theorists. To allow for variance, I define "conspiracism" as one's tendency to prefer a conspiracy theory to plausible, non-conspiratorial theories, and I define a "conspiracist" as one who more often than most prefers a conspiracy theory to plausible, non-conspiratorial theories.

Lastly, I define "stigma" a mark of disgrace upon a claim that causes a person making that claim to fear social exclusion. Accordingly, the "conspiracy-theory-stigma" is the stigma associated with public endorsement of a conspiracy theory. Now that I have equipped the reader with these definitions, I turn to the rest of the chapter where I attempt to measure the conspiracy-theory-stigma and answer commonsense questions that one can only answer with such a measure in hand.

B. Purpose and Roadmap

In this chapter, I use natural language processing and machine learning to estimate the probability that stigma is associated with approximately one million tweets about fifteen conspiracy theories over fifteen years (2007-2022). Then, I use the estimates to answer three questions from the conspiracy-theory-stigma literature. First, what is the over-time trend in the stigma associated with conspiracy theories (Barkun 2015; Basham 2018; Thalmann 2019)? Second, what causes the stigma (Dean 2000; Husting and Orr 2007; Wendt and Duvall 2008)? Third, what are the consequences of the stigma (Dean 2000; Goshorn 2000; DeHaven-Smith 2013; Coady 2018; Orr and Husting 2019)?

I situate each of these questions in the context of censorship on social media. Research from psychology establishes the boomerang effect (sometimes referred to as the Streisand effect), a phenomenon whereby attempts to restrict access to information tend to increase interest in that information (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018). For example, some QAnon followers responded to being banned from Facebook by arguing that the bans were evidence of their conspiracy theories (Frenkel 2020). Contrary to censorship, rationalizing and ridiculing arguments effectively change beliefs (Orosz et al. 2016). That which faces censorship cannot face rationalizing and ridiculing rebuttal. Thus, it is plausible that a laissez-faire approach to managing conspiracy-theory-discourse on social media—an approach that avoid censorship—would reduce belief in conspiracy theories more effectively than censorship. Above, I discuss reasons to think that censorship may be ineffective. Below, I discuss evidence that social media corporations do not follow sound reasoning in determining what to censor.

According to Twitter's then-CEO Jack Dorsey, censorship of a *New York Post* article reporting on Hunter Biden's corrupt activities was a mistake (Manskar 2021). Twitter's Policy Communications Director Andy Stone conceded that the decision was based on mere suspicion, not the results of a formal investigation (@andymstone 2020). Facebook bases its misinformation-designations on *PolitiFact*'s. *PolitiFact* at first rated the Lab Leak Hypothesis "Pants on Fire!" (Funke 2020) before amending that ruling to "unsupported by evidence and in dispute" (Editor's Note 2021). Facebook's policies allow for censorship so long as the censors "have signs that a piece of content is false" (Rosen 2019). With policies like these, it is all but certain that social media corporations will occasionally censor true, important information (Thacker 2021).

Given evidence about the counter-productiveness of censorship, the effectiveness of rationalizing and ridiculing arguments, and the flimsiness of social-media-censorship-policies, we lack evidence that, relative to the laissez-faire approach, censorship reduces harm. Thus, we ought to scrutinize justifications for censorship that evoke such a reduction in harm (Meta Transparency Center: "Misinformation"; YouTube Help: "Misinformation"; Twitter Help Center: "COVID-19"; Twitter Help Center: "Civic"; Twitter Help Center: "Synthetic"). From these policies I glean several dubious assumptions on which the justifications for censorship rely. I focus on one such assumption: without censorship, misinformation thrives. This chapter tests that assumption by studying the waxing and waning of the stigma associated with conspiracy theories on Twitter. In the status quo, social media users police each other by including stigma-laden language in their tweets about conspiracy theories. This mechanism has the potential to prevent conspiracy theories from maintaining high popularity over time.

Additionally, few contemporarily popular conspiracy theories originate in the distant past despite the absence of censorship of those conspiracy theories. The classical antisemitic conspiracy theory (which sometimes includes Holocaust Denial) does originate in the distant past, but this exception supports my argument because the classical antisemitic conspiracy theory faces *more* censorship than conspiracy theories that fade *without* censorship. This logic and evidence suggests that something other than censorship causes conspiracy theories to fade over time. This chapter tests one alternative mechanism—the ability of an event that stigmatizes a conspiracy theory or its believers stigmatizes discourse related to that conspiracy theory.

Regardless of the results of analyses in this chapter, I maintain that the boomerang effect invalidates the status quo approach of censoring conspiracy-theory-discourse to reduce the harm

associated with mass, serious consideration of the censored conspiracy theories. This chapter seeks to contribute additional support to that argument.

To support my argument, I test four hypotheses. Below, I introduce each hypothesis, then I summarize the result of testing each hypothesis.

H1: The conspiracy-theory-stigma is constant from 2007-2022.

H2: Events that ought to stigmatize a conspiracy theory do not increase the predicted probability of stigma present in tweets related to that conspiracy theory.

H3: Following an event that ought to stigmatize a conspiracy theory, tweets with higher predicted probabilities of stigma do not experience a greater increase in retweets than tweets with lower probabilities of stigma.

H4: It will not be easier to reject H2 than H3.

Regarding H1, the average predicted probability of stigma associated with tweets about conspiracy theories rises by 0.002 per year since Twitter's inception. Most of this rise is concentrated in the period since Donald Trump announced his candidacy for President of the United States in June of 2015. Since then, the predicted probability of stigma associated with the average tweet increases by an average of 0.004 per year. Thus, I reject H1. This apparent rise in the conspiracy-theory-stigma mitigates concerns that conspiracy theories thrive today more than they have in the past.

Regarding H2, events that most reasonable observers would agree ought to stigmatize a conspiracy theory seem to do just that. For example, the storming of the US Capitol on January 6, 2021 ought to stigmatize conspiracy theories about the 2020 Presidential Election because those who stormed the capitol acted rashly and violently. As I hypothesize, tweets about those conspiracy theories the 24 hours after the storming had 0.05 higher average predicted probability of stigma than tweets about the same conspiracy theories in the 24 hours before the storming. This pattern holds for most events I scrutinize. Thus, I reject H2.

Regarding H3, while retweets do not always indicate agreement with the retweeted tweet, evidence shows that they tend to do so (Metaxas et al. 2015). I show that stigmatizing events have an inconsistent effect on retweets for stigmatized tweets relative to non-stigmatized tweets. I also show that, overall, stigma is inversely correlated with retweets. Thus, I fail to reject H3.

Regarding H4, because I reject H2 and fail to reject H3, I reject H4.

Overall, it seems that organic processes in social media discourse may accomplish some of what social media corporations try to accomplish via censorship. I do not measure belief in these conspiracy theories. Nor do I test my hypotheses in an experimental setting. Thus, this chapter does not end the debate on these matters. Rather, it starts the debate by 1) problematizing the status quo approach of social media corporations and 2) reorienting the debate from theory and assumption toward measurement and quantitative analysis. I hope this chapter serves as a useful first step toward measuring the conspiracy-theory-stigma and addressing pressing questions about conspiracy theories, conspiracy-theory-stigma, and censorship.

C. Literature Review

Most academic research studying conspiracy theories focuses on either the nature of conspiracy theories or the nature of *belief in* conspiracy theories. This chapter diverges from that tradition to focus on the conspiracy-theory-stigma. Before I introduce the methods I use to quantify stigma, I provide a brief review of the literature focused on the conspiracy-theory-stigma, and I show how my chapter builds on that literature.

Coady (2003) and Dentith (2019) mention in passing that the conspiracy-theory-stigma exists. Barkun (2015) and Basham (2018) claim that the stigma is decreasing, though they provide no evidence to support this claim. Thalmann (2019) uses qualitative data to support her

argument that the stigma increased from the late-1960s through the mid-1990s. I use novel methods to engage with this dispute.

Wood and Douglas (2013, 2015) dig deeper, demonstrating the presence and consequences of the stigma by showing that, in an experimental setting, people tend to be unwilling to apply the label "conspiracy theory" to their own claims. Next, Wood (2015) used another experiment to find that the presence or absence of the label "conspiracy theory" affixed to a claim is not associated with expressed belief in that claim. These studies lack generalizability beyond their experimental settings because stigma is socially constructed and creates social consequences, but these experiments did not attempt to induce such a context. The measure developed in this chapter maximizes generalizability by measuring stigma in the very context in which it is most important in modern discourse—on social media.

Attempting to induce a social context, Klein et al. (2015) ask survey-takers how they feel about conspiracy theorists. The survey-takers perceive conspiracy theorists as more "gullible, crazy, easily influenced, stupid naïve, manipulative, dishonest, assertive, and selfish than non-believers" (Abstract). Lantian et al. (2018) combine experimental design with social context by asking subjects to visualize advocating for or against a conspiracy theory about the 2015 *Charlie Hebdo* shooting. They find that those in the "for" condition exhibit greater fear of social exclusion than those in the "against" condition. Fear of social exclusion captures the essence of stigma as I study it in this chapter. The measure developed in this chapter may be used for any conspiracy theory, including those that do not yet exist, thus reducing the need to study the stigma *ad hoc* as new conspiracy theories emerge.

The remaining work in this literature studies the discursive use of the stigma (Fiske 1999, Bratich 2002, Husting and Orr 2007; Bratich 2008; Wendt and Duvall 2008; DeHaven-Smith

2013; Harambaum and Aupers 2015; Barkun 2015), structural conditions under which stigma should thrive (Wendt and Duvall 2008), and the consequences of the stigma in public discourse (Wendt and Duvall 2008, DeHaven-Smith 2013, Dean 2000, Coady 2018, Raikka and Basham 2019, Orr and Husting 2019). The measure developed in this chapter does not help us comment on the discursive use of the stigma nor the structural conditions under which it should thrive, but it does allow for an inquiry into the causes and consequences of stigma.

While no research has attempted to measure the conspiracy-theory-stigma, two papers attempt to measure the stigma associated with mental health (Robinson et al. 2019; Li et al. 2020). Notably, Li et al. (2020) developed a model very similar to the one I develop for this chapter. They used human raters to rate the stigma associated with Weibo posts related to Schizophrenia and Depression. One of their goals is to determine what features of these posts distinguish stigmatizing from non-stigmatizing posts. While their paper focuses on un-packing their model to see what drives its predictions, this chapter takes the model's predictions as given and uses the model's predictions to measure stigma across time and across conspiracy theories.

All work that engages with conspiracy-theory-stigma shares at least one of two limitations that this chapter attempts to overcome. First, most of the work in this literature uses labor-intensive qualitative measurement strategies (especially case studies) that require additional inquiry for each time period and conspiracy theory. Second, most of the work in this literature does not allow for differences across conspiracy theories. The methods in this chapter overcome these limitations by developing a model that can measure the stigma associated with any conspiracy theory and allowing for variation in stigma across conspiracy theories. Next, I introduce the reader to data and the measurement technique I employ to test hypotheses.

II. Data and Modeling

A. Data

To measure stigma in public discourse, I turn to the modern public forum—social media. In particular, I turn to Twitter because of the unrivaled access it provides for academic researchers. I used the academictwitteR package for the R programming language to collect tweets based on search queries (Barrie and Ho 2021). I used this package to extract approximately three million tweets about fifteen conspiracy theories spanning the time since Twitter's inception (2007-2022).

ii. Selection of Users

To ensure that tweets in my sample are appropriate for the measurement of stigma, I restrict my sample to include only tweets published by "verified" users of Twitter. These users provide evidence to indicate that they are authentic, notable, and active (Twitter Help Center: "Legacy verification policy"). Because verified users are authentic, tweets by verified users reflect on the reputation of the corresponding person or business to a greater extent than other tweets. This renders tweets from verified users appropriate for the measurement of stigma.

A significant proportion of non-verified Twitter-users are bots—semi-automated accounts pursuing the agendas of those who control them. Bots are not beholden to human psychology and stigma that rely on that psychology. Estimates about the proportion of users who are bots on Twitter range from Twitter's internal estimate of less than 5% (Dang 2022) to 80% (Woods 2022) and everywhere in between (Varol et al. 2017; Young 2020; Barkoukis 2022). Bots are especially active in discussions of conspiracy theories like those I address in this chapter. Carley claims to find that 34% of tweets discussing "reopening America" in early-2020 were generated by accounts that are definitely bots (Young 2020). Thus, by avoiding tweets

published by unverified users and the bots among them, I insulate my analysis against the uncertain role played by bots on Twitter.

Verified Twitter users are different from non-verified users not only in their authenticity but also their notability. Notable users are likely to be different from non-notable users in many ways, but there is no obvious reason that the influence of stigma should differ among notable and non-notable users. By excluding non-verified users, I avoid the uncertainty induced by the presence of bots among non-verified users without biasing my analysis in any obvious way. The differences between verified and non-verified users motivate H4, which I discuss in Section III.

ii. Selection of Tweets

I am interested in conspiracy theories that demonstrate extensive discourse on Twitter. I avoid focusing only on conspiracy theories that are likely to demonstrate minimal or maximal stigma. Thus, I include not only tweets about censored conspiracy theories like allegations of 2020 Presidential Election fraud, QAnon, or vaccines but also tweets about "Deflategate," the conspiracy theory that claims that National Football League Quarterback Tom Brady conspired to deflate his team's footballs and thereby gain an advantage.

To collect tweets that relate to particular conspiracy theories, I use regular expressions, the implicit language utilized by search engines when we type something into a search bar. Direct use of regular expressions allows one to filter results of a query with great specificity. For different conspiracy theories I employed different filters. For Deflategate, I collected 100% of tweets from verified users that use the word "Deflategate" and did not apply a filter. For tweets about conspiracy theories regarding the assassination of President John F. Kennedy (JFK), I used many filters to ensure that I did not engage with tweets that primarily address 1) JFK's

presidency, 2) the anniversary of JFK's death, 3) JFK International Airport, or 4) the "JFK Terror Plot" of 2007. See Table 2 (below) for details regarding the queries I use to select tweets.

First, I query Twitter's API with a word, phrase, or list of words and/or phrases. These queries, reflected in the column "Query" in Table 2, use "or" logic such that all tweets containing at least one of the listed words and/or phrases appear in my initial sample. The token "" in the "Query" column serves as a placeholder for the word or phrase in the column "Conspiracy Theory" immediately to the left of the column "Query." For example, ""' conspiracy' in the third row of the table means that I searched for tweets containing the phrase "JFK conspiracy." All queries were case insensitive.

Second, I scrutinize the tweets corresponding to each cluster of conspiracy theories and determine rules that ensure that almost all of my tweets engage with a conspiracy theory of interest. I exclude from my sample any tweet that includes at least one of the words or phrases in the "Excludes" column. Additionally, I only include tweets that contain at least one of the words or phrases in the "Includes" column. For example, tweets containing the phrase "hunter biden" but which contain references to drugs or marital infidelity are likely to refer to Hunter Biden's drug use or marital infidelity and not to his involvement in the "Laptop from Hell" set of conspiracy theories. But tweets containing "hunter biden" and something about a laptop or Ukraine are likely to refer to the set of conspiracy theories I seek to study in this chapter.

In addition to narrowing the sample in the manner described in Table 2, I also modified the text of the tweets to render them amenable to the machine learning. I describe these modifications as well as the modeling process in sub-section B below.

