UNIVERSITY OF CALIFORNIA,
IRVINE

Conspiracy Theories and Resistance to Evidence

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Philosophy

by

Maria Giulia Napolitano

Dissertation Committee:
Professor Sven Bernecker, Chair
Distinguished Professor Margaret Gilbert
Distinguished Professor Duncan Pritchard
Assistant Professor Katherine Ritchie

2022

DEDICATION

To my parents

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to begin by thanking my committee chair and advisor Sven Bernecker for his energetic and earnest support of my work and progress, first when I was a member of CONCEPT, and then as a graduate student at UC Irvine. Without a doubt, his encouragement and guidance made this dissertation possible. But more than that, Sven has been a genuine and kind mentor who has remained a steadfast source of warmth and positivity over the past four years.

I am grateful to Duncan Pritchard for his support during my studies at UCI, and for his comments on several drafts of the chapters contained in this dissertation. The discussions I have had with him have been important to the development of my ideas.

I would like to thank Margaret Gilbert for the fun and insightful discussions on group belief. I'm grateful for the opportunity I've had to attend her seminar, and to explore new and uncertain ideas with her.

I would like to thank Kate Richie for her incisive and thoughtful guidance in my research and beyond. Her encouragement and advice have been especially meaningful and helpful in this new phase of my life.

I'd also like to thank Eyal Tal for his invaluable guidance and friendship, and the other people who have helped me and supported me with this project at the University of Cologne: Lisa Benossi, Sofia Bokros, Anna-Maria Eder, Thomas Grundmann, Stefanie Grüne, Steffen Koch, Jakob Olhorst, Francesco Praolini, Luis Rosa, Sibel Schmidt, Paul Silva, and Lucia Volonté.

And thank you to my friends and fellow graduate students in the Department of Philosophy at UCI.

Special thanks to Taylor, for always being there to support me, and for making all of it worth it. Finally, thanks to my parents, who have encouraged me to do what I love and have supported me every step of the way.

The text of Chapter 1 is a reprint of the material was published with co-author Kevin Reuter in *Erkenntnis*, Springer. The text of Chapter 2 is a reprint of the material as it appears in *The Epistemology of Fake News*, Oxford University Press.

VITA

Maria Giulia Napolitano

EDUCATION

Ph.D., Philosophy **2018 - 2022**
University of California, Irvine 2020 - 2022
University of Cologne (CONCEPT) 2018 - 2020

Research Masters (MA), philosophy **2016 - 2018**
Utrecht University

Laurea Magistrale (MA), philosophy **2014 - 2016**
Università degli Studi di Padova

Laurea Triennale (BA), philosophy **2011 - 2014**
Università degli Studi di Padova

PUBLICATIONS

"Conspiracy Theories and Evidential Self-insulation", in Sven Bernecker, Amy Flowerree and Thomas Grundmann (eds.) *The Epistemology of Fake News.* Oxford University Press, 2021.

"What is a Conspiracy Theory?" *Erkenntnis* 1-28, 2021 (co-authored with Kevin Reuter).

ABSTRACT OF THE DISSERTATION

Conspiracy Theories and Resistance to Evidence

By

Maria Giulia Napolitano

Doctor of Philosophy in Philosophy

University of California, Irvine, 2022

Professor Sven Bernecker, Chair

I defend a new account of conspiracy theories, which identifies the epistemological features that make them a socially relevant—and worrisome—phenomenon in our society. On my view, a conspiracy theory is an individual or group belief in a conspiracy which is resistant to disconfirmation by evidence. My approach diverges from the dominant accounts discussed in the literature in two major ways. First, rather than discussing conspiracy theories as a type of theory about a conspiracy, I focus on conspiracy theories as a type of belief in a conspiracy, namely, one that resists revision in light of new evidence. Secondly, conspiracy theories on my view are not only an individual phenomenon, but also, importantly, a collective one. The account I propose, and the shift in focus it produces, provides a new framework for understanding and addressing the phenomenon of conspiracy theories as a problem in our society with deep implications for our political environments.

Introduction

The problem of conspiracy theories

In October 2016, a fake news report went viral associating members of the Democratic Party with a pedophilia sex ring. Shortly before the 2016 presidential elections, a theory started making its way from the alt-right fringes to the mainstream internet linking Hillary Clinton and other democrats to child trafficking taking place in the basement of a pizzeria in Washington DC, the Comet Ping Pong. Despite the lack of robust evidence, and in the absence of any witnesses, the theory spread widely in just a few weeks. When, on December 1st, 2016, a man entered the Comet Ping Pong armed with a rifle in order to investigate and expose the conspiracy, the Pizzagate theory became an example of just how dangerous conspiracy theories can be.

The Pizzagate conspiracy theory exhibits some of the distinctive traits that make conspiracy theories a complex and pressing problem to address, and a puzzling phenomenon to understand. First, Pizzagate is an example of how, no matter how far-fetched and ridiculous, and no matter how poorly supported by the evidence, conspiracy theories are not just a fringe phenomenon. They often manage to spread and gain widespread support. Pizzagate also suggests that, at least in some cases, the support that these theories have is not simply performative: people are not just pretending to believe conspiracy theories, or acting as if they did. They seem to sometimes genuinely believe these theories—enough to act on them, and do so despite the incredibly costly consequences (the Pizzagate shooter, for instance, served 4 years in prison).[1]

[1] https://www.bbc.com/news/world-us-canada-40372407

Secondly, conspiracy theories seem to stand in a particular relation to evidence against them. Conspiracy theories appear to be very resistant, if not immune, to being disconfirmed by countervailing evidence. The Pizzagate shooter did not find any evidence of child trafficking in the pizzeria. In fact, he discovered that the pizzeria did not have a basement at all. One would expect that, after learning from a supporter of the theory that there was no child trafficking happening in the basement of the Comet Ping Pong, Pizzagate would have lost traction. Instead, Pizzagate still has many supporters today, some of whom periodically harass the pizzeria is Washington DC.[2]

Another feature of conspiracy theorizing which is illustrated by Pizzagate has to do with the people who believe conspiracy theories. It seems to be the case that supporters of conspiracy theories operate in groups unified by their joint belief in the conspiracy. Supporters of Pizzagate developed as an online community (especially on Reddit) interested in uncovering the truth regarding Comet Ping Pong and child trafficking.[3] Later on, Pizzagate has become integral to a wider net of conspiracy theories associated with the Qanon group. Conspiracy theories seem to operate not only at the level of individuals, but also at the level of groups.

Finally, conspiracy theories are often politically charged and seem to work exceptionally well in support of political agendas. For instance, Pizzagate started spreading in alt-right environments close to the presidential election in connection to the spread of fake news about Hillary Clinton and members of her party, and it was immediately embraced and developed by Trump supporters. Conspiracy theories are often political in nature, and they seem to flourish when they function to support pre-existing ideological commitments (see Cassam 2019).

This dissertation is about conspiracy theories and the problem they pose to our society in virtue of the features outlined here.

[2] https://www.washingtonian.com/2021/01/19/conspiracy-theorists-target-comet-ping-pong-on-trumps-last-night-in-office/

[3] https://www.buzzfeed.com/craigsilverman/fever-swamp-election

A novel perspective

This approach to the philosophy of conspiracy theories is in tension with the dominant tendency in the philosophical debate to focus on conspiracy theories as explanations that involve conspiracies in order to determine what, if anything, is wrong with them. Most authors engaged in this debate have argued for some version of a descriptive, neutral account of conspiracy theories as explanations involving conspiracies and have tried to determine why conspiracy theories have such a bad reputation. They conclude, more often than not, that nothing is in fact wrong with conspiracy theories as explanation types.[4] In this dissertation, I approach the issue in the opposite way. I start with the assumption that conspiracy theories like Pizzagate are a problem, and I give an account of what the problem is exactly, and how we should best understand conspiracy theories in order to address it.

The reason for this perspectival shift has to do with the methodology of the philosophical debate about conspiracy theories. Conspiracy theories outside of philosophy are typically assumed to be a worrisome phenomenon linked to misinformation, fake news, and mistrust in experts and institutions. It is commonly assumed that belief in conspiracy theories is irrational and that it ought to be minimized, and conspiracy theorists are often ridiculed and ostracized. Many authors in philosophy have conducted their analyses of conspiracy theories by questioning the grounds for the disparaging attitude that people have towards conspiracy theories and those who believe them. Some have argued that this disparaging attitude is more dangerous than conspiracy theorizing itself, as it may lead the dominant opinion to dismiss warranted accusations of conspiring as ridicule conspiracy theories.[5] By focusing on conspiracy theories as mere explanations involving conspiracies, they have attempted to free the label 'conspiracy theory' from the negative connotations—both epistemic and moral—attached to it.