Table 2:

Conspiracy Theory	Query	Excludes	Includes
Hunter Biden	hunter biden	drug; drugs; crack; cocaine; smoke; smoking; smoked; alcohol; paternity; arkansas; baby; love child; widow; wife; wedding; marry; married; marrying; marriage; divorce; divorced; divorcing; promise me, dad; ashley madison	ukraine; ukrainian; laptop; laptops; scandal; scandals; conspiracy; conspiracies; corrupt; corruption; deal; deals; dealed; dealing; dealings; fbi; investigate; investigated; investigating; investigation; investigations; profit; profits; profited; profiting; lobby; lobbied; lobbyist; lobbying; burisma; energy; oil; gas; producer; producers; conglomerate; conglomerates; rudy; guiliani; social media; facebook; twitter
JFK	"" conspiracy; "" conspiracies; "" oliver stone; "" hsca; "" assassination; "" assassinations; "" mafia; "" cia; "" magic bullet; "" classified; "" declassified; "" de classified; "" hide; "" hides; "" plot; "" plots; "" scheme; "" schemes; "" coverup; "" cover up; "" truther; "" truthers	memory; memories; anniversary; anniv; airport; airports; airplane; airplanes; plane; planes; aviation; pilot; pilots; crew; crews; delay; delayed; delays; traffic; crash; crashes; flight; flights; takeoff; takeoffs; landings; departure; departures; arrival; arrivals; officer; officers; terror; terrorism; bomb; bombing	
9/11	"" conspiracy; "" conspiracies; "" conspire; "" conspires; "" conspired; "" coverup; "" cover up; "" cover ups; "" inside job; "" truth; "" truths; "" truther; "" truthers; "" theory; "" theories; "" theorist; "" theorist; "" secret; "" secrets; "" lie; "" lies; "" lied		
Anti-Vaccine	vaccine; vaccines; vaccination; vaccinations; vaccinated; vax; vaxx; vaxer; vaxxer; vaxed; vaxxed; jab; jabs; booster; boosters		anti-vaxx; antivaxx; anti-vax; antivax; anti-vaxxer; antivaxxer; anti-vaxer; antivaxer; conspiracy; conspiracies; anti-science; antiscience; scam; scams; plot; plots; poison; poisons; nano; fetus; fetal; misinformation; misinformed; disinformation; disinformed; danger; dangers; dangered; dangerous; endanger; endangers; endangered; endangerous; autism; autistic; deny; dinies; denier; deniers; denial; denials; chip; chips; magnet; magnetic; 5g; sterile; sterilize; sterilizes; sterilization; infertile; infertility; infertiliztion

Table 2 cont'd:

Conspiracy Theory	Query	Excludes	Includes
Moon Landing	"" conspiracy; "" conspiracies; "" conspired; "" hoax; "" fake; "" faked; "" lie; "" lies; "" lied; "" coverup; "" cover up; "" covered up; "" stage; "" staged; "" sibrel; "" theory; "" theories; "" theorist; "" theorists; "" secret; "" secrets; "" truth; "" truths; "" truther; "" truthers		
Lab Leak Hypothesis	lab leak		
2020 Election	"" fraud; "" frauds; "" stolen; "" steal; "" steals; "" rig; "" rigs; "" rigged; #stopthesteal		
Climate Change	climategate; climate hoax; climate hoaxes; climate conspiracy; climate conspiracies; climate coverup; climate cover up; climate cover ups; climate covered up; warminggate; warming hoax; warming hoaxes; warming conspiracy; warming conspiracies; warming coverup; warming cover up; warming cover ups; warming covered up		
Birtherism	birther; birthers; birtherism; obama birth; obama born; obama muslim; obama islam theory; obama islam theories; obama islam conspiracy; obama islam conspiracies; obama islamic theory; obama islamic theories; obama islamic conspiracy; obama islamic conspiracies; obama koran theory; obama koran theories; obama koran conspiracy; obama koran conspiracies; obama quran theory; obama quran theories; obama quran conspiracy; obama quran conspiracies; obama qu'ran theory; obama qu'ran theories; obama qu'ran conspiracy; obama qu'ran conspiracies	birthday; forced	birther; is a muslim; born in kenya; born outside the us; born outside the united states; not born in the us; not born in the united states; is not american; is not from the us; is not from the united states; is not an american
QAnon	qanon; q anon		
Russiagate	russiagate		
UFOs	ufo; uap		ufo
Flat Earth	flat earth	flat earth news	
Deflategate	deflategate		
Chemtrails	chemtrail; chemtrails	country club; cc	

B. Modeling

This chapter is a contribution to the conspiracy-theory-literature, not the machine learning literature. Thus, in this sub-section I emphasize my application of machine learning rather than the details of machine learning itself. I focus on the domain-specific variables that distinguish my application of machine learning from other applications. Those variables are a) cleaning; b) splitting; c) labeling; and d) details related to the model itself.

i. Cleaning

I clean the tweets to ensure that the computer can make sense of them. This process involves many small steps, the goal of which is to prevent the computer from being distracted by characters and words that should not help it identify stigma.

First, I convert all letters to lower-case so that the computer does not think that upper-case words are substantively distinct from lower-case words. Second, I replace urls with "HTTP" and Twitter usernames with "TwitterHandle." While the meaning of words tends to endure for decades or centuries, the stigma-content of urls and Twitter usernames is likely to ebb and flow as urls cease to exist and certain Twitter users become less active (and eventually die). Thus, a good model should not depend on these temporally unstable markers. Third, I remove punctuation. Punctuation marks like periods and commas are present in most tweets, so I don't want my model to depend on punctuation for its predictions. Fourth, I decompose contractions into their constituent parts. This ensures that the model does not treat "don't" and "do not" as distinct words. Fifth, I remove stop-words such as "a", "the", "but", "in", "how", etc. Like punctuation, these words are common yet convey almost no information relevant to stigma.

The last major step of text-cleaning is lemmatization. A word's lemma is its root. For example, the words "tweet," "tweets," "tweeted," and "tweeting" all share the lemma "tweet." In

most cases, different words with the same lemma have very similar meanings. Thus, it is better for a model to learn to discriminate between lemmas than between words. Accordingly, I lemmatize all cleaned words in my sample, thereby replacing words with their lemmas. There are several other components of the cleaning process that are not important for this chapter.

ii. Splitting

With clean tweets in hand, I remove from the full sample a training/validation set (8,789 tweets) and a test set (2,000 tweets). The training/validation set should be representative of the bulk of my sample, and the test set should be representative of future tweets to which the model may be applied. My subject presents unique challenges that lead me to diverge from best practice in some respects. I explore those divergences below, turning first to the training/validation set, and second to the test set.

To ensure that tweets in the training/validation set are amenable to training a model and assigning labels with confidence, I exclude tweets that reply to other tweets as well as tweets that contain urls. These tweets include context not available to my model, as the model receives only the words in the tweet itself. I retain retweets because 1) the entirety of the tweet is contained in the text available via Twitter's API and 2) retweets tend to indicate agreement with the retweeted content, so it is appropriate to treat them as duplicates of the original tweet (Metaxas et al. 2015).

To ensure that the model is generalizable across conspiracy theories, I over-sampled tweets referencing less popular conspiracy theories (e.g., Lab Leak Hypothesis) and under-sampled tweets referencing more popular conspiracy theories (e.g., QAnon). The number of tweets in my full sample, the training/validation set, and the test set are reported below in Table 3. Note that the sum of the "Train/Validation Count" column in Table 3 is 9,310, but I only use 8,789 tweets as the train/validation set in the modeling process. The larger sum in Table 3

reflects double counting. For example, a tweet that says, “The Hunter Biden laptop conspiracy theory and Lab Leak Hypothesis are both stupid!” is represented twice in the table, however my model only observes that tweet once.

Lastly, in line with previous work that uses methods like mine, I ensure that there are enough tweets with stigma by selecting only tweets that contain at least one word that most people would agree conveys stigma—words like “crazy,” “lunatic,” “wacko,” “nutjob,” “racist,” etc. (Maas et al. 2011). Without this criterion in place, 70-80% of tweets lack stigma, so the training/validation set may lack enough stigma-tweets from which it could learn to predict stigma.

Table 3

Conspiracy Theory	Full Sample Count	Train/Validation Count	Test Count
Hunter Biden	105,584	11	1,099
JFK	37,822	689	17
9/11	18,331	735	28
Anti-Vaccine	96,842	629	103
Moon Landing	9,934	716	6
Lab Leak Hypothesis	21,831	347	54
2020 Election	62,245	630	259
Climate Change	39,853	699	33
Birtherism	86,344	707	9
QAnon	183,168	716	300
Russiagate	42,119	672	27
UFOs	152,963	676	137
Flat Earth	47,829	722	34
Deflategate	106,934	669	4
Chemtrails	17,156	692	24

For the test set, I do not filter to avoid tweets with urls nor to gather tweets with obvious stigmatizing language. The model’s performance on the test set should represent its predicted performance on unlabeled tweets. For my model to remain useful over time, it must be able to predict the stigma associated with conspiracy theories that do not yet exist. To approximate this condition, about half of the tweets in the validation set reference conspiracy theories about Hunter Biden and his infamous “Laptop from Hell.” Tweets about these conspiracy theories are

absent from the training/testing set. This provides a difficult challenge for the model. If the model can successfully predict the stigma associated with tweets about a conspiracy theory it hasn't seen, then the model is likely to perform well when predicting the stigma associated with tweets about conspiracy theories that do not yet exist.

The other half of the validation set is representative of the full sample, though only from the time period following the tweets in the training/validation set. While tweets in the training/validation set span 2007 through March 27, 2022 and omit Hunter Biden tweets (unless those tweets also reference another conspiracy theory of interest), tweets in the test set are all from the period after March 27, 2022. Thus, when I use the model to make predictions for the test set, I approximate the best practice condition whereby models are tested for their ability to predict observations from the future—observations it could not possibly use to train.

iii. Labeling

With tweets split into a training/validation set and a test set, I label the original versions of those tweets according to the presence of stigma. I mark tweets with "1"—stigma—if I would fear social exclusion after tweeting endorsement of the conspiracy referenced in the tweet. I mark remaining tweets "0"—no stigma. Tweets vary in how well they conform to my labels. Tweets that ridicule a conspiracy theory or its believers are obvious 1s. Obvious 0s include tweets that endorse a conspiracy theory, tweets that claim that the tweeter is taking the conspiracy theory seriously, and tweets that merely advertise an upcoming article, interview, podcast, etc. that will discuss the conspiracy theory. Most tweets fit these obvious labeling rules.

To label remaining tweets, I rely on more subtle rules. For example, if a tweet claims that somebody is pushing or peddling a conspiracy theory, the tweet usually gets a 1. But if the tweet claims that somebody is advancing or offering a conspiracy theory, the tweet usually gets a 0. If

a tweet claims that a conspiracy theory is false or debunked, the tweet usually gets a 1. But if a tweet claims that a conspiracy theory is unproven or uncertain, the tweet usually gets a 0. And if a tweet claims that believers probably also believe many other conspiracy theories, the tweet usually gets a 1. But if the tweet does not dispute the conspiracy theory and merely proposes that believers should wait for more evidence to accumulate before they endorse the conspiracy theory, the tweet usually gets a 0.

iv. The Model

With labels in hand, I develop a classification model that trains on the tweets and labels in the training/validation set and tests itself on the test set. I use a bag of words approach to process tweets in such a way that a computer can make sense of them. The bag of words approach converts text data into a sparse matrix of 0s and 1s. Each row is a tweet, and each column is a word. Each tweet (row) has 1s for all of the words (columns) that appear in the tweet and 0s for all other words. Thus, my sample of tweets takes the form of approximately 10,000 binary features (words) used to predict approximately 10,000 binary labels (stigma). Two classification models are especially well suited to such a classification problem—Random Forest and Bernoulli Naïve Bayes'.

To decide between these models, I determine the strongest version of each model for my domain and then compare the strongest version of each model. To determine the strongest version of each model, I use repeated k-fold cross-validation while varying parameters likely to affect the model's accuracy. Ultimately, I find that default parameters (using the sklearn package in the python programming language) are suitable with three exceptions. The Random Forest model performed best using a maximum tree depth of 64 and 1000 trees. Higher values strain my

computer beyond usability. The Naïve Bayes' model performed best using an alpha (a smoothing parameter) of 10.

To compare the strongest versions of the Random Forest and Naïve Bayes' models, I again use repeated k-fold cross-validation. Performances are equivalent. Errors are more balanced (similar number of false positives (FPs) and false negatives (FN)) for the Random Forest model, so I select that model.

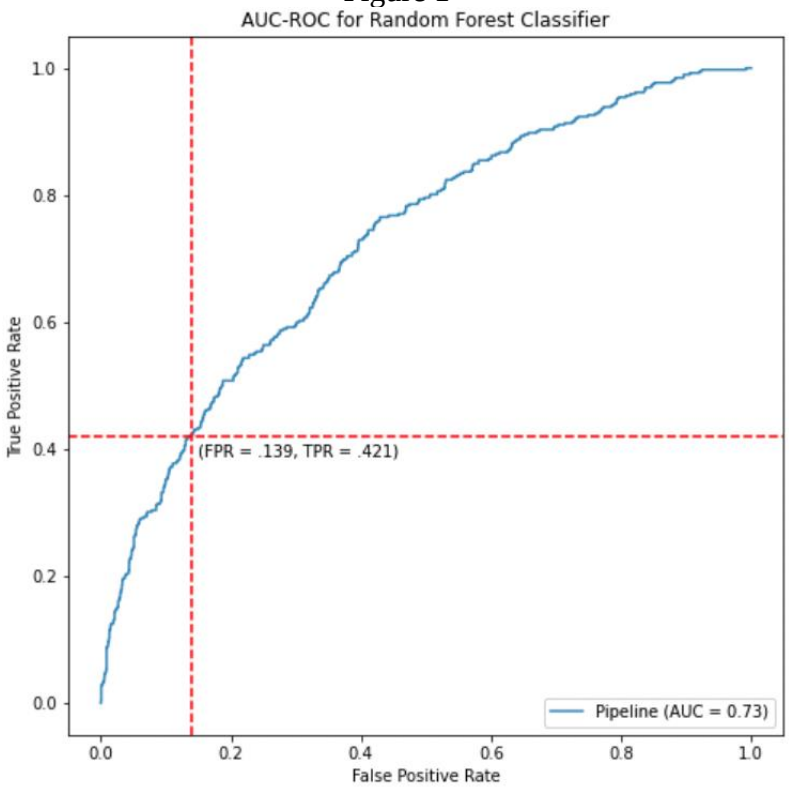
Because a Random Forest model utilizes randomness independent from the random state feature in sklearn's train-test-split function, repeated fittings of the Random Forest yield slightly different results. I fit the model repeatedly until I achieve a fitting with 1) an area under the receiver operating curve (AUC-ROC) greater than 0.72, 2) balanced errors, and 3) a ratio of 0s to 1s among my predictions of the test set that resembles the ratio of labels in the test set. Table 4 summarizes the model's performance in a confusion matrix.

Table 4

Confusion Matrix: Test Set		Predictions		
		0	1	Total
Labels	0	TN: 1387	FP: 223	1610
	1	FN: 227	TP: 165	392
	Total	1614	388	2002

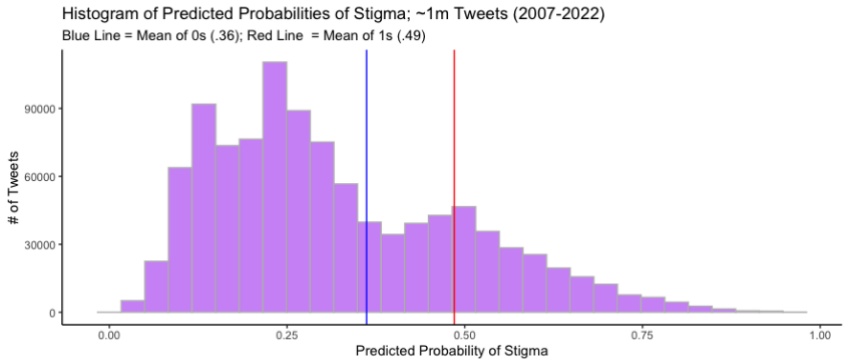
The ROC in Figure 1 generalizes from the confusion matrix to visualize the tradeoffs between the true positive rate (TPR: TP / (TP + FN)) and false positive rate (FPR: FP / (FP + TN)). The red lines on the plot mark the false positive rate and true positive rate when the model makes predictions for my validation set. Each point on the curve indicates a dyad of FPR and TPR that my model could yield if some probabilistic cutoff other than 0.5 were used to round predictions.

Figure 1



The model's performance is good enough for my purposes. No individual predicted probability of stigma should be trusted without validation by a human, but estimates for groups of tweets are trustworthy. A Student's t-test comparing the predicted probabilities of stigma among 0s and 1s in the test set shows a large and statistically significant difference (p-value approximately equal to 0) between the two groups. The average predicted probability of stigma among 0s in the test set is 0.362, and the average among 1s is 0.485. The 95% confidence interval for the difference between 0s and 1s ranges from 0.137 to 0.109. In Figure 2, I present a histogram of the predicted probability of stigma for all tweets in my sample. The blue line shows the average of the 0s in the test set, and the red line shows the average of the 1s in the test set.

Figure 2



In Table 5, I present a range of examples to demonstrate the model's performance. None of the tweets I review is in the training or testing set. The model predicts that tweets 1-4 are 0s, the model is uncertain about tweets 5-10, and the model predicts that tweets 11-14 are 1s. I display predicted probabilities rounded to the nearest thousandth in the column "Prediction." The column "Truth" displays how these tweets would be labeled by most humans. The model predicted green rows correctly and red rows incorrectly. I review seven of these rows below.

Tweet 1 advocates for the Lab Leak Hypothesis, so the model correctly predicts a probability of stigma close to 0. Tweet 3 advocates against the Lab Leak Hypothesis, but the model incorrectly predicts a probability close to 0. Tweet 5 weighs in on Twitter's censorship of the Lab Leak Hypothesis rather than the conspiracy theory itself, so the model correctly predicts a probability close to 0.5. Tweet 9 advocates against the Lab Leak Hypothesis, but the model incorrectly predicts a probability close to 0.5. Tweet 11 does not disparage the 2020 Presidential Election conspiracy theory, but the model incorrectly predicts a probability close to 1. Tweet 13 describes the Birther conspiracy theory as a racist witch hunt, so the model correctly predicts a probability close to 1.

Overall, consideration of the model's individual predictions in Table 5, the confusion matrix in Table 4, the ROC in Figure 1, and the t-test highlight the limitations and strengths of the model. On the one hand, review of Table 5 confirms the un-reliability of predictions for

individual tweets. On the other hand, review of Table 4, Figure 1, and the t-test confirms that the model provides value when estimating the stigma associated with a group of tweets. Analysis in this chapter relies on predictions for groups of tweets, thus avoiding the limitations of the model.

Table 5

Tweet	Prediction	Truth
1. If you consider yourself a reasonable person, favor a Lab! Leak! theory of Covid's origins, and haven't read this, you SHOULD. (tl;dr version: "New data presented by the WHO study provides clear findings in support of the natural origin of SARS-CoV-2.")	0.097	< 0.5
2. Amazing how many people savaged anyone who dared consider obvious potential for Wuhan lab leak theory or worse, evidently…	0.085	< 0.5
3. Experts say Wuhan lab leak theory is baseless and politically-motivated as some countries are blaming China without any scientifically validated evidence. #GLOBALink	0.093	> 0.5
4. It is a Sinophobic lie to even consider that a lab leak happened... Western Scientists: Right on! A na…	0.079	> 0.5
5. Facebook stopped banning, and started to allow, dissent on the lab-leak theory as soon as the Democratic Party decided it was a viable theory rather than a crazy conspiracy. Tech monopolies are censoring to aggrandize the demands of the dominant Party:	0.496	< 0.5
6. This thin gruel is evidence only of the increasing desperation of those peddling the false narrative that science shows SARS-CoV-2 entered humans through natural spillover.	0.501	< 0.5
7. In recent months, the idea that the pandemic started somehow in a laboratory — and perhaps involved an engineered virus — has gained traction, especially with President Joe Biden ordering a review of U.S. intelligence to assess the possibility in May.	0.492	~0.5
8. With Biden's order for a fresh investigation into the pandemic's origins, Facebook has opened the floodgates for new claims about the coronavirus.	0.503	~0.5
9. "That's the epicentre, that's where the virus definitely started spreading." A University of Sydney virologist has released new research that debunks conspiracy theories about Covid-19's origins.	0.508	> 0.5
10. So it's annoying to hear Bill Maher promptly pivot to former CDC director Robert Redfield's baseless belief that it came from a lab & complain "that became the conservative view." Hence, imaginary Twitter censorship. Fox News agrees with Bill: not cool	0.507	> 0.5
11. Republican Rep. Dan Crenshaw heckled for saying the 2020 election wasn't stolen.	0.871	< 0.5
12. What does it mean to be a Republican, post-Trump? A majority of GOP voters polled say that claiming the 2020 election was stolen is a defining characteristic.	0.912	< 0.5
13. If @realDonaldTrump isn't white supremacist, why did he use racial slurs against Elizabeth Warren, lead a racist birther witch hunt against Obama, push a Muslim ban, attack the NFL, defend Nazis, inspire the Klan, pardon a racial profiler, and call Omarosa a "dog"?	0.925	> 0.5
14. For all the delusional climate change deniers (they tend to think earth is flat, too)...read this & weep. Literally	0.925	> 0.5

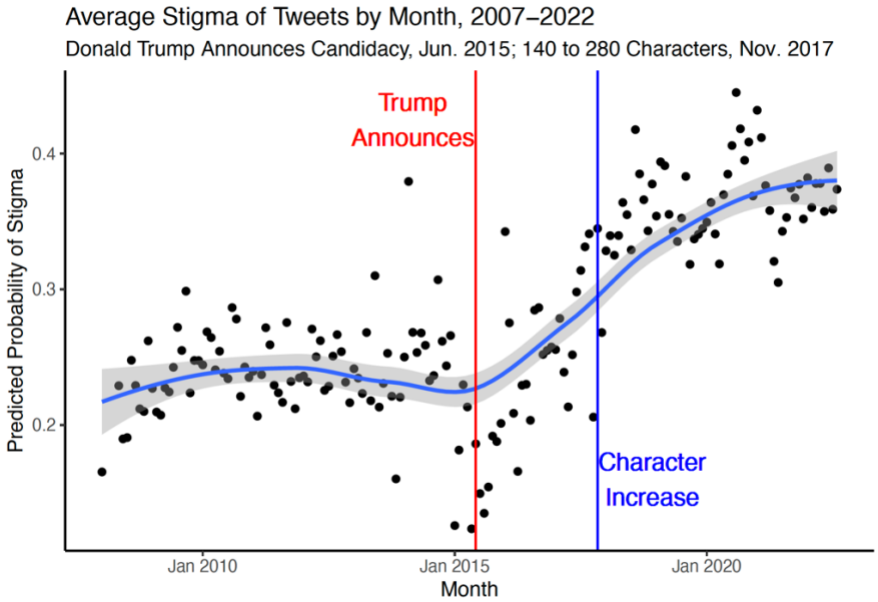
III. Analysis

A. Stigma Over Time

Barkun (2015) and Basham (2018) claim that the conspiracy-theory-stigma is decreasing. Thalmann (2019) tests this suggestion with detailed, qualitative review of the conspiracy-theory-stigma in the United States starting with JFK's assassination and ending in the mid-1990s. She finds that the stigma increased over that period. She points to elite-rhetoric, especially that of popular publications and public intellectuals, to demonstrate the trend. Given the nuances of the Internet Age that arrived on the heels of Thalmann's analysis, measurement of the conspiracy-theory-stigma morphs into a different task than Thalmann's. On the one hand, it is more difficult to measure the stigma now because today's society lacks central pillars of elite-opinion around which public discourse forms. On the other hand, it is easier to measure stigma now because Twitter provides researchers with access to all tweets published on its website. This allows researchers to measure stigma where it lives—in public discourse—rather than inferring it from elite-discourse as Thalmann and others do. I use predictions emerging from the model developed in Section II to test H1, H2, H3, and H4. I begin with H1: The conspiracy-theory-stigma is constant from 2007-2022.

Figure 3 shows the average predicted probability of stigma for every tweet in the sample, grouped by month. Many see Donald Trump's emergence on the political scene as a turning point in the centrality of conspiracy theorizing in the US (Uscinski 2016; DelReal 2016; Cillizza 2017; Beckwith 2018; Williams 2019), so I add the vertical red line to the graph to note the month during which he announced his candidacy for President. Additionally, the number of characters and words in a tweet is positively associated with predicted probability of stigma, so I add the vertical blue line to the graph to note the month during which Twitter doubled its character-limit from 140 to 280 characters. The blue curve and corresponding gray confidence interval indicate a polynomial of best fit relating month to predicted probability of stigma.

Figure 3



While the association between the length of tweets and the predicted probability of stigma complicates my inquiry into H1, the association is not a weakness of the model. Focusing on tweets that I labeled, the 1s averaged two more words than the 0s. Thus, the model is not in error when it assigns greater predicted probabilities of stigma to tweets of greater length. Rather, higher-stigma tweets tend to be longer than lower-stigma tweets.

Turning to the results of my test of H1, Table 6 (below) shows the influence of time on stigma. Holding constant the number of words in a tweet and the conspiracy theory with which the tweet is associated (fixed effects with "UFOs" as the reference-conspiracy-theory), predicted probability of stigma increases over time. Specifically, an additional year is associated with an additional 0.002 predicted probability of stigma. Since Trump announced his candidacy, the annual rate of increase doubled to 0.004. While these rates of increase are small, the finding provides some evidence against the claim that the stigma is decreasing (Barkun 2015; Basham 2018) and provides some evidence for the claim that the stigma is increasing. This allows me to reject H1.

Table 6

	Dependent variable:
	Predicted Probability of Stigma
Years Since 1st Tweet	0.002*** (0.00005)
Number of Words	0.005*** (0.00002)
2020 Election	0.237*** (0.001)
9/11	0.297*** (0.001)
Anti-Vaxx	0.108*** (0.0005)
Birtherism	0.141*** (0.0004)
Chemtrails	0.170*** (0.001)
Climate Change	0.119*** (0.001)
Deflategate	−0.046*** (0.0004)
Flat Earth	0.398*** (0.001)
Hunter Biden	0.140*** (0.0005)
JFK	−0.002*** (0.001)
Lab Leak	−0.023*** (0.001)
Moon Landing	0.126*** (0.001)
QAnon	0.290*** (0.0004)
Russiagate	0.064*** (0.001)
Constant	0.085*** (0.001)
Observations	1,028,891
R^2	0.643
Adjusted R^2	0.643
Residual Std. Error	0.104 (df = 1028874)
F Statistic	115,722.100*** (df = 16; 1028874)
Note:	*p<0.1; **p<0.05; ***p<0.01

Now that I have demonstrated the value of the model and its predictions in the context of a simple question like "Is the stigma increasing?" I turn my focus to the main aim of this chapter: analysis of a potential cause of the stigma and potential consequences of the stigma.

B. Causes and Consequences of Conspiracy-Theory-Stigma

Here, I test two hypotheses relevant to scholars of the conspiracy theory stigma as well as social media corporations:

H2: Events that ought to stigmatize a conspiracy theory do not increase the predicted probability of stigma present in tweets related to that conspiracy theory.

H3: Following an event that ought to stigmatize a conspiracy theory, tweets with higher predicted probabilities of stigma do not experience a greater increase in retweets than tweets with lower probabilities of stigma.

In the status quo, social media corporations censor social media posts, users, and groups according to the harm that they think that post, user, or group causes. Absent from these considerations is discussion of whether organic processes—such as the stigmatization of conspiracy theories following an event like the January 6 storming of the US Capitol—accomplish the goals of censorship. While many papers analyze the effectiveness with which censorship reduces interest in the censored content (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018), no research analyzes the effectiveness of organic processes of conspiracy-theory-stigmatization following an event that ought to stigmatize the conspiracy theory. Below, I test a cause (H2) and consequence (H3) of conspiracy-theory-stigma to gain insight into the extent to which organic processes alter conspiracy theory discourse.

Regarding H3, most retweets come from non-verified users. Non-verified users include bots along with real accounts. Accordingly, I expect non-verified users (and the retweets that they produce with respect to H3) to be less influenced by stigma than verified users (and the posts they produce with respect to H2). Henceforth emerges H4: It will not be easier to reject H2 than H3.

To test H2 and H3, I identify three sets of conspiracy theories that experience censorship on social media. For each set of conspiracy theories, I identify at least one stigmatizing event. Events must 1) be prominent enough to be perceived by many Twitter users and 2) seem likely to influence stigma in an obvious direction. Regarding 2020 Presidential Election conspiracy

theories, I focus on the effects of January 6, 2021. I expect that this event stigmatizes 2020 Presidential Election conspiracy theories by associating believers with the anti-social behaviors of the January 6 rioters. Regarding anti-vaccine conspiracy theories, I focus on the effect of Food and Drug Administration (FDA) approval of the Pfizer mRNA vaccine against COVID-19. I expect this event stigmatizes anti-vaccine conspiracy theories by discrediting claims that the vaccine is dangerous or ineffective. For QAnon, I spread my focus across seven events, which I review below in the sub-section dedicated to QAnon.

I define a "threshold-event" as an event that ought to stigmatize a conspiracy theory and its believers. I use threshold-events to divide tweets into those that come before the threshold-event (0s in my dataset) and those that come after the threshold-event (1s in my dataset). The storming of the Capitol and FDA approval of the Pfizer mRNA vaccine are examples of threshold-events. I define a "threshold-time" as the precise time, day, year, and month that correspond to a threshold event. January 6, 2021 at 14:17 is the accurate threshold-time for the threshold-event "January 6, 2021." For each threshold-event, I test the effects of various inaccurate threshold-times. I can reject H2 and H3 only if effects of different threshold-times vary in predictable ways—with stronger effects close to the true threshold-time or obvious explanations where that pattern does not hold.

After reviewing the effects of threshold-events, I provide one final test of H3 by analyzing the association between the predicted probability of stigma and Twitter-retweets across the full sample.

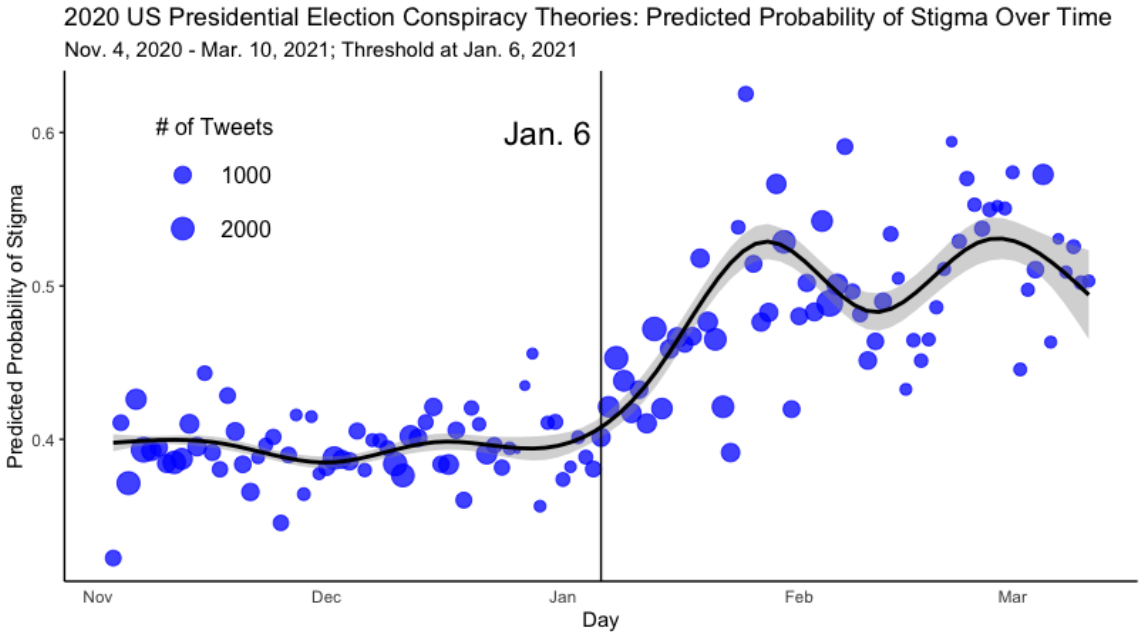
i. 2020 Presidential Election Conspiracy Theories

The violence and general disregard for rule of law practiced by those who stormed the US Capitol on January 6, 2021 should stigmatize conspiracy theories about the 2020 Presidential

Election. Thus, I expect the predicted probability of stigma associated with tweets about 2020 Presidential Election conspiracy theories to increase in the immediate aftermath of January 6 (H2). Additionally, while I expect all tweets about these conspiracy theories to experience a surge in retweets following January 6 due to increased attention the threshold-event draws to the conspiracy theories, I expect the size of that surge to be positively associated with the predicted probability of stigma for a tweet (H3). Relative to less stigmatized tweets, more stigmatized tweets should receive more additional retweets after January 6 than they did before.

Figure 4 shows that H2 is plausible. Each dot in Figure 4 is a day's tweets related to 2020 Presidential Election conspiracy theories. The location of the dot on the x-axis represents the day. The location of the dot on the y-axis represents the average predicted probability of stigma for tweets that day. The size of the dot represents the number of tweets contained within that dot. The vertical line marks January 6, 2021. The black curve is a polynomial of best fit relating date to predicted probability of stigma. The gray band surrounding the black line is a 95% confidence interval for the polynomial of best fit. The polynomial is agnostic about January 6, yet it clearly

Figure 4



identifies January 6 as a meaningful threshold.

Table 7 (below) presents regression results as well as sensitivity analysis. The highlighted row refers to the accurate threshold-time for the January 6 threshold-event. 14:17 is one hour after the first tweet by a verified user that mentions the storming of the capitol. Only after this time do the majority of tweets tagged "2020 Presidential Election" in my sample discuss the storming. Other rows in the table display results using inaccurate threshold-times: 6, 12, 24, 36, and 48 before and after the accurate threshold-time. If the effects in the rows closer to the accurate threshold-time are statistically significant and larger than effects in other rows, then I can confidently reject H2 and H3. Each row contains 48 hours of tweets: from 24 hours before the threshold-time to 24 hours after the threshold-time.

Equation 1 describes the "Stigma" column.

Equation 1: $$Y_{i,S,E,T} = \alpha_{E,T} + \beta_{E,T} X_{E,T,i} + \epsilon$$

For each threshold-time T corresponding to the threshold-event E in question, I display in the column "Stigma" the effect $\beta_{E,T}$ of proximity to (before or after) T on the predicted probability of stigma $Y_{i,S,E,T}$ for all tweets i within 24 hours of T. $X_{E,T,i}$ equals 0 if the tweet i was published after the threshold-time T, and $X_{E,T,i}$ equals 1 otherwise (Equation 1). For example, if tweets in the 24 hours after a threshold-time T have an average predicted probability of stigma of 0.6 and tweets in the 24 hours before the threshold-time have an average of 0.5, then the value in the "Stigma" column in the row corresponding to threshold-time T would be 0.1.

If the threshold-events I study indeed stigmatize belief in the corresponding conspiracy theories, then I expect the event to affect not only the average stigma-content of tweets (the "Stigma" column in the below tables; H2) but also the extent to which Twitter users retweet

stigma-tweets (relative to non-stigma tweets) after the threshold-event (the "Retweet" column in the below tables; H3). This encapsulates H3. Equations 2, 3, 4, and 5 codify H3.

Equation 2: $$Y_{i,R,E,T} = \alpha_{E,T} + \beta_{E,T}X_{E,T,i} + \beta_{S,T}X_{S,T,i} + \beta_{E\times S,T}(X_{E,T,i} \times X_{S,T,i}) + \epsilon$$

Equation 3: $$Y_{i,R,E,T,X_{S,T,i}=1} = \alpha_{E,T} + \beta_{E,T}X_{E,T,i} + \beta_{S,T} \times 1 + \beta_{E\times S,T}(X_{E,T,i} \times 1) + \epsilon$$

Equation 4: $$Y_{i,R,E,T,X_{S,T,i}=0} = \alpha_{E,T} + \beta_{E,T}X_{E,T,i} + \beta_{S,T} \times 0 + \beta_{E\times S,T}(X_{E,T,i} \times 0) + \epsilon$$

Equation 5: $$Retweets = Y_{i,R,T,X_{S,T,i}=1} - Y_{i,R,T,X_{S,T,i}=0}$$

Regarding the column "Retweets," for each threshold-time T corresponding to the threshold-event E in question, I first estimate the interaction effect $\beta_{E\times S,T}$ and direct effects of the threshold-time, $\beta_{E,T}$, and stigma, $\beta_{S,T}$, on the predicted number of retweets $Y_{i,R,E,T}$ for all tweets i within 24 hours of T (Equation 2). Then, I predict $Y_{i,R,E,T}$ where $X_{S,T,i}$ equals 1 (Equation 3) and where $X_{S,T,i}$ equals 0 (Equation 4). Thus, $Y_{i,R,T,X_{S,T,i}=1}$ is the effect of proximity to T for a tweet i with predicted probability of stigma equal to 1, and $Y_{i,R,T,X_{S,T,i}=0}$ is the effect of proximity to T for a tweet i with predicted probability of stigma equal to 0. In the tables below, I display in the column "Retweets" the estimand in Equation 3 minus the estimand in Equation 4. Consider the following example from Table 7.