[4] See, for instance Dentith 2014, 2018b; Basham & Dentith 2016; Pigden 1995
[5] For instance, Basham & Dentith 2016; Coady 2007

Here, I advocate for a different approach to the philosophy of conspiracy theories, one that strives to keep the philosophical analysis of conspiracy theories nearer to the public debate around this phenomenon. This approach is meant to provide a framework to talk about the problem of conspiracy theories and the danger they pose to our society. The aim of such a framework is twofold. First, the framework proposed in this dissertation provides adequate grounds for treating conspiracy theories as a problem to be addressed, while avoiding the dangers identified by other philosophers associated with employing the label 'conspiracy theory' to silence accusations of conspiracies. Second, this framework advances a specific diagnosis of the problem of conspiracy theorizing as a problem of individual and collective resistance to evidence.

I argue that conspiracy theories are a problem of evidence-resistant beliefs of individuals and groups. Conspiratorial beliefs, rather than theories about conspiracies, are the focus of my dissertation. By focusing on beliefs in conspiracies which are resistant to evidence, I argue that we can center our analysis on the same target as ordinary discussions of conspiracy theorizing. At the same time, we are able to make this target more precise and to provide an explanation of what goes wrong with conspiracy theories.

Preview

The first chapter, "What is a conspiracy theory" (joint paper with Kevin Reuter) lays the methodological foundation for the account of conspiracy theory defended in the other chapters. We start by identifying two assumptions commonly made in the philosophical debate by those authors who argue that 'conspiracy theory' ought to be understood neutrally as a theory about a conspiracy. It is assumed that either: (i) *conspiracy theory* is a neutral concept in ordinary language, and does not have a strong evaluative component—except perhaps as a pragmatic implication; or (ii) that *conspiracy theory* is predominantly a negatively

loaded term, but it ought to be engineered neutrally to minimize the risk of exploiting the label to silence warranted conspiracy accusations. We argue against the first assumption by presenting the results of four empirical studies which suggest that *conspiracy theory* in ordinary language has a negative connotation. We also argue, against the second assumption, for an empirically informed engineering strategy which explicates *conspiracy theory* as an evaluative concept with a negative epistemic connotation. I develop an account of *conspiracy theory* which follows this engineering strategy in chapters two and three.

In the second chapter, I follow the engineering strategy identified in chapter one, and I argue that we should understand *conspiracy theory* as a epistemically evaluative concept. In particular, I argue for an account of conspiracy theories as a type of irrational mental state—namely, a belief in a conspiracy which is immune to being revised in light of disconfirming evidence. I adopt a general Bayesian notion of rationality to show that a belief in a conspiracy—where belief stands for any credence above a certain threshold—which is immune to being disconfirmed in light of incoming evidence is irrational. While my proposal is a radical shift from the standard accounts of *conspiracy theory* as an explanation involving a conspiracy, the results of the discussion in this chapter has consequences for many of these accounts. In fact, it is often argued that conspiracy theories are unfalsifiable, but that this is not a problematic feature because a theory about a conspiracy predicts misleading evidence. In this chapter, I argue that theories about conspiracies cannot be rationally unfalsifiable, and that if we are interested in unfalsifiable conspiracy theories, then we need to focus on irrationally held conspiratorial beliefs.

Chapter three offers an account of a collective phenomenon that arises with conspiracy theories, which I call *group conspiracy theory*. A group conspiracy theory is the belief in a conspiracy by a group which is held together solely by the commitment to believing in the conspiracy. Group conspiracy theories are hard for groups to abandon when evidence against them arises because of the special obligations that members of the group have towards

each other in virtue of their belief, without which the group would cease to exist. This analysis of conspiracy theorizing as a collective phenomenon helps to explain the resilience of conspiracy theories in the public domain, and how internet communication has made it easier for collective conspiracy theories to form and survive despite mounting evidence against them. It also suggests that in order to fully address the problem of conspiracy theorizing in our society, it is necessary to devise strategies that target these collective dynamics, in addition to individuals' reasoning and assessment of evidence and authorities.

The fourth chapter presents an analysis of the relation between conspiracy theories and political propaganda. I argue that conspiracy narratives work as propaganda when they function to spread conspiracy theories—in the sense defended in chapter two—in the audience. I focus on the features that make conspiracy beliefs prone to becoming evidentially insulated to show how propagandistic messages can facilitate the spread of irrational conspiracy theories. Conspiracy theories, in turn, work as powerful ways to shape the relations of epistemic trust of their believers, polarize their opinions, and ultimately close off rational debate with dissenting voices.

Chapter 1

What is a Conspiracy Theory?*

1.1 Introduction

Discussions on conspiracy theories pervade internet forums, social media, and the news. In both public and academic discourse, conspiracy theories are often taken to undermine trust in institutions and to hinder the spread of information (for instance, Sunstein & Vermeule 2008; Dieguez et al. 2016), and to be used as instruments of political propaganda (Muirhead & Rosenblum 2019; Cassam 2019). Not surprisingly, the expression 'conspiracy theory' seems to carry with it a negative value. This manifests itself in various ways. Conspiracy theorists are often portrayed as irrational people (Coady 2007), and few people are willing to apply the label 'conspiracy theory' to their own views (Wood & Douglas 2013:7). In fact, expressions such as 'this is just a conspiracy theory!' seem to be often employed to dismiss certain theories as rumors or speculations. In other words, 'conspiracy theory', at least on the face of it, seems to be a negatively loaded expression.

This feature is reflected in the academic discussion of conspiracy theories. Scholars

* This chapter is a reprint of the joint paper with Kevin Reuter as it appears in *Erkenntnis* (2021). This work is fully collaborative.

have been interested in finding ways to minimize belief in conspiracy theories (for instance, Swami et al. 2014) and to understand which psychological factors drive belief in conspiracy theories (for instance, Bilewicz, Cichocka & Soral 2015; Douglas et al. 2016; Swami et al. 2010; Swami 2012). Conspiracy theories are assumed in many academic discussions to be something which should not be believed, and they are considered akin to rumors (Berinsky 2015), false beliefs, and misinformation (Lewandowsky et al. 2012).

However, philosophers working on this topic have been reluctant to include an evaluative element when analyzing or engineering the concept *conspiracy theory*, instead advocating for purely neutral definitions. This paper challenges the methodological approach that has led philosophers to focus on neutral definitions, and suggests a way forward that relies on empirical investigations of the ordinary concept *conspiracy theory*. Given how heavily the discussion over defining conspiracy theories relies on empirical assumptions about the ordinary meaning of the expression, especially about the relation between its evaluative and descriptive dimensions, it comes as a surprise that such empirical investigations have not been conducted on the ordinary use of the expression 'conspiracy theory'.[1] In this paper, we present empirical data on the nature of the ordinary concept *conspiracy theory* in order to foster a more careful discussion of how it should be defined or engineered.

We start in section 2 with a survey of the most prevalent accounts of *conspiracy theory* in the philosophical debate, and identify two different assumptions that have been made when defining *conspiracy theory*: (1) the ordinary concept *conspiracy theory* is predominantly descriptive; (2) the evaluative *conspiracy theory* serves the function of silencing warranted accusations of conspiring. In section 3, we present five studies on the ordinary meaning of 'conspiracy theory'. Our studies show that, while some people seem to have a descriptive concept of *conspiracy theory*, the most widespread uses of the expression 'conspiracy the-

[1] Wood (2016) investigates the consequences of labelling a theory 'conspiracy theory'. While this might provide some insight into the *consequences* of using this term, it does not constitute an analysis of it. Some remarks about the negative connotation of *conspiracy theory* have been made by Wood & Douglas (2013).

ory' indicate the predominance of an evaluative concept. Moreover, they suggest a double dissociation between conspiracy and conspiracy theory: referring to a conspiracy is neither necessary nor sufficient for attributions of 'conspiracy theory'. In section 4, we discuss some implications of these results for the existing accounts of *conspiracy theory*, and present our own strategy for an epistemically evaluative engineering.

1.2 Defining *Conspiracy Theory*: Philosophical Methodology

The philosophical literature on conspiracy theories includes a variety of attempts at defining *conspiracy theory.* We identify two general approaches, which we label as *descriptive conceptual analysis* and *conceptual engineering.* By *descriptive conceptual analysis*, we mean those analyses which aim to provide the meaning of a term, and which are acceptable only if they are consistent with our ordinary intuitions. These proposals typically make use of the method of cases—e.g., they consider cases that are publicly known as 'conspiracy theories'—and then try to identify the necessary and sufficient conditions that make something a conspiracy theory. We use the label *conceptual engineering* to include all of those revisionary approaches to defining 'conspiracy theory' which do not aim to match our intuitions about cases, but rather *improve on the ordinary concept*, by defining *conspiracy theory* in a way that serves a certain theoretical or practical goal. Framing the issue in terms of conceptual engineering allows us to draw from a rich discussion about this philosophical methodology and the different approaches that fall under this label. We understand conceptual amelioration, or ameliorative analysis (Haslanger 2012; 2020a) as a type of conceptual engineering aimed at improving social reality by focusing on the purposes or functions of concepts. Another popular method, explication, proceeds from an often vague, informal concept, to provide a

more exact and fruitful one, with the aim of improving a theoretical discussion.[2] Within the literature on conspiracy theories, we can recognize engineering attempts akin to both kinds. In this section, we employ the labels of analysis and engineering to review the different claims and methods from prominent authors, in order to pin down the empirical assumptions that help to justify their conclusions for defining *conspiracy theory* in a descriptive, neutral way.[3]