The highlighted row in Table 7 (below) corresponds to the accurate threshold-time for the January 6 storming of the Capitol. The estimate in the "Retweets" column 2775 means that I expect the increase in retweets for tweets with a predicted probability of stigma of 1 to exceed the increase in retweets for tweets with a predicted probability of stigma of 0 by a margin of 2775 retweets. While all tweets about 2020 Presidential Election conspiracy theories got more retweets after the threshold-event than before due to the attention that the event drew to those tweets, highly stigmatized tweets experienced a greater increase in retweets. This helps us reject H3.

Table 7

Effect of January 6 on Stigma and Retweets

Rows Contain 48 Hours of Tweets; Threshold-Times EST

Threshold_Time	Count	Stigma	Retweets
-48 hrs: Jan. 4, 14:17	257	0.037*	2752**
-36 hrs: Jan. 4, 02:17	307	0.001	21
-24 hrs: Jan. 5, 14:17	419	-0.01	-246
-12 hrs: Jan. 6, 02:17	514	0.032**	17
-6 hrs: Jan. 6, 08:17	519	0.058***	691
Jan. 6, 14:17	495	0.049***	1158
+6 hrs: Jan. 6, 20:17	487	0.012	2407***
+12 hrs: Jan. 7, 02:17	454	0.02	2830***
+24 hrs: Jan. 7, 14:17	348	0.02	-774
+36 hrs: Jan. 8, 02:17	210	0.028	-2428*
+48 hrs: Jan. 8, 14:17	172	-0.016	2

Note: *p<0.1; **p<0.05; ***p<0.01

For three reasons, it is acceptable to observe large, statistically significant effects outside the highlighted row. First, for all sets of conspiracy theories and the corresponding tables below, all rows except the top two and bottom two represent a set of tweets that overlap with the set tweets represented by the highlighted row. My sets of tweets include 48 hours of tweets, but some of my threshold-times are less than 48 removed from the accurate threshold-time.

Second, regarding the 2020 Presidential Election conspiracy theories and Table 7 in particular, Twitter users at this time encountered events related to 2020 Presidential Election conspiracy theories but unrelated to the accurate threshold-time. On the morning of January 6, before anybody stormed the Capitol, President Donald Trump, Rudy Giuliani, Senator Mitch McConnell (R – KY), and Senator Ted Cruz (R – TX) all gave speeches related to 2020 Presidential Election conspiracy theories. Just as I expect the behavior of those who stormed the Capitol to stigmatize the conspiracy theories, I expect these speeches to stigmatize the conspiracy theories because they failed to advance evidence to support their allegations of fraud.

Third, for threshold-events that capture the news cycle (primarily January 6), a threshold-event can extend beyond the accurate threshold-time I use in the highlighted rows. This can make it seem as if an event like the storming of the Capitol transpired over multiple days rather than just a few hours. Those who followed the unfolding of the January 6 threshold-event likely followed not only the initial breaking news but also commentary over the following days. With these caveats in mind, I argue that Table 7 helps me reject H2 and H3.

Regarding H2, the strongest effects in the "Stigma" column are in response to use of the accurate threshold-time in the highlighted row and the "-6 hrs" threshold-time that takes place just before the January 6 speeches were made. It is tolerable that the effect for the "-12 hrs" threshold-time is statistically significant because half of the tweets in the "after" group for that row are the same as the tweets in the "after" group in the highlighted row. It is most important that effects closer to the highlighted row are more positive and more significant than effects further away. Regarding H2, Table 7 is inconclusive but helps me reject H2.

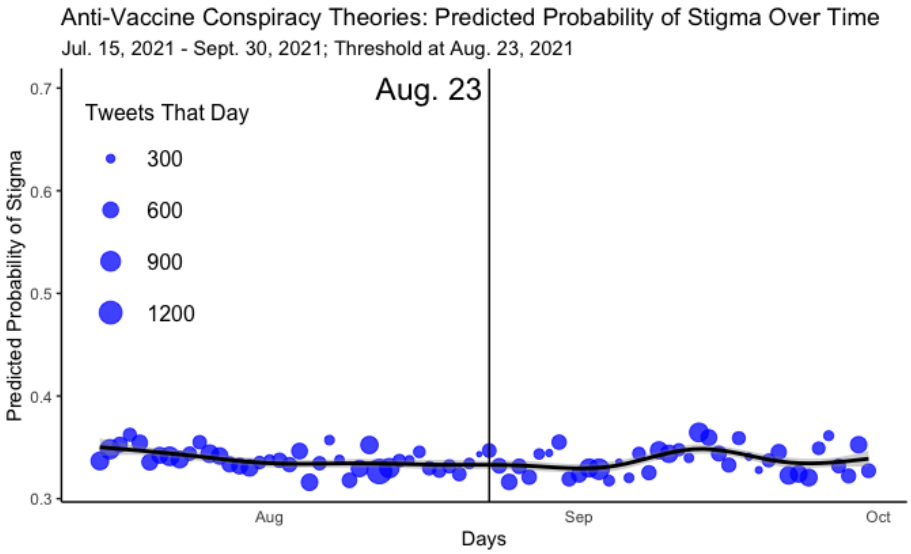
Regarding H3, we see that while strong, positive effects are close to the highlighted row, there is also such an effect in the "-48 hrs" row. This inhibits my ability to reject H3. On its own, Table 7 offers mixed, uncertain results. I can say little about H2 and H3 with reference to Table 7 alone. Next, I turn to analysis of anti-vaccine conspiracy theories on Twitter amid FDA approval of Pfizer's mRNA vaccine.

ii. Anti-Vaccine Conspiracy Theories

FDA approval of Pfizer's mRNA vaccine for COVID-19 serves as evidence against conspiracy theories that claim that the vaccine is dangerous and/or ineffective. Thus, I expect the stigma associated with tweets about Anti-Vaccine conspiracy theories to increase in the immediate aftermath of FDA approval (H2). Additionally, while I expect all tweets about these

conspiracy theories to experience a surge in retweets following FDA approval due to increased attention that the threshold-event draws to the conspiracy theories, I expect the size of that surge to be associated with the predicted probability of stigma for a tweet (H3). Relative to less

Figure 5



stigmatized tweets, more stigmatized tweets should receive more additional retweets after FDA approval than they did before. Figure 5 does not show a threshold-oriented pattern, but Table 8 shows a modest pattern.

Table 8

Effect of FDA Approval on Stigma and Retweets

Rows Contain 48 Hours of Tweets; Threshold-Times EST

Threshold_Time	Count	Stigma	Retweets
-48 hrs: Aug. 21, 08:41	366	0.035***	25**
-36 hrs: Aug. 21, 20:41	291	0.001	1280**
-24 hrs: Aug. 22, 08:41	304	-0.033***	-561
-12 hrs: Aug. 22, 20:41	409	0.007	-1138***
-6 hrs: Aug. 23, 02:41	406	0.034***	-1311***
Aug. 23, 08:41	460	0.023*	21
+6 hrs: Aug. 23, 14:41	533	0	-151
+12 hrs: Aug. 23, 20:41	580	-0.015*	-183
+24 hrs: Aug. 24, 08:41	579	-0.009	-193
+36 hrs: Aug. 24, 20:41	556	-0.012	-505
+48 hrs: Aug. 25, 08:41	536	-0.022**	-449*

Note: *p<0.1; **p<0.05; ***p<0.01

Regarding H2, the "Stigma" column of Table 8 shows that the highlighted row, the "-6 hrs" row, and the "-48 hrs" row show the strongest positive effects of the associated threshold-time on predicted probability of stigma. The strength of the estimates toward the middle of the table helps me reject H2. The strength of the "-48 hrs" estimate also helps us reject H2 because around that threshold-time several newspapers including the *New York Times* (Weiland and LaFraniere 2021) reported that the FDA "is aiming" to approve the vaccine.

Regarding H3, I observe no pattern in the "Retweets" column related to the accurate threshold-time. The only positive, large, statistically significant estimate is in the "-36 hrs" row. Thus, I conclude that use of the FDA approval threshold and results in Table 6 helps me reject H2 but does not help me reject H3. This helps me reject H4.

iii. QAnon

Regarding QAnon, I focus on seven distinct events that should affect stigma in an obvious direction. Below, I review the seven thresholds and associated results in the order in which the threshold-events occurred: 1) Sarah Ashcraft accused Tom Hanks of pedophilic sex-crimes; 2) Jeffrey Epstein was arrested for various pedophilic sex-crimes; 3) QAnon-inspired rumors circulating on Twitter caused "Oprah Winfrey" to trend; 4) Facebook's internal investigation showed that Facebooks hosts millions of pro-QAnon accounts; 5) President Donald Trump stated during a Town Hall that QAnon is "…very strongly against pedophilia…"; 6) the January 6, 2021 storming of the US Capitol; and 7) QAnon believers congregated in Dealey Plaza in Dallas, Texas—where JFK was assassinated—to await the appearance of the deceased JFK Jr.

These thresholds span three types which are denoted in the "Type" column of Table 6. First, five of the thresholds, like January 6, 2021 in relation to 2020 Presidential Election

conspiracy theories, demonstrate "Negative Association." These thresholds associate the conspiracy theories and belief therein with people, behaviors, and/or ideas that it is reasonable to assume most Twitter users consider negative (violence, absurdity, etc.).

Second, threshold 5—Trump's televised association of QAnon with anti-pedophilic advocacy—demonstrates "Positive Association" because Trump associates QAnon with anti-pedophilic advocacy. Accordingly, I expect the stigma associated with QAnon-related conspiracy theories to decrease following Trump's declaration.

Third, threshold 2—Jeffrey Epstein's arrest—is "Positive Evidence" because Epstein's arrest constitutes authoritative circumstantial evidence in favor of the QAnon-associated claim that powerful actors engage in pedophilia. Since I expect stigma against Anti-Vaccine conspiracy theories to increase following the revelation of evidence against Anti-Vaccine conspiracy theories (FDA approval), then I must also expect stigma against QAnon-related tweets to decrease following evidence in favor of QAnon-related conspiracy theories (Epstein's arrest).

Table 9

Effect of 7 QAnon Thresholds on Stigma and Retweets

Intervals Vary; All Times EST

Event	Threshold_Time	Interval	Count	Type	Stigma	Retweets
1. Sarah Ashcraft Accuses Tom Hanks; MSM Pokes Fun	2018-07-30 11:23	+/- 48 hrs	353	Negative Association (+)	0.058*	534
2. Jeffrey Epstein Arrested	2019-07-06 20:10	+/- 72 hrs	259	Positive Evidence (-)	-0.005	-781**
3. QAnon Rumors Cause Oprah to Trend	2020-03-17 23:44	+/- 48 hrs	174	Negative Association (+)	0.052***	184
4. FB Investigation Shows Millions of QAnon Accounts; MSM Pokes Fun	2020-08-10 15:15	+/- 24 hrs	644	Negative Association (+)	0.013	72
5. Trump Says QAnon is '...very strongly against pedophilia...'	2020-10-15 19:15	+/- 24 hrs	3345	Positive Association (-)	-0.022***	-562***
6. Jan. 6 Storming of the U.S. Capitol	2021-01-06 14:17	+/- 24 hrs	1945	Negative Association (+)	-0.011	252*
7. JFK Jr. Resurrection	2021-11-01 20:00	+/- 24 hrs	601	Negative Association (+)	-0.008	352**

Note: *p<0.1; **p<0.05; ***p<0.01

All seven cases yield promising results for my ability to reject H2 and H3. There are no large nor statistically significant effects in the opposite direction of my expectations. In four of the seven cases (1, 2, 4, and 5), effects of stigma and retweets are both in the expected direction. I review each case below.

Tom Hanks: On July 30, 2018, Sarah Ashcraft accused Tom Hanks of pedophilic sex-crimes. Ashcraft alleged that her parents sold her to Hanks when she was 13 years old. Ashcraft provided no concrete evidence to support her claim. This lack of evidence, Hanks's lack of criminal history, the severity of the alleged crimes, and the virality of the allegations seemed to trigger a wave of mainstream journalistic coverage of QAnon (Wong 2018; Bank, Stack, and Victor 2018; Stanley-Becker 2018; Carter 2018; Cassin and Wendling 2018; Hall 2018; Cillizza 2018; May 2018; Woodruff 2018; Bort 2018). This wave—confined to a 48-hour period—introduced the mass public to QAnon as a false, incoherent, pro-Trump conspiracy theory. Thus, I expect this threshold-event to stigmatize tweets about QAnon. After the threshold, we find that the stigma associated with tweets about QAnon increased (H2) and that retweets for stigmatized tweets increased more than retweets for non-stigmatized tweets (H3). This helps me reject H2 and H3. The higher statistical significance for the estimate related to H2 relative to the estimate for H3 helps me reject H4.

Jeffrey Epstein: On July 6, 2019, Jeffrey Epstein was arrested for alleged sex trafficking and sex with minors. Epstein was a popular New York City socialite, thus placing him into contact with other public figures who some QAnon-believers claim are also guilty of pedophilic sex-crimes (Levenson 2019). While Epstein's arrest and social network should not on their own motivate belief in QAnon's claims, the arrest nonetheless evidence in favor of the conspiracy theory. Therefore, I expect the arrest to de-stigmatize QAnon-related conspiracy theories. After

the threshold, we find that the predicted probability of stigma associated with tweets about QAnon decreased to a minute and statistically insignificant extent (H2). On the other hand, retweets for stigmatized tweets decreased more than retweets for non-stigmatized tweets to a large and statistically significant extent (H3). This result helps me reject H3 but not H2. It also inhibits my ability to reject H4.

Oprah Winfrey: Late on March 17, 2020, Oprah Winfrey began to trend on Twitter in association with QAnon-instigated rumors that claimed she was involved with sex-related human trafficking of minors. While Epstein 1) pleaded guilty to a state charge of procuring for prostitution a girl below age 18 (Goldsmith 2008) and 2) was arrested on related charges in 2019, Winfrey had no such track record, and allegations of her wrongdoing contained no non-circumstantial evidence. Thus, allegations against Winfrey sound *prima facie* absurd, and I expect that her trending on Twitter should stigmatize QAnon-related conspiracy theories by associating QAnon with absurd, unevidenced allegations. After the threshold, we find that the stigma associated with tweets about QAnon increased (H2) and that retweets did not change in a statistically significant way (H3). This result helps me reject H2 but does not help me reject H3. This result also helps me reject H4.

Facebook Internal Investigation: On August 10, 2020, Facebook revealed the results of an internal investigation. Facebook found that millions of its users were members of pro-QAnon groups. This sparked another wave of mainstream coverage of the QAnon-phenomenon (Sen and Zadrozny 2020; Canales 2020; Choudhary 2020; Rodrigo 2020; Alexander 2020; Wong 2020; Ecarma 2020; Coleman 2020; Mihalcik 2020). Just as in the case of the Tom Hanks threshold and associated wave of mainstream coverage, this wave uniformly criticized QAnon. Again, this wave was confined to a 48-hour period. Thus, I expect that Facebook's announcement and

associated mainstream coverage should increase the presence of stigma in tweets about QAnon by associating its claims and its believers with traits that most Twitter users consider negative. After the threshold, we find that the stigma associated with tweets about QAnon increased (H2) and that retweets for stigmatized tweets increased more than retweets for non-stigmatized tweets (H3). Neither estimate is statistically significant, but this result does not inhibit my ability to reject H2 and H3. This result provides no evidence related to H4.

President Donald Trump: On October 15, 2020, President Donald Trump said in a Town Hall that QAnon is "[…]very strongly against pedophilia[…]" In so doing, Trump associated QAnon with an agreeable, anti-pedophilia stance. Thus, I expect that Trump's statement should de-stigmatize QAnon-related conspiracy theories. After the threshold, we find that the stigma associated with tweets about QAnon decreased (H2) and that retweets for stigmatized tweets decreased more than retweets for non-stigmatized tweets (H3), however only the estimate related to stigma was statistically significant. This helps me reject H2 and H3 and does not help me reject H4.

January 6, 2021: The presence of QAnon symbology in the coverage associated with the storming of the US Capitol should stigmatize QAnon-related conspiracy theories in the aftermath of January 6. After the threshold, we find that the stigma associated with tweets about QAnon decreased insignificantly (H2) and that retweets for stigmatized tweets significantly increased more than retweets for non-stigmatized tweets (H3). This result does not help me reject H2, but it does help me reject H3. Thus, this result inhibits my ability to reject H4.

JFK Jr.: On November 1, 2021, QAnon-supporters began to congregate in Dealey Plaza in Dallas, Texas, the locus of JFK's assassination. The supporters awaited the appearance of the deceased JFK Jr., whom they expected to announce that he would be Trump's running-mate for

the 2024 Presidential Election. JFK Jr. never appeared and the impetus for the congregation was *prima facie* absurd, so I expect that this event should stigmatize QAnon-related conspiracy theories. I find that the stigma associated with tweets about QAnon decreased insignificantly after the threshold (H2). On the other hand, retweets for stigmatized tweets increased more than retweets for non-stigmatized tweets (H3), and this change was statistically significant. Thus, while this result helps me reject H3, it does not help me reject H2. Again, this result inhibits my ability to reject H4. Above, I tested H2, H3, and H4 one set of conspiracy theories at a time. Below, I ignore boundaries between conspiracy theories and test H3 on the entire sample of tweets.

iv. Full Sample

The threshold-oriented approach provides inconsistent evidence that we can reject H3, but the balance of evidence across many thresholds supports rejection of H3. The threshold-oriented approach is difficult in the context of social media activity because the researcher cannot control for events aside from the threshold that took place in a Twitter user's life around the time of the threshold. Below, I provide a threshold-agnostic approach to testing H3. I perform a linear regression to test whether increases in a tweet's predicted probability of stigma are positively associated with the retweets that tweet receives.

Holding constant time (because tweets get more retweets as Twitter ages and more users join), the number of words in the tweet (because the number of words is associated with the predicted probability of stigma), and fixed effects for each conspiracy theory (with UFO-related conspiracy theories as the reference), I find that increasing the predicted probability of stigma from 0 to 1 is associated with a decrease of 51 retweets for the tweet. Overall fit for this model is low, with an R-squared of .018, but the combination of this broad association and the mixed

results for H3 when scrutinizing one set of conspiracy theories at a time causes me to fail to reject H3.

Table 10:

	Dependent variable:
	Twitter Retweets
Pred. Prob. of Stigma	−50.594*** (6.958)
Years Since 1st Tweet	7.181*** (0.326)
Number of Words	1.871*** (0.136)
2020 Election	109.073*** (4.126)
9/11	40.194*** (6.117)
Anti-Vaxx	4.635 (3.484)
Birtherism	75.197*** (3.306)
Chemtrails	5.568 (6.039)
Climate Change	75.357*** (4.221)
Deflategate	2.350 (2.971)
Flat Earth	39.894*** (4.760)
Hunter Biden	302.285*** (3.381)
JFK	7.664* (4.224)
Lab Leak	10.242* (5.535)
Moon Landing	17.114** (7.662)
QAnon	42.875*** (3.500)
Russiagate	172.187*** (4.113)
Constant	−64.636*** (3.826)
Observations	1,028,891
R^2	0.018
Adjusted R^2	0.018
Residual Std. Error	734.953 (df = 1028873)
F Statistic	1,116.486*** (df = 17; 1028873)
Note:	*p<0.1; **p<0.05; ***p<0.01

IV. Discussion

Social media corporations censor what they identify as misinformation related to the 2020 Presidential Election, vaccines, and QAnon (among other topics). In this chapter, I find

evidence that this censorship may not be as necessary as the censors claim. I introduce, test, and tentatively reject three hypotheses to arrive at this conclusion:

> H1: The conspiracy-theory-stigma is constant from 2007-2022.
>
> H2: Events that ought to stigmatize a conspiracy theory do not increase the predicted probability of stigma present in tweets related to that conspiracy theory.
>
> H3: Following an event that ought to stigmatize a conspiracy theory, tweets with higher predicted probabilities of stigma do not experience a greater increase in retweets than tweets with lower probabilities of stigma.
>
> H4: It will not be easier to reject H2 than H3.

In testing H1, I find that the conspiracy-theory-stigma is on the rise. I find that much of this rise is concentrated after President Trump announced his candidacy for President of the US in 2015. Thus, it is at least incomplete if not incorrect to state that the Trump-era is the Golden Age of conspiracy theories (Stanton 2020; Willingham 2020; *Economist* 2021; Schaefer 2022). That is, while belief in, the breadth of, and the importance of conspiracy theories may be on the rise, the stigma against conspiracy theories is *also* on the rise.

In a world of high conspiracy-theory-stigma, public endorsements of conspiracy theories would become opportunities for public shaming that, more than censorship, would reduce belief in and disincentivize subsequent endorsement of conspiracy theories. While censorship likely steers conspiracy theorists toward platforms that do not censor (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018), stigma paired with lack of censorship allows for rationalizing and ridiculing that evidence shows can reduce conspiracy-theory-belief (Orosz et al. 2016). This chapter presents further evidence that these forces may be at work, but censorship interferes with the ability to study the matter further.

I test H2 and H3 by utilizing thresholds that should be associated with changes in the extent of stigma associated with conspiracy theories. I show that when an event ought to stigmatize a conspiracy theory, tweets related to that conspiracy theory contain more stigma after the event than before the event. I show the reverse pattern as well (less stigma in tweets following a de-stigmatizing event). Thus, I reject H2. Further, I show that when an event ought to stigmatize a conspiracy theory, tweets that contain more stigma do not experience greater increases in performance (measured in retweets) than tweets that contain less stigma. I also show a negative correlation between stigma and retweets regardless of thresholds. Thus, I fail to reject H3. By rejecting H2 and failing to reject H3, I reject H4.

This evidence shows that verified Twitter users (though perhaps not non-verified users) respond to novel information approximately how Twitter would like them to. Just as Twitter responds to certain events with a desire to censor discussion thereof, verified Twitter users seem to respond to the same events by stigmatizing that which Twitter would like to censor. Thus, to some extent at least, verified Twitter users' organic responses to information reduce the need for censorship. Non-verified users, however, fail to respond to novel information as Twitter would like them to. Most Twitter users are non-verified, so I cannot conclude that a laissez-faire approach to censorship will cause the reduced access to and interest in conspiracy theories that social media corporations seek to cause via censorship.

Regardless, if social media censorship of misinformation continues, users who disagree with this policy or its application will likely migrate toward less restrictive platforms. Thus, it is not clear how censorship reduces the harms that social media corporations claim conspiracy theories cause. Rather, it seems that censorship merely relocates that harm. One may argue that pushing conspiracists to smaller social media websites reduces the alleged harm by reducing the

number of users who see tweets that endorse conspiracy theories. To the contrary, I make two arguments.

First, if smaller social media websites become places where a high proportion of users endorse these conspiracy theories, believers are less likely to encounter ridiculing and rationalizing arguments against conspiracy theories. Those anti-conspiracist arguments effectively reduce belief in conspiracy theories (Orosz et al. 2016). Second, studies show that attempts to restrict access to information tend to increase access to and interest in that information (Ashmore, Ramchandra, and Jones 1971; Brehm and Brehm 1981; Nabi 2014; Kwon, Moon, and Stefanone 2015; Hobbs and Roberts 2018). Thus, even if the laissez-faire approach (non-censorship) does not accomplish censors' goals, censorship is even less likely to accomplish those goals. The censors offer no rationale that incorporates these criticisms. Why, then, do they censor? I offer one possible answer.

In the status quo, journalists and others imply a false dichotomy that may motivate the censorship practices of social media corporations. Journalists argue or imply that social media's share of the blame for the spread of conspiracy theories is inversely associated with the extent to which they censor that misinformation (Whitten 2015; Menczer 2016; Carey 2017; Meserole 2018; Yaraghi 2019; Bond 2020; Suciu 2021; Myers 2022). I propose that we take a more nuanced approach to the complicated matter of censorship of conspiracy theories. In so doing, we may create the incentive for social media corporations to censor rationally and in alignment with academic scholarship.

Bibliography

@andymstone. October 14, 2020. *Twitter*. Date of access: May 16, 2022. https://twitter.com/andymstone/status/1316395902479872000?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E1316395902479872000%7Ctwgr%5E%7Ctwcon%5Es1_&ref_url=https%3A%2F%2Fwww.washingtonexaminer.com%2Fnews%2Ffacebook-reduces-visibility-on-hunter-biden-story-before-fact-checking

Alexander, Harriet. August 10, 2020. "Facebook internal investigation uncovers thousands of QAnon pages on the site followed by MILLIONS of people." *Daily Mail*. Date of access: October 3, 2022. https://www.dailymail.co.uk/news/article-8613357/Facebook-internal-investigation-uncovers-millions-QAnon-conspiracy-theory-members.html.

Amazeen, Michelle. 2015. "Revisiting the epistemology of fact-checking." Critical Review 27 (1): 1-22. https://doi.org/10.1080/08913811.2014.993890.

"An update to how we address movements and organizations tied to violence." Updated: November 9, 2021. Original publication date: August 19, 2020. *Meta*. Date of access: May 24, 2022. https://about.fb.com/news/2020/08/addressing-movements-and-organizations-tied-to-violence/.

Andrew, Scottie. September 18, 2019. "The US Navy just confirmed these UFO videos are the real deal." *CNN*. Date of access: January 29, 2022. https://www.cnn.com/2019/09/18/politics/navy-confirms-ufo-videos-trnd/index.html.

Ashmore, Richard D., Vasantha Ramchandra, and Russell A. Jones. 1971. "Censorship as an attitude change induction." Eastern Psychological Association Conference. [Speech given before Eastern Psychological Association Annual Meeting (42nd, New York, N.Y., April 15-17, 1971)] https://eric.ed.gov/?id=ED053411.

Alfonesca, Kiara. September 11, 2021. "20 years after 9/11, Islamophobia continues to haunt Muslims." *ABC News*. Date of access: May 19, 2022. https://abcnews.go.com/US/20-years-911-islamophobia-continues-haunt-muslims/story?id=79732049.

Badshah, Mashood, Jackson Shriver, Beatrice Rynders, Alexander Sjovold, Muhammad Hamza Saad Shaukat, and Naveen Rajpurohit. 2021. "MODERNA mRNA-1273 vaccine-associated myopericarditis in a patient with a subclinical autoimmune predisposition." Journal of Cardiology Cases 24 (5): 227-229. https://doi.org/10.1016/j.jccase.2021.09.007.

Bank, Justin, Liam Stack, and Daniel Victor. August 1, 2018. "From 2018: Explaining QAnon, the Internet conspiracy theory that showed up at a Trump rally." *The New York Times*. Date of access: October 2, 2022. https://www.nytimes.com/2018/08/01/us/politics/what-is-qanon.html.

Barkoukis, Leah. May 18, 2022. "Audit reveals shocking percentage of Biden's Twitter followers are fake." *Townhall*. Date of access: November 18, 2022.

https://townhall.com/tipsheet/leahbarkoukis/2022/05/18/audit-half-of-bidens-twitter-followers-are-fake-n2607432.

Barkun, Michael. 2015. “Conspiracy theories as stigmatized knowledge.” Diogenes 62 (3-4): 114-120. https://doi.org/10.1177/0392192116669288.

Basham, Lee. 2003. “Malevolent global conspiracy.” Journal of Social Philosophy 34 (1): 91-103. https://doi.org/10.1111/1467-9833.00167.

Basham, Lee. 2017. “Pathologizing open societies: a reply to the Le Monde social scientists.” Social Epistemology Review and Reply Collective, 6 (2): 59-68.

Basham, Lee and M R. X. Dentith. 2018. “The psychologists’ conspiracy panic: They seek to cure everyone.” In *Taking Conspiracy Theories Seriously*, ed. M R. X. Dentith. Collective Studies in Knowledge and Society, eds. James H. Collier, 79-92. Lanham, MD: Rowman & Littlefield International. Kindle.

Baurmann, Michael. 2007. “Rational fundamentalism? An explanatory model of fundamentalist beliefs.” Episteme 4 (2): 150-166. https://doi.org/10.3366/epi.2007.4.2.150.

BBC Trending. June 18, 2018. “Galloway's war of words with a mystery Wikipedia editor.” *BBC News*. Date of access: February 12, 2022. https://www.bbc.com/news/blogs-trending-44495696.

Beckett, Lois. October 16, 2020. “QAnon: A timeline of violence linked to the conspiracy theory.” *The Guardian*. Date of access: April 13, 2022. https://www.theguardian.com/us-news/2020/oct/15/qanon-violence-crimes-timeline.

Beckwith, Ryan Teague. August 2, 2018. “Donald Trump Weaponized Conspiracy Theories. QAnon Shows We're Now in an Arms Race.” *Time*. Date of access: October 5, 2022. https://time.com/5356443/trump-conspiracy-theories-qanon/.

Benxi, Aria. April 27, 2022. “Fauci says US is transitioning out of 'pandemic phase'.” *NBC News*. Date of access: May 24, 2022. https://www.nbcnews.com/health/health-news/fauci-says-us-moving-pandemic-stage-control-phase-rcna26236.

Bernstein, Carl. October 20, 1977. “The CIA and the media: How Americas most powerful news media worked hand in glove with the Central Intelligence Agency and why the Church Committee covered it up.” *The Rolling Stone*. Date of access: February 5, 2022. https://www.carlbernstein.com/the-cia-and-the-media-rolling-stone-10-20-1977.

Biddle, Sam. “Facebook's Tamil censorship highlights risks to everyone.” *The Intercept*. January 11, 2022. Date of access: January 19, 2022. https://theintercept.com/2022/01/19/facebook-tamil-censorship-sri-lanka/.

Biesecker, Michael, Michael Kunzelman, Gillian Flaccus, and Jim Mustian. January 10, 2021. “Records show fervent Trump fans fueled US Capitol takeover.” *Associated Press*. Date

of access: February 28, 2022. https://apnews.com/article/us-capitol-trump-supporters-1806ea8dc15a2c04f2a68acd6b55cace.

Bond, Shannon. September 22, 2020. “What can social media do to slow down the spread of misinformation?” *NPR*. Date of access: October 4, 2022. https://www.npr.org/2020/09/22/915555286/what-can-social-media-do-to-slowdown-the-spread-of-misinformation.

Bort, Ryan. August 1, 2018. “As QAnon goes mainstream, Trump’s rallies are turning darker.” *Rolling Stone*. Date of access: October 2, 2022. https://www.rollingstone.com/politics/politics-news/trump-qanon-705425/.

Bratich, Jack. 2002. “Injections and truth serums: AIDS conspiracy theories and the politics of articulation.” In *Conspiracy Nation: The Politics of Paranoia in Postwar America,* ed. Peter Knight, 133-156. New York, NY: New York University Press.

Bratich, Jack Z. 2008. *Conspiracy Panics: Political Rationality and Popular Culture*. Albany, NY: State University of New York Press.

Brehm, Sharon S. and Jack W. Brehm. 1981. *Psychological Reactance: A Theory of Freedom and Control*. New York, NY: Academic Press.

Brotherton, Robert, Christopher C. French, and Alan D. Pickering. 2013. “Measuring belief in conspiracy theories: The generic conspiracist beliefs scale.” Frontiers in Psychology 4 (279): 1-15. https://doi.org/10.3389/fpsyg.2013.00279.

Brown, Alleen. March 23, 2019. “The Green Scare.” *The Intercept*. Date of access: May 19, 2022. https://theintercept.com/2019/03/23/ecoterrorism-fbi-animal-rights/.

Bryant, Nick. January 22, 2015. “Flight logs put Clinton, Dershowitz on pedophile billionaire’s sex jet.” *Gawker*. Date of access: May 26, 2022. https://www.gawker.com/flight-logs-put-clinton-dershowitz-on-pedophile-billio-1681039971.

Buenting, Joel and Jason Taylor. 2010. “Conspiracy theories and fortuitous data.” Philosophy of the Social Sciences 40 (4): 567-578. https://doi.org/10.1177/0048393109350750.

Bush, George W. November 10, 2001. “President Bush speaks to United Nations.” Remarks by the President to United Nations General Assembly, UN Headquarters, New York, NY. Date of access: February 23, 2022. https://georgewbush-whitehouse.archives.gov/news/releases/2001/11/20011110-3.html.

Bush, George W. January 29, 2022. “State of the Union address.” *Miller Center*. https://web.archive.org/web/20111011053416/http://millercenter.org/president/speeches/detail/4540.

Canales, Katie. August 10, 2020. “An internal Facebook audit reportedly shows QAnon groups have millions of members, but some employees who ran the investigation fear the company won't take any action.” *Business Insider*. Date of access: October 3, 2022.

https://www.businessinsider.com/facebook-qanon-internal-audit-millions-of-members-groups-report-2020-8.

Carey, Benedict. October 20, 2017. “How fiction becomes fact on social media.” *The New York Times*. Date of access: October 4, 2022. https://www.nytimes.com/2017/10/20/health/social-media-fake-news.html.

Carter, Brandon. August 2, 2018. “What is QAnon? The conspiracy theory tiptoeing into Trump world.” *NPR*. Date of access: October 2, 2022. https://www.npr.org/2018/08/02/634749387/what-is-qanon-the-conspiracy-theory-tiptoeing-into-trump-world.

Cassam, Quassim. 2015. “Stealthy vices.” Social Epistemology Review and Reply Collective 4 (10): 19-25. http://wp.me/p1Bfg0-2na.

Cassam, Quassim. 2019. *Conspiracy Theories*. Cambridge: Polity Press. Kindle.

Cassin, Elizabeth and Mike Wendling. August 2, 2019. “QAnon: What's the truth behind a pro-Trump conspiracy theory?” *BBC*. Date of access: October 2, 2022. https://www.bbc.com/news/blogs-trending-45040614.