1.2.1 Descriptive Conceptual Analysis

The method of descriptive conceptual analysis for *conspiracy theory*, understood as the attempt to devise a definition which matches our folk intuitions about conspiracy theories, has most explicitly been defended by Räikkä (2018)—though under a different label. In providing a survey of the different proposals that have been put forward for defining conspiracy theories, Räikkä identifies three different approaches: to narrow, expand, or preserve the ordinary meaning of 'conspiracy theory' (2018:207). He argues that, when focusing on the question of our handling of conspiracy theories in our societies, philosophers' understanding of conspiracy theories should try to approximate the ordinary language meaning of 'conspiracy theory'. The idea is that, given that the public interest for conspiracy theories is driven by the practical and political problems with those theories that are commonly labeled 'conspiracy theories', we should neither expand nor narrow the content of the ordinary concept, but rather we should try and give a definition of *conspiracy theory* that picks out the same object as ordinary language, to then determine whether conspiracy theories so understood are in fact a problem. According to Räikkä, conspiracy theories are explanations which usually satisfy two conditions: (i) the *conflict criterion*, i.e., an explanation is a conspiracy

[2] For a comparison between explication and ameliorative analysis see Dutilh Novaes (2020).

[3] Our use of the notions of conceptual analysis and conceptual engineering relies on one possible understanding of these philosophical methods. While we employ these labels as a way to organize the recent philosophical debate on conspiracy theories, we are aware that the way in which we characterize the two approaches are neither exhaustive nor uncontroversial. For a discussion of conceptual analysis along the lines that we propose see Daly (2010). For a discussion of conceptual engineering approaches, see Burgess & Plunkett (2013a, 2013b); Cappelen (2018); and the papers in Cappelen, Plunkett, & Burgess (2020).

theory, only if it is in conflict with a received explanation of the same event, and (ii) the *conspiracy criterion*, i.e., an explanation is a conspiracy theory only if it refers to a conspiracy or plot (Räikkä 2018: 210-213). These criteria are taken to be descriptive, and *conspiracy theory* is not taken to necessarily encode in ordinary language a negative evaluation.[4]

Even though in Räikka (2018) we find the most explicit defense of the method of descriptive conceptual analysis as opposed to conceptual engineering of *conspiracy theory*, analyses of *conspiracy theory* have been proposed in the literature since its infancy. For instance, Coady (2003) argues that *conspiracy theory* should be defined as:

> A conspiracy theory is a proposed explanation of an historical event, in which conspiracy (i.e., agents acting secretly in concert) has a significant causal role. Furthermore, the conspiracy postulated by the proposed explanation must be a conspiracy to bring about the historical event which it purports to explain. Finally, the proposed explanation must conflict with an "official" explanation of the same historical event. (Coady 2003: 201).

Similar to Räikkä's, Coady's definition includes a version of the conspiracy criterion, and a version of the conflict criterion.[5] Another widely endorsed definition of *conspiracy theory* is Keeley's. For Keeley, a conspiracy theory is 'a proposed explanation of some historical event (or events) in terms of the significant causal agency of a relatively small group of persons—the conspirators—acting in secret' (1999: 116). Keeley's proposal does not include a conflict criterion, and only takes some version of the conspiracy criterion to be necessary and sufficient for *conspiracy theory*.[6]. These definitions have in common the fact that they

[4] More recently Ichino & Räikkä (2020) have included an additional *evidence criterion*, according to which a conspiracy theory 'offers insufficient evidence in support of the alternative explanation, so that the view is not considered as a competitive scientific theory or anything like that'. This addition makes their definition epistemically evaluative, and more in line with the findings of our study.

[5] The conflict criterion might be thought to encode an evaluation. However, both Coady and Räikka seem to understand the conflict criterion purely descriptively. Levy (2007) has taken the conflict criterion to be normative, putting it in relation to epistemic authorities.

[6] Other attempts to descriptively analyse the concept *conspiracy theory* which include a conspiracy crite-

take *conspiracy theory* to be a descriptive concept. In fact these authors agree that from the criteria they identify, it does not follow that conspiracy theories are necessarily irrational, or that they should not be believed. The negative connotation ordinarily attached to conspiracy theories has often been interpreted in these discussions as a pragmatic feature of *conspiracy theory*. Pigden (2007) calls it the 'conventional wisdom'— the widespread, and yet mistaken, belief that we have an epistemic duty not to believe or investigate theories about conspiracies. Some other authors who have adopted neutral definitions have identified a subclass of conspiracy theories as irrational and have introduced negatively loaded expressions to refer to them—such as Unwarranted Conspiracy Theories (Keeley 1999), and Counterfact Conspiracy Theories (Feldman 2011).

Another element that the different analyses of *conspiracy theory* have in common is the methodology employed to identify the criteria. From a methodological point of view, these proposals identify the defining features of *conspiracy theory* by analysing popular theories which are usually labeled 'conspiracy theories'—such as: the theory that climate change is a hoax, and the theory that genetically manipulated foods cause health problems (Räikka 2018); official and alternative accounts of 9/11 attacks (Coady 2003); conspiracy theories associated with the Oklahoma City bombing (Keeley 1999); the theory according to which Obama was not born in the US (Feldman 2011).

Even though these philosophers might be correct in identifying instances of conspiracy theories in the public discussion, the methodological choice of relying on putative representative examples of theories which are generally labeled 'conspiracy theories' risks being misleading when it comes to understanding the possible evaluative component of *conspiracy theory*. By looking at allegedly representative cases of conspiracy theories, and the properties that these theories share, it is difficult to identify speaker-sensitive evaluations. If, as

rion, sometimes supplemented with a descriptive conflict criterion, include Cohnitz (2018); Feldman (2011); Harris (2018); Mandik (2007); Pigden (2007). A different proposal comes from Cassam (2019). He introduces the label 'Conspiracy Theory' with capital C and capital T to refer to a subset of theories about conspiracies which display additional epistemically problematic features.

we hypothesize, *conspiracy theory* is an evaluative concept, it is necessary to focus on what drives speakers' attributions of 'conspiracy theory', rather than on the features that famous conspiracy theories have in common.

The methodological choice of relying on famous theories known as 'conspiracy theories' might be one of the factors that led the authors discussed in this section to focus on neutral definitions. That is why, in this paper, we opted for different methods, chosen specifically to test for the existence of an evaluative dimension, and investigate its prevalence and characteristics.

1.2.2 Conceptual engineering

While projects of descriptive conceptual analysis have largely taken the expression 'conspiracy theory' to pick out a descriptive concept in ordinary language, the situation is different for engineering proposals. In fact, some philosophers have acknowledged that *conspiracy theory* has a negatively evaluative meaning, but have argued that the ordinary meaning ought to be changed.[7]

Coady argues, on the basis of examples from the scholarly and public debate, that the expressions 'conspiracy theory' and 'conspiracy theorist' have multiple meanings, which are typically used equivocally in the academic and political debate. In particular, according to his analysis, 'conspiracy theory' is both employed pejoratively to dismiss certain theories and descriptively to indicate theories about conspiracies. The equivocation that follows from this semantic ambiguity has negative consequences both for our theorizing about conspiracy

[7] None of these philosophers have used the expression 'conceptual engineering' to describe their approach to *conspiracy theory*, and have instead talked about stipulative definitions. We believe that the stipulative definitions discussed here can be seen as instances of engineering, in the very broad sense of the expression adopted here, as assessing and improving representational devices (Cappelen 2018). Moreover, our notion of engineering includes both conceptual revision for practical purposes, such as improving our social and political environment, and for theoretical purposes, such as providing a more fruitful notion to improve a scientific debate.

theories and for the social environment in which these terms are used (Coady 2018a: 292). In fact, it allows for dismissing theories about conspiracies, even when these theories are epistemically justified. For Coady, the pejorative expression 'conspiracy theory' is employed to dismiss conspiracy accusations, and it constitutes a form of epistemic injustice against people who profess believing theories about conspiracies (2018a: 300). For this reason, he argues that we should refrain from using the expression 'conspiracy theory' and neighboring ones (see also Coady 2007; 2012; 2018b).