Chouchana, Laurent, Alice Blet, Mohammad Al-Khalaf, Tahir S. Kafil, Girish Nair, James Robblee, Milou-Daniel Drici, Marie-Blanche Valnet-Rabier, Joëlle Micallef, Francesco Salvo, Jean-Marc Treluyer, and Peter P. Liu. 2021. “Cardiac inflammation after COVID-19 mRNA vaccines: a global pharmacovigilance analysis.” Clinical Pharmacology & Therapeutics 111 (3): 605-613. https://doi.org/10.1002/cpt.2499.

Choudhary, Vidhi. August 10, 2020. “Facebook reportedly has over 3 million QAnon users, earned $12,000 from QAnon ads over past 30 days.” *Forbes*. Date of access: October 3, 2022. https://www.forbes.com/sites/vidhichoudhary/2020/08/10/facebook-reportedly-has-over-3-million-qanon-users-earned-12000-from-qanon-ads-over-past-30-days/?sh=5ed5edd257f8.

Cillizza, Chris. March 4, 2017. Donald Trump was a conspiracy theory candidate. Now he’s on the verge of being a conspiracy theory president.” *The Washington Post*. Date of access: October 5, 2022. https://www.washingtonpost.com/news/the-fix/wp/2017/03/04/donald-trump-was-a-conspiracy-theory-candidate-now-hes-on-the-verge-of-being-a-conspiracy-theory-president/.

Cillizza, Chris. August 1, 2018. “QAnon is the one conspiracy theory to rule them all.” *CNN*. Date of access: October 2, 2022. https://www.cnn.com/2018/08/01/politics/qanon-trump.

“Civic integrity policy.” *Twitter Help Center*. Date of access: May 9, 2022. https://help.twitter.com/en/rules-and-policies/election-integrity-policy.

Clarke, Steve. 2002. “Conspiracy theories and conspiracy theorizing.” Philosophy of the Social Sciences 32 (2): 131-150. https://doi.org/10.1177/004931032002001.

Coady, David. 2003. “Conspiracy theories and official stories.” International Journal of Applied Philosophy 17 (2): 197-209. https://doi.org/10.5840/ijap200317210.

Coady, David. 2006. *Conspiracy Theories: The Philosophical Debate*, ed. David Coady. London, England: Routledge.

Coady, David. 2006. “The pragmatic rejection of conspiracy theories.” In *Conspiracy Theories: The Philosophical Debate*, ed. David Coady. 167-170. London, England: Routledge.

Coady, David. 2018. “Anti-rumor campaigns and conspiracy-baiting as propaganda.” In *Taking Conspiracy Theories Seriously*, ed. M R. X. Dentith. Collective Studies in Knowledge and Society, eds. James H. Collier, 171-186. Lanham, MD: Rowman & Littlefield International. Kindle.

Coady, David. 2018. “Conspiracy theory particularism, both epistemic and moral, versus generalism.” In *Taking Conspiracy Theories Seriously*, ed. M R. X. Dentith. Collective Studies in Knowledge and Society, eds. James H. Collier, 39-58. Lanham, MD: Rowman & Littlefield International. Kindle.

Coleman, Aaron Ross. August 11, 2020. “Tuesday’s Georgia runoff could help launch the Congressional QAnon Caucus.” *Vox*. Date of access: October 3, 2022. https://www.vox.com/2020/8/11/21362725/georgia-runoff-trump-congressional-qanon-caucus-marjorie-taylor-greene.

Collins, Ben and Brandy Zadrozny. July 21, 2020. “Twitter bans 7,000 QAnon accounts, limits 150,000 others as part of broad crackdown.” *NBC News*. Date of access: April 29, 2022. https://www.nbcnews.com/tech/tech-news/twitter-bans-7-000-qanon-accounts-limits-150-000-others-n1234541.

“Complete Report of the Chairman of the Committee on Public Information: 1917, 1918, 1919.” 1920. Washington D.C.: Government Printing Office. https://archive.org/details/CompleteReportCommitteeOnPublicInformation191719181919/mode/2up.

“Coordinated harmful activity.” January 2021. *Twitter Help Center*. Date of access: May 19, 2022. https://help.twitter.com/en/rules-and-policies/coordinated-harmful-activity.

“COVID-19 misleading information policy.” *Twitter Help Center*. Date of access: May 9, 2022. https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy.

Dang, Sheila. “Twitter estimates spam, fake accounts comprise less than 5% of users -filing.” *Reuters*. Date of access: November 18, 2022. https://www.reuters.com/technology/twitter-estimates-spam-fake-accounts-represent-less-than-5-users-filing-2022-05-02/.

Darwin, Hannah, Nick Neave, and Joni Holmes. 2011. “Belief in conspiracy theories: the role of paranormal belief, paranoid ideation, and schizotypy.” Personality and Individual Differences 50 (8): 1289-1293. https://doi.org/10.1016/j.paid.2011.02.027.

Dean, Jodi. 1998. *Aliens in America: Conspiracy Cultures from Outerspace to Cyberspace*. Ithaca, NY: Cornell University Press.

Dean, Jodi. 2000. "Declarations of independence." In *Cultural Studies & Political Theory*, ed. Jodi Dean, 285-304. Ithaca, New York: Cornell University Press.

deHaven-Smith, Lance. 2013. *Conspiracy Theory in America*. Austin, Texas: University of Texas Press.

DelReal, Jose A. September 16, 2016. "Here are 10 more conspiracy theories embraced by Donald Trump." *The Washington Post*. Date of access: October 5, 2022. https://www.washingtonpost.com/news/post-politics/wp/2016/09/16/here-are-10-more-conspiracy-theories-embraced-by-donald-trump/.

Dentith, M R. X. 2014. *The Philosophy of Conspiracy Theories*. New York, NY: Palgrave Macmillan.

Dentith, M R. X. 2016. "When inferring to a conspiracy might be the best explanation." Social Epistemology 30 (5-6): 572-591. https://doi.org/10.1080/02691728.2016.1172362.

Dentith, M R. X. 2018. *Taking Conspiracy Theories Seriously*, ed. M R. X. Dentith. Collective Studies in Knowledge and Society, eds. James H. Collier. Lanham, MD: Rowman & Littlefield International. Kindle.

Dentith, M R. X. 2019. "Conspiracy theories on the basis of evidence." Synthese 196: 2243-2261. https://doi.org/10.1007/s11229-017-1532-7.

Devine, Miranda. November 28, 2021. "Media helped hide the real Joe Biden by censoring Hunter stories: Devine." *New York Post*. Date of access: May 4, 2022. https://nypost.com/2021/11/28/media-helped-hide-the-real-joe-biden-by-censoring-hunter-stories-devine/.

DiGrazia, Joseph. 2017. "The social determinants of conspiratorial ideation." Socius: Sociological Research for a Dynamic World 3:1-9. https://doi.org/10.1177/2378023116689791.

Douglas, Karen M. 2021. "Are conspiracy theories harmless?" The Spanish Journal of Psychology 23, e13: 1-7. https://doi.org/10.1017/SJP.2021.10.

Duster, Chandelis. April 15, 2021. "Defense Department confirms leaked video of unidentified aerial phenomena is real." *CNN*. Date of access: January 29, 2022. https://www.cnn.com/2021/04/15/politics/unidentified-aerial-phenomena-defense-department/index.html.

Ecarma, Caleb. August 11, 2020. "Is social media's QAnon crackdown too late?" *Vanity Fair*. Date of access: October 3, 2022. https://www.vanityfair.com/news/2020/08/is-social-media-facebook-qanon-crackdown-too-late.

Economist. September 4, 2021. "From Congo to the Capitol, conspiracy theories are surging." Date of access: October 1, 2022. https://www.economist.com/international/2021/09/04/from-congo-to-the-capitol-conspiracy-theories-are-surging.

Editor's note. May 17, 2021. *PolitiFact*. https://www.politifact.com/li-meng-yan-fact-check/.

Edsall, Thomas B. February 3, 2021. "The QAnon delusion has not loosened its grip." *The New York Times*. Date of access: April 13, 2022. https://www.nytimes.com/2021/02/03/opinion/qanon-conspiracy-theories.html.

Ekeberg, Bjorn. April 30, 2019. "Cosmology has some big problems." *Scientific American*. Date of access: February 9, 2022. https://blogs.scientificamerican.com/observations/cosmology-has-some-big-problems/.

Enders, Adam M., Steven M. Smallpage, and Robert N. Lupton. 2018. "Are all 'Birthers' conspiracy theorists? On the relationship between conspiratorial thinking and political orientations." British Journal of Political Science 50 (3): 849-866. https://doi.org/10.1017/S0007123417000837.

"Ex post facto." *Legal Information Institute*. https://www.law.cornell.edu/wex/ex_post_facto.

Facebook Help Center. "COVID-19 and vaccine policy updates & protections." Date of access: May 24, 2022. https://www.facebook.com/help/230764881494641.

FBI Phoenix Field Office. 2019. "Anti-government, identity based, and fringe political conspiracy theories very likely motivate some domestic extremists to commit criminal, sometimes violent activity." *Federal Bureau of Investigation: Intelligence Bulletin*. View the document for free at https://news.yahoo.com/fbi-documents-conspiracy-theories-terrorism-160000507.html.

Fenster, Mark. 2008. *Conspiracy Theories: Secrecy and Power in American Culture*. Minneapolis, MN: University of Minnesota Press.

Ferreira, Fernanda. August 26, 2021. "The development of mRNA vaccines was a collaborative effort; Robert Malone contributed to their development, but he is not their inventor." *Health Feedback*. Date of access: May 10, 2022. https://healthfeedback.org/claimreview/the-development-of-mrna-vaccines-was-a-collaborative-effort-robert-malone-contributed-to-their-development-but-he-is-not-their-inventor/.

Fiske, John. 1999. "Blackstream knowledge: genocide." In *Media Matters: Race and Gender in US Politics*, Revised Edition, 191-216. Minneapolis, MN: University of Minnesota Press.

Funke, Daniel. September 16, 2020. *PolitiFact*. https://www.politifact.com/li-meng-yan-fact-check/.

Garcia, Feliks. October 6, 2016. “US government spent over $500m on fake Al-Qaeda propaganda videos that tracked location of viewers.” *Independent*. Date of access: May 12, 2022. https://www.independent.co.uk/news/world/us-government-pentagon-fake-al-qaeda-propganda-videos-a7348371.html.

Gilbert, David. 1996. “Tracking the real genocide: AIDS: conspiracy or unnatural disaster?” Covert Action Quarterly 58 (Fall): 55-65. http://lemegalodon.net/a3648-tracking-the-real-genocide-aids-conspiracy.html.

Grimes, David Robert. 2016. “On the viability of conspiratorial beliefs.” PLOS ONE 11 (3): 1-17. https://doi.org/10.1371/journal.pone.0151003.

Goertzel, Ted. 1994. “Belief in conspiracy theories.” Political Psychology 15 (4): 731-742. https://doi.org/10.2307/3791630.

Goffman, Erving. 1963. “Stigma and social identity.” *Stigma: Notes on the Management of Spoiled Identity*, 1-40. Englewood Cliffs, NJ: Pretince-Hall Inc.

Goshorn, Keith. 2000. “Strategies of deterrence and frames of containment: on critical paranoia and anti-conspiracy discourse.” Theory & Event 4 (3). https://muse.jhu.edu/article/32597

Groh, Deiter. 1987. “The temptation of conspiracy theory, or: Why do bad things happen to good people? Part I: Preliminary draft of a theory of conspiracy theories.” In *Changing Conceptions of Conspiracy,* eds. Carl F. Graumann and Serge Moscovici, Spring Series in Social Psychology, 1-13. New York, NY: Springer-Verlag.

Habedank, Dirk, Antonia Lagast, Monica Novoa-Usme, and Iskandar Atmowihardjo. 2022. “A case of myocarditis in a 60-year-old man 48 h after mRNA vaccination against SARS-CoV2.” Clinical Research in Cardiology 111 (2): 230-232. https://doi.org/10.1007/s00392-021-01946-4.

Hagen, Kurtis. 2010. “Is infiltration of ‘extremist groups’ justified?” International Journal of Applied Philosophy 24 (2): 153-168. https://doi.org/10.5840/ijap201024215.

Hagen, Kurtis. 2011. "Conspiracy theories and stylizing facts.” Journal for Peace and Justice Studies 21 (2): 3-22. https://doi.org/10.5840/peacejustice20112121.

Hagen, Kurtis. 2018. “Conspiracy theorists and monological belief systems.” Argumenta 3 (2): 303-326. https://www.argumenta.org/wp-content/uploads/2018/05/7-Argumenta-Kurtis-Hagen-Conspiracy-Theorists-and-Monological-Beliefs-Systems-BIS.pdf. doi: 10.23811/57.arg2017.hag.

Hagen, Kurtis. 2020. “Is conspiracy theorizing *really* epistemically problematic?” Episteme 2020: 1-23. https://doi.org/10.1017/epi.2020.19.

Hagen, Kurtis. 2022. “Are ‘Conspiracy Theories’ so unlikely to be true? A critique of Quassim Cassam’s concept of ‘Conspiracy Theories.’” Social Epistemology Ahead-of-Print: 1-15. https://doi.org/10.1080/02691728.2021.2009930.

Hall, Ellie. August 2, 2018. “What is QAnon? Here's what you need to know about the baseless mega-conspiracy theory.” *Buzzfeed News*. Date of access: October 2, 2022. https://www.buzzfeednews.com/article/ellievhall/qanon-trump-rally-conspiracy-theory.

Hall, Jeffrey A. 2006. “Aligning darkness with conspiracy theory: The discursive effect of African American interest in Gary Webb’s ‘Dark Alliance.’” The Howard Journal of Communication, 17 (3): 205-22. https://doi.org/10.1080/10646170600829626.

Harambam, Jaron and Stef Aupers. 2015. “Contesting epistemic authority: Conspiracy theories on the boundaries of science.” Public Understanding of Science 24 (4): 466-480. https://doi.org/10.1177/0963662514559891.

Hardin, Russell. 2002. “The crippled epistemology of extremism.” Political Extremism and Rationality. Ed. Albert Breton, Gianluigi Galeotti, Pierre Salmon, and Ronald Wintrobe. Cambridge, UK: Cambridge University Press.

Harris, Keith. 2019. “What’s epistemically wrong with Conspiracy theorising?” Royal Institute of Philosophy Supplements Vol. 84: 235-257. https://doi.org/10.1017/S1358246118000619.

Hart, Robert. June 2, 2021. “Parler’s popularity plummets as data reveals little appetite for returning ‘free speech’ app favored by conservatives.” *Forbes*. Date of access: May 26, 2022. https://www.forbes.com/sites/roberthart/2021/06/02/parlers-popularity-plummets-as-data-reveals-little-appetite-for-returning-free-speech-app-favored-by-conservatives/?sh=28de40545e13.

Hedges, Chris. 2010. *Death of the Liberal Class*. New York, NY: Bold Type Books, Inc. Ch.3.

Hibbet, Maia and Ryan Grim. January 12, 2022. “House republicans release text of redacted Fauci emails on COVID origins.” *The Intercept*. Date of access: May 3, 2022. https://theintercept.com/2022/01/12/covid-origins-fauci-redacted-emails/.

Hlavac, Marek. 2022. stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.3. https://CRAN.R-project.org/package=stargazer.

Hobbs, William R. And Margaret E. Roberts. 2018. “How sudden censorship can increase access to information.” American Political Science Review 112 (3): 621-636. https://doi.org/10.1017/S0003055418000084.

Hofstadter, Richard. 1996. “The paranoid style in American politics.” In *The Paranoid Style in American Politics and Other Essays*, 3-40. Cambridge, MA: Harvard University Press.

House Select Committee on Assassinations. *Final Report of the Select Committee on Assassinations: Summary of Findings and Recommendations*. Section I., pp. 3. https://history-matters.com/archive/contents/hsca/contents_hsca_report.htm.

“How fact-checking works.” May 9, 2022. *Meta Transparency Center*. Date of access: May 9, 2022. https://transparency.fb.com/en-gb/features/how-fact-checking-works/.

Howell, Thomas. 1997. “The Writers' War Board: US domestic propaganda in World War II.” The Historian 59 (4): 795-813. https://www.jstor.org/stable/24451817?seq=1.

Husting, Ginna and Martin Orr. 2007. “Dangerous machinery: ‘Conspiracy theorist’ as a transpersonal strategy of exclusion.” Symbolic Interaction 30 (2): 127–150. https://doi.org/10.1525/si.2007.30.2.127.

Imhoff, Roland and Pia Karoline Lamberty. 2017. “Too special to be duped: Need for uniqueness motivates conspiracy beliefs.” European Journal of Social Psychology 47 (6): 724-734. https://doi.org/10.1002/ejsp.2265.

Istampoulouoglou, Ionna, Georgios Dimitriou, Selina Spani, Andreas Christ, Barbara Zimmermanns, Sarah Koechlin, Oliver Stoeckmann, Clemens Winterhalder, David Marono, Valeriu Toma, and Anne B. Leuppi-Taegtmeyer. 2021. “Myocarditis and pericarditis in association with COVID-19 mRNA-vaccination: Cases from a regional pharmacovigilance centre.” Global Cardiology Science & Practice 3: 1-16. https://doi.org/10.21542/gcsp.2021.18.

Jameson, Frederic. “Cognitive Mapping.” In *Marxism and the Interpretation of Culture*, ed. Cary Nelson and Lawrence Grossberg, 347-360. Urbana, Illinois: University of Illinois Press.