Other philosophers seem to share Coady's worry that the label 'conspiracy theory' can lead to the illegitimate dismissal of warranted theories about conspiracies. Basham & Dentith (2016) have claimed that the pejorative meaning of 'conspiracy theory' is routinely abused by politicians and institutions to dismiss unwanted conspiracy allegations. They write:

> Much contemporary media, most political leaders and some social scientists insist that "conspiracy theory" must mean something automatically false or irrational. Yet our historians show it does not and never did. The pejorative use of "conspiracy theory" is a use of mere convenience. The official account of 9/11 is, after all, a conspiracy theory: the hijackers conspired to fly airplanes into buildings in New York City, Washington, and elsewhere. That's a conspiracy theory. Was it called that? Not by mainstream media, or most political leaders. But it was, just the same. Any pejorative use of "conspiracy theory" is intellectually suspect, as is its convenient absence when governmental institutions use conspiracy theories to promote their goals. We are facing a phrase of social manipulation, one which some academics wish to portray and empower in a way so that it cannot impugn our hierarchies of power, but only defend them. The only conspiracy theories permitted will be official conspiracy theories. They will not be called "conspiracy theories." But their explanatory method will be indistinguishable. (Basham & Dentith 2016:15)

Differently from Coady, who argues that the ambiguity of 'conspiracy theory' calls for eliminating the expression altogether, Basham and Dentith advocate for conceptual change towards a neutral definition of *conspiracy theory*. They argue that, in both academic and public discussions of conspiracy theories, the evaluative meaning should be abandoned in favor of the descriptive sense of *conspiracy theory* as any explanation of events that cites a conspiracy (for instance, Dentith 2014; Basham 2018b).

The arguments for changing the meaning of 'conspiracy theory' reviewed so far have in common that the expression 'conspiracy theory' is assumed to have a negative effect on our political environments.[8] 'Conspiracy theory' is taken to be a powerful label that can be exploited to ridicule accusations of conspiring, allowing conspiracies to continue unnoticed. In this respect, Basham, Coady, and Dentith's proposals for engineering *conspiracy theory* can be considered as ameliorative in their intent: they aim at conceptual change primarily to improve the effects of the use of this expression on our society.[9]

Dentith (2014, 2018c) also offers a different argument for engineering *conspiracy theory* neutrally. They argue that a neutral and minimal definition of conspiracy theory as an explanation of an event which involves a conspiracy is best suited to promote theoretical discussions about the rationality of believing conspiracy theories (2014:123). Dentith's motivation for proposing such a definition is theoretical, rather than practical. Their aim is to devise a stipulative definition which can better be employed within the academic discussion of conspiracy theories; and they seem to have the academic community as a target for their definition. In this sense, Dentith's proposal is different from the other engineering approaches discussed so far. However, Dentith does not seem concerned with matching our folk intuitions about the concept *conspiracy theory*, and thus we take their proposal to be an instance of engineering in the sense adopted in this paper—albeit one closer to

[8] See also Basham & Dentith 2016; Dentith 2018b, 2018c; Basham 2018a; Hagen 2018; Orr & Dentith 2018.
[9] For a discussion of these effects on society, see Stokes (2018), and the responses to his paper in the same volume by Basham and Dentith.

an *explication* rather than an *amelioration*. In the rest of the paper, we will focus on the ameliorative approaches which are guided by social and political concerns, and leave out Dentith's theoretically motivated engineering. But their proposal will be discussed again in section 4.2.2.

Even though the engineering proposals defended by authors such as Coady, and Basham and Dentith (in their joint work) are ameliorative, and explictly presented as a shift from the way in which the expression is ordinarily used, their proposals are still bound by assumptions regarding the ordinary concept *conspiracy theory* and its function in ordinary language. First, they rely on the assumption that the current uses of the label 'conspiracy theory' are defective in that they make people prone to dismissing charges of conspiracies without evaluating them—possibly because the expression is ambiguous. Moreover, they seem to be assuming that the main function that the evaluative *conspiracy theory* plays in society is that of serving the interest of the powerful by discouraging people from investigating conspiracies. If they did not subscribe to this idea, one might argue against their proposal that, even though eliminating the evaluative concept would help promoting societal goods, it would also generate undesirable conceptual loss in our representational resources. Moreover, adopting a neutral definition for investigating conspiracy theories in academic settings would run the risk of driving a wedge between the academic and the public discussion of this phenomenon. And, given the relevance of the phenomenon of conspiracy theories to the public debate, this detachment of the academic discussion from ordinary talk of conspiracy theories could lead to further problems of equivocation and confusion. So, even though the ordinary meaning of 'conspiracy theory' is only indirectly relevant to these ameliorative proposals, they still are relying on crucial empirical assumptions regarding the ordinary use of this expression.

While the academic and political uses of the expression have been more thoroughly documented (Husting & Orr 2007; Coady 2012; 2018a), the ordinary use of *conspiracy theory* has not been systematically investigated empirically, and these philosophers have primarily

relied on examples and personal observations on the use of the expression.

Basham (2018) takes Wood (2016) to provide empirical evidence that *conspiracy theory* does not have a negative connotation:

> Michael Wood (2016) shows what we already know; "conspiracy theory" possesses no negative connotation except as residue among certain academic, media, and political elites. (Basham 2018: 40)

However, Wood's study only shows that labelling a theory a 'conspiracy theory' does not lead people to reduce their belief in that theory, compared to when that theory, with the same exact content, is called in a different way. The experiments found 'no evidence of a negative effect of calling something a conspiracy theory' (Wood 2016: 702). But this result does not necessarily imply that the meaning of the expression is neutral in ordinary language. In any case, this study speaks *against* the worry expressed by Basham and others that the label 'conspiracy theory' could be used to dismiss warranted conspiracy accusations. If labelling a theory 'conspiracy theory' does not affect belief in that theory, it is hard to see how this label could be employed for inducing people to dismiss certain warranted theories about conspiracies.

The empirical studies we present in the next section suggest that *conspiracy theory* has an evaluative meaning which is prevalent in ordinary language. They also confirm the worry that *conspiracy theory* is ambiguous. Ultimately, our studies provide empirical support for the claim that the authors reviewed in this section make, that *conspiracy theory* should be engineered. However, they also provide insight into the function of the ordinary evaluative concept, and end up putting significant pressure on the specific proposal of engineering *conspiracy theory* neutrally, or eliminating the label altogether.

1.3 Empirical Studies

While many philosophers have either taken the meaning of 'conspiracy theory' to be descriptive, or argued that the evaluative meaning should be eliminated, no one has so far investigated the folk concept of *conspiracy theory* empirically. Instead, philosophers have largely focused on individual cases that they have often considered to be representative of the general phenomenon. In what follows, we opted for a different methodological approach. In five individual studies, we empirically investigated the uniqueness, prevalence, and the form of the evaluative sense of *conspiracy theory*. Our studies are divided into two sections. In section 3.1 we aim to show that the dominant meaning of 'conspiracy theory' is evaluative. To show this, we used three different experimental paradigms (semantic feature production task, vignette study, corpus analysis). In section 3.2 we designed two studies that indicate a double dissociation between conspiracy and conspiracy theory. In other words, we show that neither do the application conditions for the term 'conspiracy theory' include a conspiracy condition, nor do people think a conspiracy warrants the application of the term 'conspiracy theory'.

1.3.1 The Evaluative Sense of *Conspiracy Theory*

In three individual studies, we empirically investigated the evaluative sense of *conspiracy theory*.

Study 1a: Semantic features of *conspiracy theory*

Semantic feature production tasks are standardly used to collect those semantic features that are encoded within a concept (McRae et al. 2005) . For instance, common features that people come up with for the term 'robin' are 'flying', 'has wings', 'lays eggs', 'a bird',

etc. The aim of Study 1a was to collect those features that are encoded within the concept conspiracy theory.

While standard semantic feature production tasks have been argued to provide access into word meaning (Medin 1989) and predict semantic processing in various tasks (Pexman et al. 2003), the results cannot straightforwardly be taken as evidence that the dominantly produced features are necessary for the application of the concept. Instead, some features might be merely salient (but not necessary) features of the concept at stake. To overcome this limitation, we decided to use a slight variation of the semantic feature production task (see e.g., Reuter et al. 2020), in which participants are asked to produce necessary features, and not just any features that come to mind.

50 participants ($M_{age} = 33.94$; 20 females, none indicating non-binary gender, all English native speakers) were recruited on Amazon's Mechanical Turk and were paid a small fee for taking part in the study. They were asked to note down three necessary conditions for something to be regarded a conspiracy theory.

Vignette Think for a moment about what it means to be a conspiracy theory. Please state three conditions that you believe are necessary for something to be a conspiracy theory.

In total, 150 responses were collected. Table 1 shows the responses for the first 25 participants.

We were particularly interested in features expressing descriptive aspects of conspiracy theories, like 'some sort of cover up' (or simply 'cover up' or likewise), 'it involves a conspiracy' (or simply 'conspiracy' or likewise), etc. Additionally, we wanted to know how many people would note down features indicating an evaluative sense of *conspiracy theory*. When we scanned the responses for evaluative features, we noticed that people came up with terms

Participant	1st Term	2nd Term	3rd Term
Person1	questions	vagueness	coverup
Person2	shady dealings ...	lies	manipulation of an event ...
Person3	more theory than fact	ties to government coverup	farfetched
Person4	lack of evidence	it involves someone powerful	hard to believe
Person5	realistic	evil	factual
Person6	UNTRUE	ILLOGICAL	IRRATIONAL
Person7	no answer to the question	many people believe it	no evidence to refute it
Person8	controversial	mysterious	weird
Person9	needs to be a conspiracy	powerful people	unlikely to be true
Person10	... secret group involved	power must be involved ...	deceit or deception ...
Person11	facts	obsession	curiousity
Person12	Not a popular belief.	Science doesn't back it up.	Controversial.
Person13	1 group believe something	2nd group is against the 1st ...	Published / known fact
Person14	Cult Following	Grasping evidence	outlandish claim
Person15	unproven info	no consensus	lack scientific data
Person16	over the top	crazy	out there
Person17	made up	paranoid	outlandish
Person18	Multiple people believe it	It's controversial	Has some evidence to back it
Person19	Uncertainty	Lack of concrete facts	People believe it
Person20	Extreme theory	Intrigue	Espionage
Person21	Not provable	insane	random
Person22	sensational	substantial coverup	crazy or crazy sounding
Person23	lots of people believe it	the government is ... part of it	at least 3 people follow it
Person24	not widely believed	hidden actors/motiives	some sort of cover up
Person25	motive	challenge	crazy people

Table 1.1: Responses of the first 25 participants in the semantic feature production task.

that fall very roughly into two different categories: (i) terms that indicate epistemic deficiencies, like 'untrue', 'no evidence', 'far-fetched', and 'not provable'; (ii) terms that are more generally dismissive and pejorative, like 'crazy', 'paranoid', and even moral terms like 'evil'. Features from both evaluative categories were so common among many participants that we decided to ask three independent coders (x / y / z) to classify all 150 terms into three groups: terms indicating epistemic deficiencies, strongly disparaging terms, and all others.

The high number of participants responding with evaluative terms, including terms expressing epistemic deficiencies as well as derogatory terms, suggests that the predominant sense of *conspiracy theory* is evaluative: A substantial majority (66% / 74% / 70%) does not entertain a notion of *conspiracy theory* that is primarily descriptive. Instead, most people seem to encode evaluative features in their concept of *conspiracy theory*. However, not all participants who think normatively about conspiracy theories, go so far in entertaining a sense of the term 'conspiracy theory' that indicates a strong pejorative attitude. Nonetheless, (32% / 38% / 32%) of the participants not only thought that conspiracy theories are

epistemically deficient theories, but considered them to be crazy or ridiculous explanations, or similarly. In regards to the descriptive terms, eleven participants wrote that a conspiracy theory involves a cover-up, four participants responded with 'conspiracy' or likewise, and a mere single person stated that a conspiracy theory is claimed to be false by officials. This suggests that the cover-up criterion might be a more prevalent descriptive criterion compared to the more widely discussed conspiracy and conflict criteria. While the results of Study 1a provide us with a window into the semantic representation of the concept of conspiracy theory, we need to be careful not to overstate the evidence, but rather take the results as one important piece in our argument. We therefore investigated the extent to which participants hold an evaluative concept of conspiracy theory using a classical vignette study.

Study 1b: The prevalence of the evaluative meaning

The aim of Study 1b was to examine more directly the extent to which people entertain an evaluative meaning of that concept. 111 participants were recruited on Amazon's Mechanical Turk. 10 participants had to be excluded, because they either indicated that English was not their native tongue, or they did not finish the survey. The average age of the remaining 101 participants was $M_{age} = 37.81$ and consisted of 46 females, 55 males, and none indicating non-binary gender. All participants were randomly presented with one of the following two short vignettes:

Imagine Peter makes the following statement:

Peter: Maria says that the theory I've heard about is a conspiracy theory. / Maria says that the theory I've heard about is a scientific theory.

The vignette featuring 'scientific theory' was included as a control condition, in order to

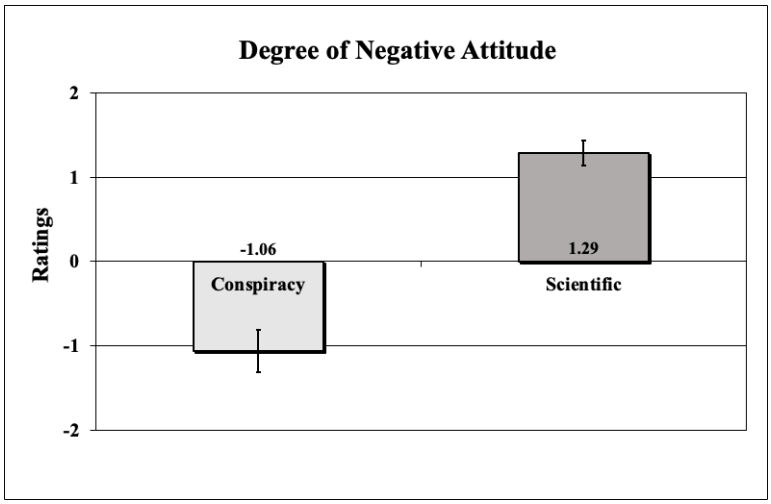

Figure 1.1: Mean values of the ratings in Study 2. Error bars indicate standard error around the means.

eliminate the possibility that the way we asked the question would bias people into thinking negatively about Maria's attitude. The participants were then presented with the following statement "Maria's attitude towards the theory that Peter has heard about is:" and asked to rate Maria's attitude on a 7-point Likert scale anchored at '-3' meaning 'negative', '0' meaning 'neutral', and '3' meaning 'positive'. The average rating for the condition featuring 'conspiracy theory' was $M_{av} = -1.06$ (SD = 1.77), and was significantly below the midpoint of '0', t(51)=-4.32, p<0.001. The mean value for the control condition (scientific theory) was $M_{av} = +1.29$ (SD = 1.02), and was significantly above the midpoint: t(48)=8.81, p<0.001. Of the 52 participants who were randomly assigned to the conspiracy vignette, 33 participants gave a negative rating, and 11 participants gave a positive rating. Figure 1 shows the average ratings for both conditions.

The results of Study 1b reveal that a substantial majority of the participants selected negative values. Comparing the results of Study 1a with Study 1b, a rather uniform pictures evolves, according to which around 2/3rds of the population seem to think that the concept *conspiracy theory* is negatively evaluative.

Study 1c: Corpus Analysis

Studies 1a and 1b have two shortcomings: First, they test people's use of terms in isolation from their natural context. Second, the experimental setup might have interfered with getting reliable results on the meaning of the target expressions. Within the last decade, philosophers have started to use tools from corpus linguistics to examine terms within large corpora (see, e.g., Sytsma et al. 2019). These tools avoid both these shortcomings by investigating how terms like 'conspiracy theory' are used by ordinary language speakers in natural contexts.

In Studies 1a and 1b, we collected data showing that many people entertain evaluative features when thinking about conspiracy theories. If the dominant meaning of 'conspiracy theory' were indeed evaluative, we should be able to find further support for this claim when analyzing large corpora. As we are primarily interested in the way that ordinary people use the term in natural contexts, a corpus using posts and discussion notes from reddit, the social news website, was built (no specific subreddits were selected, such that the selected comments are unlikely to have been tilted towards a certain readership and/or distorted by specific jargon). We then searched for and collected comments featuring the term 'conspiracy theory/ies/ists' in a time span of 120 days (25th December 2019 - 22nd April 2020), and ended up with 68'640 comments.[10]

In order to investigate the evaluative dimension of *conspiracy theory*, we recorded all adjectives that appear directly in front of 'conspiracy theory' like 'crazy conspiracy theory'. If the expression is used in a predominantly neutral way, then we should find a balanced representation of positive, neutral and negative adjectives, e.g., 'plausible/political/crazy conspiracy theory'. In contrast, if the concept is mostly used negatively, we would expect

[10]Given the limited time span of our corpus analysis, we cannot be sure that the way people talked about conspiracy theories in 2019 and 2020 is representative of the years before and the years to come. Some conspiracy theories most talked about in these two years might have features that other conspiracy theories discussed at other times do not have.

those adjectives to be strongly negative as well.

Among the 50 most frequent adjectives — covering 42% of all uses we collected — 25 adjectives were negative, 17 neutral, and 8 positive. Table 2 (left hand side) lists the 10 most frequent adjectives that occur in front of 'conspiracy theory'. These results strongly suggest that conspiracy theories are not predominantly considered to be neutral theories, but very often considered to be inherently negative.

However, against this conclusion, a plausible objection might be made. The evaluative terms we recorded might not be indicative of the meaning of 'conspiracy theory' but rather suggest that the meaning was modified by those terms in order to express an evaluation that is not part of 'conspiracy theory' itself. However, we believe we have good reasons to dismiss this alternative reading. If negative adjectives merely modify but don't intensify the evaluative aspect of 'conspiracy theory', then we should not find a major difference between 'conspiracy theory' and a clearly neutral term like 'theory'. We therefore conducted an analysis of the term 'theory' using the same method as above. The results are markedly different (see also Table 2 (right hand side)). Among the 50 most frequent adjectives, only 6 were negative, 13 positive, and 31 neutral.

Conspiracy Theory			Theory		
Term	Number	Percentage	Term	Number	Percentage
crazy	684	4.1%	good	872	5.3%
good	460	2.8%	interesting	752	4.6%
stupid	377	2.3%	economic	704	4.3%
ridiculous	342	2.1%	scientific	586	3.6%
wild	272	1.6%	political	456	2.8%
insane	267	1.6%	great	450	2.7%
dumb	238	1.4%	personal	387	2.4%
racist	230	1.4%	critical	336	2.0%
favorite	208	1.3%	different	252	1.5%
weird	208	1.3%	popular	247	1.5%

Table 1.2: A list of the 10 most frequent adjectives in front of 'conspiracy theory' and 'theory'.