Jin, Kang-Xing. 2020. “Keeping people safe and informed about the Coronavirus.” Date of access: February 28, 2022. https://about.fb.com/news/2020/12/coronavirus/.

Jolley, Daniel and Jenny L. Paterson. 2020. “Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence.” British Journal of Social Psychology 59 (3): 628-640. https://doi.org/10.1111/bjso.12394.

Jolley, Daniel and Karen M. Douglas. 2014. “The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one’s carbon footprint.” British Journal of Psychology 105 (1): 35–56. https://doi.org/10.1111/bjop.12018.

Jolley, Daniel and Karen M. Douglas. 2017. “Prevention is better than cure: Addressing anti-vaccine conspiracy theories.” Journal of Applied Social Psychology 47 (8): 459-469. https://doi.org/10.1111/jasp.12453.

Jones, Vincent. 1985. *Manhattan: The Army and the Atomic Bomb*. Washington, D.C.: United States Army Center of Military History. Date of access: February 12, 2022. https://history.army.mil/html/books/011/11-10/CMH_Pub_11-10.pdf.

Kaufmann, Eric. 2021. “Academic freedom in crisis: Punishment, political discrimination, and self-censorship.” Center for the Study of Partisanship and Ideology, Report No. 2. Date of access: February 2, 2022. https://cspicenter.org/wp-content/uploads/2021/03/AcademicFreedom.pdf.

Keeley, Brian L. 1999. “Of conspiracy theories.” The Journal of Philosophy 96 (3): 109-126. https://doi.org/10.2307/2564659.

Keeley, Brian L. 2003. “Nobody expects the Spanish Inquisition! More thoughts on conspiracy Theory.” Journal of Social Philosophy 34 (1): 104-110. https://doi.org/10.1111/1467-9833.00168.

Klein, Oliver, Nicolas Van der Linden, Myrto Pantazi, and Mikhail Kissine. 2015. “Behind the screen conspirators: Paranoid social cognition in an online age.” In *The Psychology of Conspiracy: A Festschrift for Mirosław Kofta*, ed. Michal Bilewicz, Aleksandra Cichoka, and Wiktor Soral, 162-182. Hove, UK: Routledge.

Kwon, K. Hazel, Shin-Il Moon, and Michael A Stefanone. 2015. “Unspeaking on Facebook? Testing network effects on self-censorship of political expressions in social network sites.” Quality & Quantity 49: 1417–1435. https://link.springer.com/article/10.1007/s11135-014-0078-8.

LaFraniere, Sharon. May 5, 2022. “Citing safety concerns, F.D.A. further limits J.&J. COVID vaccine.” *The New York Times*. Date of access: May 9, 2022. https://www.nytimes.com/2022/05/05/us/politics/johnson-vaccine-fda-covid.html?searchResultPosition=1.

Lakatos, Imre. 1970. “The history of science and its rational reconstructions.” *PSA: Proceedings of the biennial meeting of the philosophy of science association*. D. Reidel Publishing. https://www.journals.uchicago.edu/doi/abs/10.1086/psaprocbienmeetp.1970.495757.

Lamberty, Pia and David Leiser. 2019. “Sometimes you just have to go in: The link between conspiracy beliefs and political action.” Preprint from PsycArXiv. Preprint doi: 10.31234/osf.io/bdrxc.

Lantian, Anthony, Dominique Muller, Cécile Nurra, Olivier Klein, Sophie Berjot, and Myrto Pantazi. 2018. “Stigmatized beliefs: Conspiracy theories, anticipated negative evaluation of the self, and fear of social exclusion.” European Journal of Social Psychology 48 (7): 939-954. https://doi.org/10.1002/ejsp.2498.

Lantian, Anthony, Virginie Bagneux, Sylvain Delouvée, and Nicolas Gauvrit. 2021. “Maybe a free thinker but not a critical one: High conspiracy belief is associated with low critical thinking ability.” Applied Cognitive Psychology 35, (3): 674-684. https://doi.org/10.1002/acp.3790.

“Legacy verification policy.” *Twitter Help Center*. [Policy covers the time period through November 9, 2022, including all tweets in my sample]. Date of access: November 18, 2022. https://help.twitter.com/en/managing-your-account/legacy-verification-policy.

Levenson, Eric. August 13, 2019. “Who's who of Jeffrey Epstein's powerful friends, associates and possible co-conspirators.” *CNN*. Date of access: October 4, 2022. https://www.cnn.com/2019/08/12/us/jeffrey-epstein-associates-possible-accomplices/index.html.

Levy, Neil. 2007. “Radically socialized knowledge and conspiracy theories.” Episteme 4 (2): 181-192. https://doi.org/10.3366/epi.2007.4.2.181.

Li, Ang, Dongdong Jiao, Xiaoqian Liu, and Tingshao Zhu. 2020. “A comparison of the psycholinguistic styles of schizophrenia-related stigma and depression-related stigma on social media: Content analysis.” Journal of Medical Internet Research 22 (4): 1-10. http://dx.doi.org/10.2196/16470.

Lim, Yongwhan, Min Chul Kim, Kye Hun Kim, In-Seok Jeong, Yong Soo Cho, Yoo Duk Choi, and Jong Eun Lee. 2021. “Case report: Acute fulminant myocarditis and cardiogenic shock after messenger RNA coronavirus disease 2019 vaccination requiring extracorporeal cardiopulmonary resuscitation.” Frontiers in Cardiovascular Medicine 8 (758996): 1-6. https://doi.org/10.3389/fcvm.2021.758996.

Lima, Cristiano. May 27, 2021. “Facebook no longer treating 'man-made' COVID as a crackpot idea.” *Politico*. Date of access: May 4, 2022. https://www.politico.com/news/2021/05/26/facebook-ban-covid-man-made-491053.

Luhmann, Niklas. 1993. *Risk: A Sociological Theory*. Translated by Rhodes Barrett. Berlin, Germany: Walter de Gruyter.

MacIntyre, Alasdair. 2007. *After Virtue: A Study in Moral Theory*. London: Duckworth.

Mandavilli, Apoorva. Updated: November 11, 2021. Original publication date: August 18, 2021. “Vaccine effectiveness against infection may wane, C.D.C. studies find.” *New York Times*. Date of access: May 9, 2022. https://www.nytimes.com/2021/08/18/health/covid-cdc-boosters-elderly.html?searchResultPosition=2.

Mandik, Pete. 2007. “Shit happens.” Episteme 4 (2): 205-218. https://doi.org/10.3366/epi.2007.4.2.205.

Manskar, Noah. October 14, 2020. “Twitter, Facebook censor *Post* over Hunter Biden exposé.” *New York Post*. Date of access: May 4, 2022. https://nypost.com/2020/10/14/facebook-twitter-block-the-post-from-posting/.

Manskar, Noah. March 25, 2021. “Jack Dorsey says blocking *Post*’s Hunter Biden story was ‘total mistake’—but won’t say who made it.” *New York Post*. Date of access: May 16, 2022. https://nypost.com/2021/03/25/dorsey-says-blocking-posts-hunter-biden-story-was-total-mistake/.

Matta, Abhishek, Rajesh Kunadharaju, Marcus Osman, Christy Jesme, Zachary McMiller, Erika M. Johnson, Danielle Matta, Rekha Kallamadi, and Dinesh Bande. 2021. “Clinical presentation and outcomes of myocarditis post mRNA vaccination: A meta-analysis and systematic review.” Cureus 13 (11): 1-11. 10.7759/cureus.19240.

May, Ashley. August 1, 2018. “QAnon: The conspiracy theory explained, after Q posters spotted at Trump's Florida rally.” *USA Today*. Date of access: October 2, 2022. https://www.usatoday.com/story/news/nation-now/2018/08/01/qanon-trump-florida-rally-what-q-signs-mean/878642002/.

Media Lens. October 17, 2018. “Phillip Cross: The mystery Wikipedia editor targeting anti-war sites.” *Mintpress News*. Date of access: February 12, 2022. https://www.mintpressnews.com/phillip-cross-the-mystery-wikipedia-editor-targeting-anti-war-sites/250824/.

Melley, Timothy. 2016. *Empire of Conspiracy: The Culture of Paranoia in Postwar America*. Ithaca, NY: Cornell University Press.

Menczer, Filippo. November 28, 2016. “Fake online news spreads through social echo chambers.” *Scientific American*. Date of access: October 4, 2022. https://www.scientificamerican.com/article/fake-online-news-spreads-through-social-echo-chambers/.

Messerole, Chris. May 9, 2018. “How misinformation spreads on social media—And what to do about it.” *Brookings*. Date of access: October 4, 2022. https://www.brookings.edu/blog/order-from-chaos/2018/05/09/how-misinformation-spreads-on-social-media-and-what-to-do-about-it/.

Metaxas, Panagiotis, Eni Mustafaraj, Kily Wong, Laura Zeng, Megan O’Keefe, and Samantha Finn. 2015. “What do retweets indicate? Results from user survey and meta-review of research.” Ninth International AAAI Conference on Web and Social Media 9 (1): 659-661. https://doi.org/10.1609/icwsm.v9i1.14661.

Mihalcik, Carrie. August 11, 2020. “QAnon groups and pages flourish on Facebook, report says.” *CNET*. Date of access: October 3, 2022. https://www.cnet.com/tech/mobile/qanon-groups-and-pages-flourish-on-facebook-report-says/.

Miller, Joanne M., Kyle L. Saunders, and Christina E. Farhart. 2016. “Conspiracy endorsement as motivated reasoning: The moderating roles of political Knowledge and trust.” American Journal of Political Science, 60 (4): 824-844. https://www.jstor.org/stable/24877458.

“Misinformation.” *Meta Transparency Center*. Date of access: May 9, 2022. https://transparency.fb.com/policies/community-standards/misinformation/.

“Misinformation policies.” *YouTube Help*. Date of access: May 9, 2022. https://support.google.com/youtube/answer/10834785?hl=en&ref_topic=10833358#zippy=%2Charmful-remedies-and-cures%2Csuppression-of-census-participation%2Cmanipulated-content%2Cmisattributed-content.

Morris, Alex. October 15, 2021. “It’s not Q. It’s you.” *Rolling Stone*. Date of access: April 13, 2022. https://www.rollingstone.com/culture/culture-features/qanon-expert-joesph-uscinski-1242636/.

Morris, Emma-Jo and Gabrielle Fonrogue. October 14, 2020. “Smoking-gun email reveals how Hunter Biden introduced Ukrainian businessman to VP dad.” *New York Post*. Date of access: May 4, 2022. https://nypost.com/2020/10/14/email-reveals-how-hunter-biden-introduced-ukrainian-biz-man-to-dad/.

Morrison, Pat. August 14, 2019. “Column: How much are conspiracy theories really shaping American life?” Date of access: April 13, 2022. https://www.latimes.com/opinion/story/2019-08-13/patt-morrison-joseph-uscinski-white-supremacy-conspiracy.

Myers, Steven Lee. October 13, 2022. “How social media amplifies misinformation more than information.” *The New York Times*. Date of access: October 13, 2022. https://www.nytimes.com/2022/10/13/technology/misinformation-integrity-institute-report.html.

Nabi, Zubair. 2014. “Resistance [strikethrough] Censorship is Futile.” First Monday 19 (11). https://arxiv.org/abs/1411.0225. doi: 10.5210/fm.v19i11.

National Education. December 17, 2015. “Signing of the convention on media and information literacy.” [Translated from French using Google Translate] Date of access: January 31, 2022. https://www.najat-vallaud-belkacem.com/2015/12/17/signature-de-la-convention-sur-leducation-aux-medias-et-a-linformation/.

North, David. August 25, 2017. “An open letter to Google: Stop the censorship of the Internet! Stop the political blacklisting of the World Socialist Web Site!” *World Socialist Web Site*. Date of access: May 10, 2022. https://www.wsws.org/en/articles/2017/08/25/pers-a25.html.

Olmsted, Kathryn S. 2009. *Real Enemies: Conspiracy Theories and American Democracy, World War I to 9/11*. New York, NY: Oxford University Press.

Orosz, Gábor, Péter Krekó, Benedek Paskuj, István Tóth-Király, Beáta Bőthe and Christine Roland-Lévy. 2016. “Changing conspiracy beliefs through rationality and ridiculing.” Frontiers in Psychology 7 (1525): 1-9. https://doi.org/10.3389/fpsyg.2016.01525.

Orr, Martin and Ginna Husting. 2019. “Media marginalization of racial minorities: ‘Conspiracy theorists’ in US ghettos and on the ‘Arab street.’” In *Conspiracy Theories and the People Who Believe Them*, ed. Joseph E. Uscinski, 82-93. New York, NY: Oxford University Press.

Orr, Martin and M R. X. Dentith. 2018. “Clearing up some conceptual confusions about conspiracy theory theorizing.” In *Taking Conspiracy Theories Seriously, ed. M R. X. Dentith*. Collective Studies in Knowledge and Society, ed. James H. Collier, 141-154. Lanham, MD: Rowman & Littlefield International. Kindle.

Paul, Kari. October 14, 2020. “Facebook and Twitter restrict controversial *New York Post* story on Joe Biden.” *The Guardian*. Date of access: May 16, 2022. https://www.theguardian.com/technology/2020/oct/14/facebook-twitter-new-york-post-hunter-biden.

Pennycook, Gordon, James Allan Cheyne, Nathaniel Barr, Derek J. Koehler, Jonathan A. Fugelsang. 2015. “On the reception and detection of pseudo-profound bullshit.”

Judgement and Decision Making 10 (6): 549-563. http://journal.sjdm.org/15/15923a/jdm15923a.html.

"Pierre Kory, MD." St. George's University, Grenada, West Indies website. https://web.archive.org/web/20210720021147/https://www.sgu.edu/graduate-profiles/pierre-kory-md/. Rotted url: https://www.sgu.edu/graduate-profiles/pierre-kory-md/.

Pigden, Charles. 1995. "Popper revisited, or What is wrong with conspiracy theories." Philosophy of the Social Sciences 25 (1): 3-34. https://doi.org/10.1177/004839319502500101.

Pigden, Charles. 2006. "Complots of mischief." In *Conspiracy Theories: The Philosophical Debate,* ed. David Coady, 139-166. London, England: Routledge.

Pigden, Charles. 2018. "Conspiracy theories, deplorables, and defectibility: A reply to Patrick Stokes." In *Taking Conspiracy Theories Seriously*, ed. M R. X. Dentith. Collective Studies in Knowledge and Society, ed. James H. Collier, 203-216. Lanham, MD: Rowman & Littlefield International. Kindle.

Pipes, Daniel. 1997. *Conspiracy: How the Paranoid Style Flourishes and Where It Comes From*. New York, NY: Simon & Schuster.

Plesch, Daniel. 2017. *Human Rights After Hitler: The Lost History of Prosecuting Axis War Crimes*. Washington, DC: Georgetown University Press.

Popper, Karl. 1994. *The Open Society and its Enemies: New One-Volume Edition*. Princeton, NJ: Princeton University Press.

Post Editorial Board. October 12, 2021. "One year later, *The Post*'s Hunter Biden reporting is vindicated—but still buried." *New York Post*. Date of access: May 14, 2022. https://nypost.com/2021/10/12/one-year-later-the-posts-hunter-biden-reporting-is-vindicated-but-still-buried/.

Powers, Elia, Michael Koliska, and Pallavi Guha. 2019. "'Shouting matches and echo chambers:' Perceived identity threats and political self-censorship on social media." International Journal of Communication 13: 3630-3649. Date of access: December 29, 2021. https://ijoc.org/index.php/ijoc/article/view/9927/2743.

Raikka, Juha and Lee Basham. 2019. "Conspiracy theory phobia." In *Conspiracy Theories and the People Who Believe Them*, ed. Joseph E. Uscinski, 178-186. New York, NY: Oxford University Press.

Ray, Siladitya. January 14, 2021. "The far-right is flocking to these alternate social media apps—Not all of them are thrilled." *Forbes*. Date of access: May 7, 2022. https://www.forbes.com/sites/siladityaray/2021/01/14/the-far-right-is-flocking-to-these-alternate-social-media-apps---not-all-of-them-are-thrilled/?sh=107b747e55a4.

Reuters. August 11, 2020. “Facebook removes seven million posts for sharing false information on coronavirus.” *NBC News*. Date of access: May 1, 2022. https://www.nbcnews.com/tech/tech-news/facebook-removes-seven-million-posts-sharing-false-information-coronav-rcna77.

Risen, James. January 3, 2018. “The biggest secret: My life as a *New York Times* reporter in the shadow of the War on Terror.” *The Intercept*. Date of access: January 10, 2022. https://theintercept.com/2018/01/03/my-life-as-a-new-york-times-reporter-in-the-shadow-of-the-war-on-terror/.

Robbins, Rebecca and Christina Jewett. Updated: December 17, 2021. Original publication date: December 16, 2021. “The C.D.C. recommends that other vaccines be preferred over J.&J.” *The New York Times*. Date of access: May 9, 2022. https://www.nytimes.com/2021/12/17/health/cdc-johnson-vaccine-blood-clot.html?searchResultPosition=4.

Robins, Robert S. and Jerrold M. Post. 1997. *Political Paranoia: The Psychopolitics of Hatred*. New Haven, CT: Yale University Press.

Robinson, Patrick, Daniel Turk, Sagar Jilka, and Matteo Cella. 2019. “Measuring attitudes towards mental health using social media: investigating stigma and trivialisation.” Social Psychiatry and Psychiatric Epidemiology 54: 51-58. https://doi.org/10.1007/s00127-018-1571-5.