Summary of the Results

In Study 1a, we used the semantic feature production task to detail the terms most commonly associated with 'conspiracy theory'. The results suggest that the concept *conspiracy theory* is often considered evaluative. In Study 1b, we asked a direct and simple question about whether the content of the concept *conspiracy theory* is evaluative. Most people believe the term 'conspiracy theory' to have a negative meaning. For Study 1c, we built a corpus from the social media agglomeration website *reddit* in order to examine the term 'conspiracy theory' in natural language use. Our data analysis revealed a strong co-occurrence of negative adjectives with the term 'conspiracy theory', thereby confirming our previous studies that involved online participants.

1.3.2 Double Dissociation of Conspiracy Theory and Conspiracy

While the first three studies indicate the existence and dominance of an evaluative meaning of 'conspiracy theory' among laypeople, we have not directly investigated the role of the descriptive features that many proponents of purely descriptive accounts have defended. Unfortunately, testing the importance of all the proposed descriptive features, is beyond the scope of this paper. We have therefore decided to examine what seems to many philosophers to be the *core descriptive aspect*. More specifically, we tested two hypotheses: First, people are willing to call an explanation a 'conspiracy theory', even in situations in which no conspiracy has taken place.[11] Second, people tend not to call a claim a conspiracy theory, even if a conspiracy has taken place. In other words, we aim to show a double dissociation between conspiracy theory and conspiracy.

[11] A similar hypothesis has been considered, though not investigated empirically, by Walker (2018).

Study 2a: Conspiracy Theory without Conspiracy

In Study 2a, we tested whether people are willing to call a claim a conspiracy theory even though no conspiracy has been taken place. We decided to manipulate two different aspects. First, we aimed to measure the effect that a real conspiracy would have on people's willingness to call a claim a conspiracy theory. Thus, in the conspiracy condition *Conspiracy*, we specified that "the Incas destroyed most of the evidence, and made it look as if they had built the temple themselves." In the no-conspiracy condition *Lost in Time*, we stated that "everyone knew about it back then, ... most of the evidence of it got lost over time." Second, the vignettes either featured an explanation *Aliens* that was epistemically highly deficient (an explanation that referred to aliens), or an explanation *Chancas* that was not the official explanation, but not too far-fetched (an explanation that referred to the Chanca tribe).

All participants were randomly assigned to one of four conditions: *Aliens Conspiracy*, *Aliens Lost in Time*, *Chancas Conspiracy*, *Chancas Lost in Time*. Here are the two vignettes (Aliens Lost in Time & Chancas Lost in Time (in Square brackets)) we used for the no-conspiracy conditions.

Imagine you overhear the following conversation:

Luke: I think that the famous Coricancha temple was built by aliens [by the Chancas] and not by the Incas.

Anna: Are you kidding? And how do you explain that we know nothing about this?

Luke: Well, I believe that everyone knew about it back then, but the temple was built such a long time ago, so most of the evidence of it got lost over time.[12]

[12]In the two conspiracy conditions, the second statement of Luke was "Well, I believe the Aliens built [Incas recruited the Chancas to build] the temple. But the Incas did not want the world to know that they did not build the Coricancha temple themselves, so they destroyed most of the evidence, and made it look as

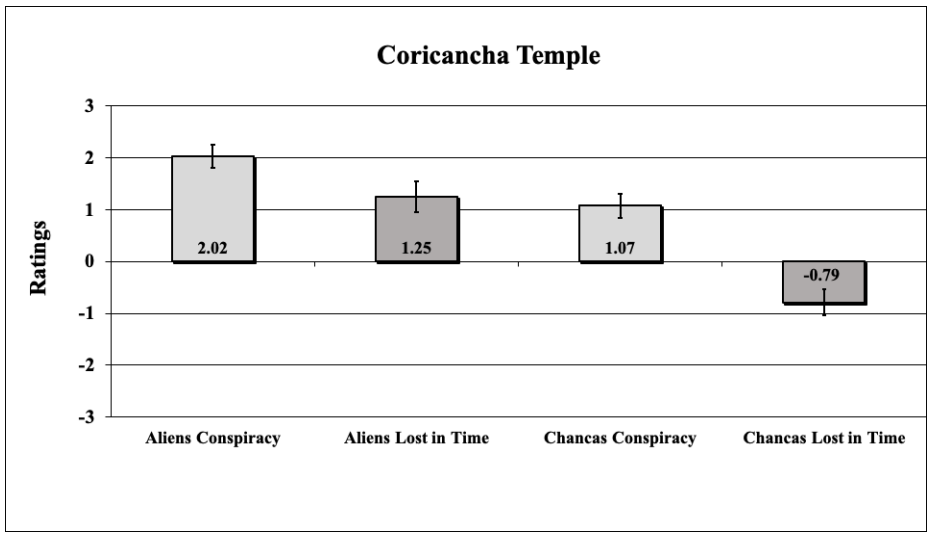

Figure 1.2: Mean values of the ratings in Study 2a. Error bars indicate standard error around the means.

Anna: This is a conspiracy theory.

Participants were then asked how fitting it is that Anna used the expression 'conspiracy theory', and rated the fittingness on a 7-point Likert scale anchored at '-3' meaning 'not at all fitting' and '3' meaning 'Absolutely fitting'. 161 participants (66 male, 91 female, 2 non-binary; $M_{age} = 34.19$) were recruited on Prolific Academic and were paid a small fee for taking part in the experiment.

The average ratings of the four vignette are depicted in Figure 2. We conducted a 2 x 2 ANOVA with *Ratings* as the dependent variable and two independent variables: *Creatures* (Aliens, Chancas) and *Condition* (Conspiracy, Lost in Time). Both these factors were significant, *Creatures*, $F(1, 156) = 35.78$, $p < 0.001$, and *Condition*, $F(1, 156) = 27.82$, $p < 0.001$. There was also a small interaction between those two factors, $F(1, 156) = 4.74$, $p = 0.031$. Furthermore, mean ratings for all four conditions were significantly different from the midpoint of '0'.

if they had built the temple themselves."

The results show that in the condition *Aliens Lost in Time*, people are willing to call an explanation a conspiracy theory ($M_{av} = 1.25$) despite the fact that no conspiracy had taken place, but "everyone knew about it". When the explanation instead featured a more plausible story *Chancas Lost in Time*, participants did not endorse the claim that the explanation was a conspiracy theory. Now, admittedly, whether or not the Incas conspired, did play a significant role in boosting the average responses. Thus, we should note that people were more likely to call an explanation a conspiracy theory if a conspiracy had taken place, ceteris paribus. Nonetheless, the results do indicate that conspiracy theories do not need to include a conspiracy.[13]

Study 2b: Conspiracy without Conspiracy Theory

In Study 2b, we intended to investigate the reverse claim, namely, whether people would be reluctant to call a claim a conspiracy theory, even if it featured a conspiracy. To this end, we used a variant of the Watergate scandal, but set in Argentina in the 1980s. We hypothesized that whether or not people call a claim a conspiracy theory was less dependent on whether the claim included a conspiracy, but rather whether the claim was true or false. We therefore manipulated two factors: The truth condition of the claim at stake was varied between true and false. We also suspected that the officialness of the story could have an important effect. Thus, we also manipulated whether or not the wide majority of Argentinians today believed the claim to be true.

163 participants were recruited on Prolific (57 male, 105 female, 1 non-binary; $M_{age} =$

[13]We would like to thank a reviewer for this journal for highlighting two concerns with this experiment. First, the story line resembles conspiracy theories we hear about the mythical island Atlantis. Second, participants might have inferred that Luke was told about the alternative explanation by people who might intend to cover-up the actual evidence. Both aspects could have led participants to make an implicit inference that there was indeed some conspiracy going on. While we have taken reasonable precaution to avoid such implications, e.g., by having Luke say that"I *believe* that ..." as if it was his own reasoning that got him to this 'insight' about the Incas, we cannot fully rule out that these aspects have had an effect on our participants.

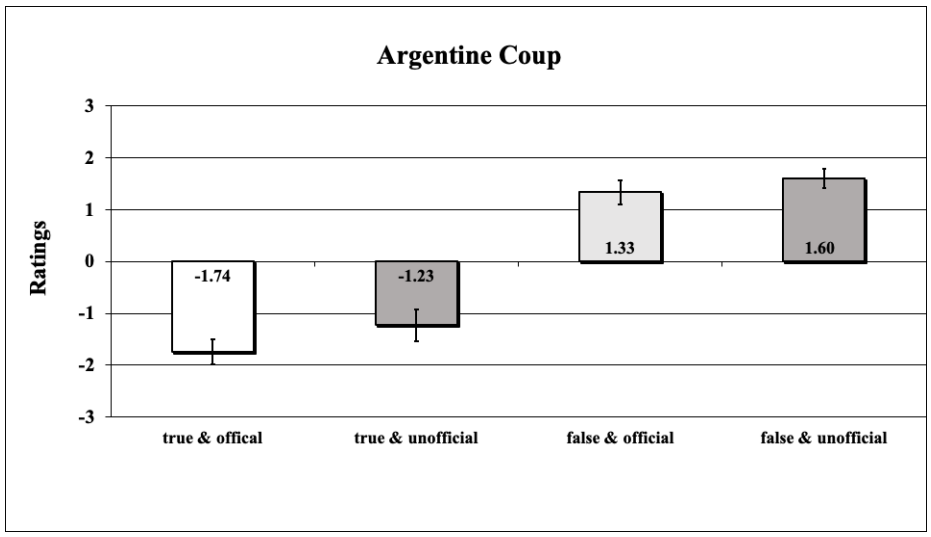

Figure 1.3: Mean values of the ratings in Study 2b. Error bars indicate standard error around the means.

34.54). They were then randomly assigned to one of four conditions: *true & official*, *true & unofficial*, *false & official*, *false & unofficial*. The vignettes read as follows:

> In the 1980s in Argentina, some people broke into the building of a political party and stole documents. Soon, the following claim was made:
>
> Claim: The president himself had ordered this crime in order to gain advantage over his opponents in the upcoming election campaign process and tried to cover this up.
>
> Here are two facts about that case.
>
> 1. The claim is true [false]. (the president ordered [did not order] the crime and tried to cover up his involvement.)
> 2. Today the wide majority of Argentinians believe the claim to be true [false].

All participants were then asked to answer the question 'Would you call the claim a conspiracy theory?' on a 7-point Likert scale ranging from '-3' labelled as 'Definitely not' to '3' labelled as 'Definitely yes'. The average results for all four conditions are displayed in

Figure 3. We conducted a 2 x 2 ANOVA with *Ratings* as the dependent variable and *Truth* (true, false) and *Officialness* (official, unofficial) as independent variables. *Truth* was a significant factor, $F(1, 159) = 134.86$, $p < 0.001$., whereas *Officialness* was not, $F(1, 159) = 6.10$, $p = 0.130$. t-tests revealed that the results of both *true & official* and *true & unofficial* were significantly below the midpoints, and both *false & official* and *false & unofficial* were significantly above the midpoint of '0' (all ps < 0.001).

In both conditions, in which the claim at stake was true, the majority of the participants would not regard the claim a conspiracy theory despite the claim involving a conspiracy at the highest level. Only when the claim was false, did people positively state that this was a conspiracy theory. Whether or not the claim was officially accepted played no significant role, although a greater number of participants might have resulted in a significant effect. Why would people call a claim a conspiracy theory if the claim was false? Presumably, it is not the truth or falsity that matters, but rather that a claim that is false does not have the right epistemic virtues, e.g., no facts, lack of evidence, far-fetched. At least this would be in accordance with the results of Studies 1a-c. Finding out more about the role these factors play, is an exciting avenue for future research.

Summary of the Results

In Study 2a, we tested a specific account on the descriptive meaning of 'conspiracy theory'. According to many theorists, conspiracy theories satisfy the conspiracy criterion, i.e., the explanation at stake features a conspiracy. The majority of the participants considered it appropriate to apply the term 'conspiracy theory' to an explanation that features no conspiracy. The outcome of Study 2b revealed that the reverse claim also holds: Even if an explanation clearly includes a conspiracy, it is not considered to be a conspiracy theory, if the claim is true. In other words, our results in Section 3.2 suggest that conspiracy and conspiracy theory can be doubly dissociated.

1.4 General Discussion

In the empirical part of this paper, we examined the content of the ordinary meaning of 'conspiracy theory'. Our main goal was to find out whether this concept is evaluative or predominantly descriptive. In order to investigate this issue, we used five studies using three different empirical methods:

The results of all five studies reveal a rather uniform picture, according to which the ordinary meaning of the expression 'conspiracy theory' is predominantly evaluative. This is not to say that the term 'conspiracy theory' is not ambiguous: around one-third of the population seem to entertain a notion of conspiracy theory that is primarily descriptive.

Importantly, the predominant evaluative meaning of 'conspiracy theory' is not just an add-on on top of some descriptive content. Instead, the results of Studies 2a & 2b suggest that the evaluative meaning is largely independent of some of the widely-discussed descriptive criteria. The independence of both usages provides further evidence that the evaluative content of the term 'conspiracy theory' is not conveyed pragmatically but is part of the meaning of the term.[14]

Of course, the results of our studies cannot be easily generalized to other cultures, languages, and other times. In fact, we consider it quite likely that the meaning of the term 'conspiracy theory' has changed substantially during the last few decades. The frequency with which the term 'conspiracy theory' is used has risen sharply since the 1980s and continues to rise (see the Google Books Ngram of 'conspiracy theory' here). A cross-temporal analysis of how the term 'conspiracy theory' might have changed during the last few decades, is thus highly desirable.

[14]This result is in line with the history of the expression 'conspiracy theory' defended by McKenzie-McHargh (2018), according to whom two different concepts—a descriptive one, and an evaluative one—emerged in the Nineteenth and Twentieth century. However, the alternative possibility that an evaluative concept emerged from the descriptive one is not ruled out.

In the remaining two sections of this paper, we discuss how best to interpret the data we collected in the empirical studies and the consequences for philosophical theorizing about conspiracy theories. We first provide an analysis of the evaluative concept *conspiracy theory* using the framework of thick concepts. We then subsequently examine the chances for various ways of defining and engineering the concept of *conspiracy theory*.

1.4.1 The evaluative meaning of *conspiracy theory*

The results of five studies revealed that the dominant meaning of 'conspiracy theory' is not purely descriptive. Instead, we have presented empirical evidence that evaluative content is encoded as part of the concept *conspiracy theory*. But what exactly then is the meaning of the term 'conspiracy theory'? The simplest suggestion would be to hold that an equivalent expression to 'conspiracy theory' is 'bad theory' or perhaps 'bad explanation'. Accordingly, in the context of talking about a theory or an explanation, we merely communicate our disapproval of the theory by calling the explanation a 'conspiracy theory'. While this suggestion is in line with the results of Study 1b, it is too simple a suggestion and cannot be squared with the results of the other studies. Our results do indeed indicate that people believe conspiracy theories to be bad theories, they do not, however, show that people think that all bad theories are properly called conspiracy theories: Conspiracy theories are bad in a more specific respect. To get a better picture of what this amounts to, let us briefly highlight some differences among evaluative concepts that have been discussed in the literature.

Evaluative concepts are usefully divided into thin concepts, thick concepts, and dual character concepts. Thin evaluative concepts like *awesome* and *bad* evaluate without specifying the way in which the target is evaluated. E.g., if Helen says that Tom's behavior is bad, you do not know (without further information) whether it is bad because he lied, because he was risking lives, because he was only in for his own advantage, etc.

Thick concepts work differently: concepts like *crazy*, *beautiful*, and *unjustified* have both evaluative and descriptive content. Their use evaluates positively or negatively, but also specifies the content that is under evaluation. E.g., if Helen states that Tom's behavior is crazy, then we can infer more than that Helen thinks of Tom's behavior as bad. She tells us that what is bad about his behavior is that his behavior makes little sense in one way or another. If she were to say that Tom's behavior is generous, then she would tell us that Tom was giving more than expected *and* that his behavior is good in virtue of giving more than expected.[15]

Given the existence and rich stock of thick evaluative concepts in our languages, a more sophisticated proposal would be to say that the composite term 'conspiracy theory' has a thick evaluative meaning. In other words, people not only express their disapproval of a theory (or an explanation) by calling it 'conspiracy theory', they also communicate the descriptive aspect under which it is considered to be bad. Fortunately, the results of Studies 1a and 1c allow us to draw some conclusions what the descriptive aspect might be that is evaluated negatively. When we asked people to tell us the necessary conditions for something to be a conspiracy theory, they did not simply note down 'bad', 'terrible' or other thin evaluative terms. Instead, many participants wrote down thick evaluative terms. Those evaluative responses were categorized into two groups: those attributes that indicate epistemic deficiency, and those indicative of a more general pejorative attitude. Very common responses in the first group were 'unjustified', 'false', and 'non-scientific'. The second group consisted of a wide variety of terms. On the one hand, terms like 'obsessed' and 'evil' express moral disapproval. Other terms like 'insane', 'crazy' and 'ridiculous' are not moral terms, but are derogatory terms that are most likely produced because of the epistemic deficiencies of conspiracy theories. Such an interpretation is further supported by Study 1c, in which adjectives were recorded that occur directly before 'conspiracy theory'.

[15]Dual character concepts are similar to thick concepts in that they encode both descriptive as well as evaluative information. In contrast to thick concepts, however, the evaluative and descriptive dimensions are doubly dissociatable. For a more detailed exposition of dual character concepts, see Reuter 2019.