Rodrigo, Chris Mills. August 10, 2020. “Internal Facebook investigation found millions of users involved with QAnon theory: report.” *The Hill*. Date of access: October 3, 2022. https://thehill.com/policy/technology/511377-internal-facebook-investigation-found-millions-of-users-supporting-qanon/.

Rosen, Guy. October 21, 2019. “Helping to protect the 2020 US elections.” *Meta*. Date of access: May 16, 2022. https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/.

Rosen, Guy. Updated: May 26, 2021. Original publication date: April 16, 2020. “An update on our work to keep people informed and limit misinformation about COVID-19.” *Meta*. Date of access: May 10, 2022. https://about.fb.com/news/2020/04/covid-19-misinfo-update/.

Rothschild, Mike. January 27, 2021. “Here is every QAnon prediction that’s failed to come true.” *The Daily Dot*. Date of access: February 27, 2022. https://www.dailydot.com/debug/qanon-failed-predictions/.

Rottweiler, Bettina and Paul Gill. 2020. “Conspiracy beliefs and violent extremist intentions: The contingent effects of self-efficacy, self-control and law-related morality.” Terrorism and Political Violence 24, (1): 1-20. https://doi.org/10.1080/09546553.2020.1803288.

Sargent, Kayla. October 20, 2020. "*New York Post* Twitter lock reaches seventeenth day." *mrcNewsBusters*. Date of access: May 16, 2022. https://www.newsbusters.org/blogs/free-speech/kayla-sargent/2020/10/30/new-york-post-twitter-lock-reaches-seventeenth-day.

Sayers, Freddie. May 31, 2021. "How Facebook censored the lab leak theory." *UnHerd*. Date of access: May 10, 2022. https://unherd.com/2021/05/how-facebook-censored-the-lab-leak-theory/.

Schaefer, Donovan. July 6, 2022. "Buying into conspiracy theories can be exciting—That's what makes them dangerous." *Salon*. Date of access: October 1, 2022. https://www.salon.com/2022/07/06/buying-into-conspiracy-theories-can-be-exciting--thats-what-makes-them_partner/.

Schott, Bryan. January 14, 2022. "Dave Bateman doubles down on antisemitic conspiracy theory." *The Salt Lake Tribune*. Date of access: January 31, 2022. https://www.sltrib.com/news/politics/2022/01/14/dave-bateman-doubles-down/.

Sen, Ari and Brandy Zadrozny. August 10, 2020. "QAnon groups have millions of members on Facebook, documents show." *NBC News*. Date of access: October 3, 2022. https://www.nbcnews.com/tech/tech-news/qanon-groups-have-millions-members-facebook-documents-show-n1236317.

Smallpage, Steven M., Hugo Drochon, Joseph E. Uscinski, and Casey Klofstad. 2020. "Who are the conspiracy theorists? Demographics and conspiracy theories." In *Routledge Handbook of Conspiracy Theories*, ed. Michael Butter and Peter Knight, from the series Conspiracy Theories, 263-277. Abingdon, England: Routledge.

Solnit, Rebecca. January 5, 2022. "Why Republicans keep falling for Trump's lies." *The New York Times*. Date of access: February 17, 2022. https://www.nytimes.com/2022/01/05/opinion/republicans-trump-lies.html?searchResultPosition=1.

Stanley-Becker, Isaac. August 1, 2018. "'We are Q': A deranged conspiracy cult leaps from the Internet to the crowd at Trump's 'MAGA' tour." *The Washington Post*. Date of access: October 2, 2022. https://www.washingtonpost.com/news/morning-mix/wp/2018/08/01/we-are-q-a-deranged-conspiracy-cult-leaps-from-the-internet-to-the-crowd-at-trumps-maga-tour/.

Stanton, Zack. June 17, 2020. "You're living in the Golden Age of conspiracy theories." *Politico*. Date of access: October 1, 2022. https://www.politico.com/news/magazine/2020/06/17/conspiracy-theories-pandemic-trump-2020-election-coronavirus-326530.

Stewart, Phil and Tom Perry. April 12, 2018. "US says air strikes cripple Syria chemical weapons program." *Reuters*. Date of access: May 1, 2022. https://www.reuters.com/article/us-mideast-crisis-syria/trump-says-ordered-precision-strikes-against-syria-chemical-weapons-capabilities-idUSKBN1HJ0ZS.

Stokes, Patrick. 2018. “On some moral costs of conspiracy theory.” In *Taking Conspiracy Theories Seriously*, ed. M R. X. Dentith. Collective Studies in Knowledge and Society, ed. James H. Collier, 189-201. Lanham, MD: Rowman & Littlefield International. Kindle.

Suciu, Peter. August 2, 2021. “Spotting misinformation on social media is increasingly challenging.” *Forbes*. Date of access: October 4, 2022. https://www.forbes.com/sites/petersuciu/2021/08/02/spotting-misinformation-on-social-media-is-increasingly-challenging/?sh=408bb4a2771c.

Sunstein, Cass R. and Adrian Vermeule. 2009. “Conspiracy theories: Causes and cures.” The Journal of Political Philosophy 17 (2): 202-227. https://doi.org/10.1111/j.1467-9760.2008.00325.x.

Swami, Viren, Martin Voracek, Stefan Stieger, Ulrich S. Tran, and Adrian Furnham. 2014. “Analytic thinking reduces belief in conspiracy theories.” Cognition, 133 (3): 572-585. https://doi.org/10.1016/j.cognition.2014.08.006.

Swami, Viren, Rebecca Coles, Stefan Stieger, Jakob Pietschnig, Adrian Furnham, Sherry Rehim, and Martin Voracek. 2011. “Conspiracist ideation in Britain and Austria: Evidence of a monological belief system and associations between individual psychological differences and real-world and fictitious conspiracy theories.” British Journal of Psychology 102 (3): 443-463. https://doi.org/10.1111/j.2044-8295.2010.02004.x.

“Synthetic and manipulated media policy.” *Twitter Help Center*. Date of access: May 9, 2022. https://help.twitter.com/en/rules-and-policies/manipulated-media.

Taibbi, Matt. June 19, 2021. “Meet the censored: Bret Weinstein.” *TK News by Matt Taibbi*. Date of access: May 10, 2022. https://taibbi.substack.com/p/meet-the-censored-bret-weinstein?s=r.

Taibbi, Matt. July 2, 2021. “If private platforms use government guidelines to police content, is that state censorship?” *TK News by Matt Taibbi*. Date of access: May 10, 2022. https://taibbi.substack.com/p/a-case-of-intellectual-capture-on?s=r.

TASS. April 13, 2018. “Russian top brass reports it has proof of UK’s involvement in Douma chemical incident.” *Russian News Agency TASS*. https://tass.com/defense/999641.

Thalmann, Katharina. 2019. *The Stigmatization of Conspiracy Theory Since the 1950s: “A Plot to Make Us Look Foolish,”* from the series Conspiracy Theories, ed. Michael Butter and Peter Knight. Abingdon, England: Routledge.

The Bryant Park Project. April 9, 2008. “Reminder: Saddam statue was toppled by Psy-Ops,” from special series The Impact of War. *NPR*. Date of access: May 12, 2022. https://www.npr.org/2008/04/09/89489923/reminder-saddam-statue-was-toppled-by-psy-ops.

Tin, Alexander. June 14, 2021. “CDC plans meeting on rare heart inflammation following COVID-19 vaccines.” *CBS News*. Date of access: May 10, 2022. https://www.cbsnews.com/news/covid-19-vaccine-cdc-meeting-myocarditis-heart-inflammation/.

Twitter. January 8, 2021. “Permanent suspension of @realDonaldTrump.” *Twitter Blog*. Date of access: May 10, 2022. https://blog.twitter.com/en_us/topics/company/2020/suspension.

Twitter Safety. January 12, 2021. “An update following the riots in Washington, DC.” *Twitter Blog*. Date of access: May 19, 2022. https://blog.twitter.com/en_us/topics/company/2021/protecting--the-conversation-following-the-riots-in-washington--.

US Department of State, Bureau of International Information Programs. “How to identify misinformation.” 2005. https://rense.com/27-595713.html.

US Senate. *Final Report of the Select Committee to Study Governmental Operations with Respect to Intelligence Activities.* Book III. April 23, 1976. Washington D.C.: US Government Printing Office. https://www.intelligence.senate.gov/sites/default/files/94755_III.pdf.

US Senate. *Final Report of the Select Committee to Study Governmental Operations with Respect to Intelligence Activities.* Book II. April 26, 1976. Washington D.C.: US Government Printing Office. https://www.intelligence.senate.gov/sites/default/files/94755_II.pdf.

US Senate. “Hearings before the Select Committee to Study Governmental Operations with Respect to Intelligence Activities: Ninety-Fourth Congress.” First Session, Vol. 7, Covert Action. December 4 and 5, 1975. Washington D.C.: US Government Printing Office. https://archive.org/details/ChurchCommittee/Church%20Committee%20Volume%207%20-%20Hearings%20on%20Covert%20Action/mode/2up.

Uscinski, Joseph E. and Ryden W. Butler. 2013. “The epistemology of fact checking.” Critical Review 25 (2): 162-180. https://doi.org/10.1080/08913811.2013.843872.

Uscinski, Joseph E. and Joseph M. Parent. 2014. *American Conspiracy Theories*. New York, NY: Oxford University Press.

Uscinski, Joseph E. 2015. “The epistemology of fact-checking (is still naïve): Rejoinder to Amazeen.” Critical Review 27 (2): 243-252. https://doi.org/10.1080/08913811.2015.1055892.

Uscinski, Joseph E. March 30, 2016. “If Trump’s rhetoric around conspiracy theories follows him to the White House, it could lead to the violation of rights on a massive scale.” *LSE Blogs*. Date of access: October 5, 2022. https://blogs.lse.ac.uk/usappblog/2016/03/30/if-trumps-rhetoric-around-conspiracy-theories-follows-him-to-the-white-house-it-could-lead-to-the-violation-of-rights-on-a-massive-scale/.

van Prooijen, Jan-Willem. 2016. “Why education predicts decreased belief in conspiracy theories.” Applied Cognitive Psychology 31 (1): 50-58. https://doi.org/10.1002/acp.3301.

van Prooijen, Jan-Willem, Talia Cohen Rodrigues, Carlotta Bunzel, Oana Georgescu, Dániel Komáromy, and André P. M. Krouwel. 2022. “Populist gullibility: Conspiracy theories, news credibility, bullshit receptivity, and paranormal belief.” Political Psychology, Forthcoming. https://doi.org/10.1002/acp.3301.

Varol, Onur, Emilio Ferrera, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. “Online human-bot interactions: Detection, estimation, and characterization.” Proceedings of the International AAAI Conference on Web and Social Media 11 (1): 280-289. https://doi.org/10.1609/icwsm.v11i1.14871.

Weinstein, Bret. June 2021. “How to save the world, in three easy steps: DarkHorse podcast with Robert Malone, Steve Kirsch & Bret Weinstein.” *DarkHorse Podcast.* https://open.spotify.com/episode/2GkYPouJqLMnMqDzMsdaeh.

Wendt, Alexander and Raymond Duvall. 2008. “Sovereignty and the UFO.” Political Theory 36 (4): 607-633. https://doi.org/10.1177/0090591708317902.

Wikipedia. “Bohemian Grove.” Date of access: February 28, 2022. https://en.wikipedia.org/wiki/Bohemian_Grove.

Wikipedia. “ExxonMobil climate change controversy.” Date of access: February 28, 2022. https://en.wikipedia.org/wiki/ExxonMobil_climate_change_controversy#Board_shakeup.

Wikipedia. “General Motors streetcar conspiracy.” Date of access: February 28, 2022. https://en.wikipedia.org/wiki/General_Motors_streetcar_conspiracy.

Wikipedia. “Jekyll Island.” Date of access: February 28, 2022. https://en.wikipedia.org/wiki/Jekyll_Island#Planning_of_the_Federal_Reserve_System.

Wikipedia. “Operation Northwoods.” Date of access: February 28, 2022. https://en.wikipedia.org/wiki/Operation_Northwoods.

Wikipedia. “Unethical human experiments in the United States.” Date of access: February 28, 2022. https://en.wikipedia.org/wiki/Unethical_human_experimentation_in_the_United_States.

Whitten, Sarah. November 16, 2015. “Rumors and misinformation circulate on social media following Paris attacks.” *CNBC*. Date of access: October 4, 2022. https://www.cnbc.com/2015/11/14/rumors-and-misinformation-circulate-on-social-media-following-paris-attacks.html.

Williams, Juan. October 14, 2019. “Juan Williams: Trump, the conspiracy theory president.” *The Hill*. Date of access: October 5, 2022. https://thehill.com/opinion/white-house/465603-juan-williams-trump-the-conspiracy-theory-president/.

Willingham, AJ. October 3, 2020. “How the pandemic and politics gave us a golden age of conspiracy theories.” *CNN*. Date of access: October 1, 2022. https://www.cnn.com/2020/10/03/us/conspiracy-theories-why-origins-pandemic-politics-trnd/index.html.

Winter, Jana. August 1, 2019. “Exclusive: FBI document warns conspiracy theories are a new domestic terrorism threat.” *Yahoo News*. Date of access: May 19, 2022. https://news.yahoo.com/fbi-documents-conspiracy-theories-terrorism-160000507.html.

Wong, Julia Carrie. July 30, 2018. “What is QAnon? Explaining the bizarre rightwing conspiracy theory.” *The Guardian*. Date of access: October 2, 2022. https://www.theguardian.com/technology/2018/jul/30/qanon-4chan-rightwing-conspiracy-theory-explained-trump.

Wong, Julia Carrie. August 11, 2020. “Revealed: QAnon Facebook groups are growing at a rapid pace around the world.” *The Guardian*. Date of access: October 3, 2022. https://www.theguardian.com/us-news/2020/aug/11/qanon-facebook-groups-growing-conspiracy-theory.

Wong, Queenie. August 18, 2021. “Facebook removed more than 20 million posts for COVID-19 misinformation.” *CNET*. Date of access: May 1, 2022. https://www.cnet.com/news/social-media/facebook-removed-more-than-20-million-posts-for-covid-19-misinformation/.

Wood, Michael J. 2015. “Some dare call it conspiracy: Labeling something a conspiracy theory does not reduce belief in it.” Political Psychology 37 (5): 695-705. https://doi.org/10.1111/pops.12285.

Wood, Michael J. And Karen M. Douglas. 2013. “‘What about building 7?’ A social psychological study of online discussion of 9/11 conspiracy theories.” Frontiers in Psychology 4 (409): 1-9. https://doi.org/10.3389/fpsyg.2013.00409.

Wood, Michael J. And Karen M. Douglas. 2015. “Online communication as a window to conspiracist worldviews.” Frontiers in Psychology 6 (836): 1-8. https://doi.org/10.3389/fpsyg.2013.00409.

Wood, Michael J. Karen M. Douglas, and Robbie M. Sutton. 2012. “Dead and alive: Beliefs in contradictory conspiracy theories.” Social Psychological and Personality Science 3 (6): 767-773. https://doi.org/10.1177/1948550611434786.

Woodruff, Judy. August 1, 2018. “How the false, fringe ‘QAnon’ conspiracy theory aims to protect Trump.” *PBS*. Date of access: October 2, 2022. https://www.pbs.org/newshour/show/how-the-false-fringe-qanon-conspiracy-theory-aims-to-protect-trump.

Woods, Dan. July 14, 2022. “I’m a former CIA cyber-operations officer who studies bot traffic. Here’s why it’s plausible that more than 80% of Twitter’s accounts are actually fake—

and Twitter is not alone." *F5*. Date of access: November 19, 2022. https://www.f5.com/company/blog/bot-traffic-percentage-fake-accounts-expert.

Yaraghi, Niam. April 9, 2019. "How should social media platforms combat misinformation and hate speech?" *Brookings*. Date of access: October 4, 2022. https://www.brookings.edu/blog/techtank/2019/04/09/how-should-social-media-platforms-combat-misinformation-and-hate-speech/.

Young, Virginia Alvino. May 27, 2020. "Nearly half of the Twitter accounts discussing 'reopening America' may be bots." *Carnegie Mellon University News*. Date of access: November 18, 2022. https://www.cmu.edu/news/stories/archives/2020/may/twitter-bot-campaign.html.

YouTube Team. January 25, 2019. "Continuing our work to improve recommendations on YouTube." *YouTube Official Blog*. Date of access: May 24, 2022. https://blog.youtube/news-and-events/continuing-our-work-to-improve/.

YouTube Team. October 15, 2020. "Managing harmful conspiracy theories on YouTube." *YouTube Official Blog*. Date of access: May 24, 2022. https://blog.youtube/news-and-events/harmful-conspiracy-theories-youtube/.

Zimmer, Carl. Updated: August 18, 2021. Original publication date: July 23, 2021. "Israeli data suggests possible waning in effectiveness of Pfizer vaccine." *The New York Times*. Date of access: May 9, 2022. https://www.nytimes.com/2021/07/23/science/covid-vaccine-israel-pfizer.html?searchResultPosition=1.

Zucchino, David. July 3, 2004. "Army stage-managed fall of Hussein statue." *Los Angeles Times*. Date of access: May 12, 2022. https://www.latimes.com/archives/la-xpm-2004-jul-03-na-statue3-story.html.