Let us start with the first set of words expressing epistemic deficiency without derogatory meaning: A plausible interpretation of the high frequency with which terms like 'not provable', 'unjustified', and 'far-fetched' were mentioned, is that a conspiracy theory is considered a theory that has insufficient epistemic justification, is not in the business of being epistemically justified, or is epistemically deficient in some other way. Consequently, an equivalent expression to 'conspiracy theory' is not just 'bad theory' but rather 'epistemically deficient theory'. Of course, ordinary people will hardly use these two expressions interchangeably, but the proposal seems to aptly summarize that people not only consider conspiracy theories bad, but provide some descriptive content under which they are considered to be bad.

The second set of words goes beyond mere epistemic deficiency. Several participants provided responses that communicate strong disapproval, sometimes of a moral type. The majority of pejorative terms were of a non-moral sort, like 'crazy' and 'ridiculous'. In fact, these terms were not only frequent in the semantic feature production task, the corpus analysis revealed a frequent and strong co-occurrence with 'conspiracy theory'. These terms are often used pejoratively in an epistemic sense. Importantly, epistemic deficiency does not automatically warrant a pejorative attitude in all cases. Many theories are unproven or unprovable (string theory), or unjustified (phlogiston theory). We suspect though that very few people would go so far in calling them crazy or ridiculous. In other words, we can distinguish theories that are (at least in some sense) bad because they are epistemically deficient (including string theory), and theories that are epistemically bad *and* likely to be disparaged.

What might these further aspects be that often trigger a derogatory attitude? A plausible suggestion is that many people consider conspiracy theories to be crazy and ridiculous, because they satisfy some additional criterion, like having no epistemic justification or resist disconfirmation in light of any counter-evidence (a criterion I propose in Chapter 2). What exactly those criteria are is an interesting and open question. It is very likely that these

criteria are not fixed but vary from person to person. Unfortunately, a more precise answer goes beyond what is inside the data we collected.

1.4.2 Towards a definition of *conspiracy theory*

The results of the empirical studies allow us to draw some conclusions regarding not only what the ordinary meaning of 'conspiracy theory' is, but also how *conspiracy theory* ought to be defined.

Our studies suggest that the ordinary language term has a predominant evaluative component. If, as Räikkä (2018) argued, philosophers should approximate the dominant ordinary language meaning of 'conspiracy theory' as much as possible in order to foster an academic and political discussion of conspiracy theories which maintains the same subject as ordinary language, *conspiracy theory* should incorporate evaluative features, and it should not incorporate a conspiracy criterion as a necessary condition. If a proposed descriptive conceptual analysis of *conspiracy theory* left out an evaluative aspect, then it would not be a satisfactory analysis of the predominant concept of *conspiracy theory*.

Engineering *conspiracy theory*

The results from our studies have consequences for determining whether descriptive conceptual analysis is indeed the best method to achieve the goals that Räikkä has in mind for fostering a rigorous theoretical and institutional discussion of conspiracy theories which is close to everyday language. The range of evaluations encoded in *conspiracy theory* and the double dissociation of conspiracy and conspiracy theory might constitute a reason against sticking to the ordinary meaning of the term—at least for a scientific or institutional discussion of the problem of conspiracy theories. While belief in bad theories is something that

academics and institutions might be interested in understanding and minimizing, it seems to be way too broad to constitute the object of targeted scientific investigation and public intervention. This is a reason to advocate for engineering the evaluative *conspiracy theory* into a sharper concept, rather than employing descriptive conceptual analysis.[16]

In addition to recommending engineering as the best method for promoting a rigorous academic and institutional discussion of the phenomenon of people believing absurd theories, the empirical results of our studies have implications for the existing ameliorative proposals which were reviewed in section 2. Those proposals were driven by the worry that the current use of 'conspiracy theory' might lead to the silencing of warranted investigations of conspiracies, and might thus allow powerful people to exploit semantic defects of the concept *conspiracy theory* at their advantage. Some results of our empirical studies seem to speak against this worry: the evaluative concept *conspiracy theory* does not seem to apply to all theories about conspiracies, and attributions of 'conspiracy theory' seem to be driven by assessments of the theory.[17] However, our studies also indicate that the concept is ambiguous. Even though only a minority of the participants in Study 1a indicated only descriptive features, this result is not irrelevant. It suggests that some people possess a purely descriptive concept of *conspiracy theory*. Moreover, our studies show that there is not a core

[16]The method of explication, famously described by Carnap (1950), is especially suited to this task, as it aims at promoting the exactness and fruitfulness of a concept while maintaining similarity with its ordinary meaning. For a discussion of Carnapian explication see Maher 2007; Carus 2008; Justus 2012; Brun 2016.

[17]While our results seem to be in tension with the findings of Wood (2016) about the lack of a negative effect of labeling a theory 'conspiracy theory', there are reasons to think that they are not. Wood's experiments indeed found 'no evidence of a negative effect of calling something a conspiracy theory' (Wood 2016: 702). However, two things should be noted: In Wood's main Study 2, the two conditions differed in whether the claims were labelled as 'conspiracy theory' and 'corruption allegation'. As allegations are also presented without any proofs, perhaps a more neutral label would have resulted in substantial differences. Second, and more importantly, the story the participants read, featured information about the "latest development in a growing political scandal" as well as a "Canadian political watchdog group calling for an investigation". Hence, participants are given information that strongly suggest that the claim under investigation is not as outrageous and far-fetched as many other claims that run under the label 'conspiracy theory'. So, they might disagree on labelling the content presented as a 'conspiracy theory', and they might be assessing the plausibility and likelihood of the conspiracy allegations independently of whether they are labeled 'conspiracy theory'. A similar explanation could be given for Wood's Study 1. While this might explain away the apparent tension between our results and Wood's (2016), more studies should be conducted to investigate under which circumstances labelling a claim a 'conspiracy theory' has an effect and under which is has no such effect.

concept *conspiracy theory* that is then supplemented with a more or less negative attitude. The descriptive and evaluative concept *conspiracy theory* seem to be distinct concepts—the evaluative concept does not include the core feature of the descriptive concept, which is involving a conspiracy. This ambiguity might lead to equivocations and unwarranted inferences when reasoning about conspiracy theories, and it would motivate a different kind of conceptual engineering approach, as amelioration aimed at promoting better political and social environments, and not just at improving theoretical discussions of conspiracy theories.

While our studies lend support to the claim made by Coady, Basham, and Dentith that *conspiracy theory* should be ameliorated, they also put pressure on their specific proposal of eliminating the evaluative concept of *conspiracy theory* altogether, even granting that in conceptual amelioration the engineered concept need not be necessarily similar to the ordinary one in meaning—amelioration opting instead for the continuity with the ordinary concept's function or functions in our practices and discourses (Haslanger 2012:224-225; Dutilh Novaes 2020). The predominance of the evaluative concept *conspiracy theory*, and its independence from the descriptive one show that the choice of some academics to focus on conspiracy theories as a problem, and the attempt to explain belief in such theories in terms of psychological or sociological factors is not necessarily a choice driven by hidden political motives to silence the investigation of conspiracies in our societies. The evaluative concept *conspiracy theory* is prevalent in ordinary thought and language, and attributions of 'conspiracy theory' seem to be driven by an assessment of the target theory, even in the absence of a conspiracy component. Thus, the function that this concept serves in academic practices and discourses cannot be silencing warranted conspiracy accusations.[18] More likely,

[18]Cf. Basham & Dentith (2016); Dentith (2018b, 2018c); Basham (2018a), Hagen (2018), Orr & Dentith (2018). One possible way to object to our analysis would be to grant that the derogatory use of *conspiracy theory* is not the exclusive use of powerful elites of academics and politicians, but claim that the derogatory use of *conspiracy theory* is the result of the repeated efforts by these groups to derogate theories about conspiracies by employing the label 'conspiracy theory'. However, our studies lend support to the claim that the evaluative component of *conspiracy theory* is a feature of the meaning of the expression, rather than a pragmatic implication of its use. So, even if the origin of the meaning of *conspiracy theory* was related to the manipulative intention of academics and politicians, this would not change the fact that the meaning of 'conspiracy theory', nowadays, is predominantly evaluative, it does not apply to all theories

the main function of the evaluative concept is to single out a phenomenon of irrational belief in certain absurd theories. While an ambiguous use of 'conspiracy theory' might lead to equivocations and silencing, and should therefore be corrected for, the evaluative concept alone does not seem to serve this function and the complete elimination of the evaluative *conspiracy theory* would interfere with the progress of outlining and understanding a phenomenon which has attracted the interest of many scholars and institutions.

Sketching a solution

One of the findings of our study suggests a way forward. We have observed that the evaluative concept *conspiracy theory* encodes a wide range of evaluations, which we classified in two categories: epistemic evaluations and derogatory attitudes. Our proposal is to reserve the label 'conspiracy theory' to refer only to the epistemically evaluative *conspiracy theory*—while introducing a new expression, such as 'conspiratorial explanation', to refer to the descriptive 'theories which involve conspiracies'.

The more pejorative sense, expressed by terms like 'crazy' and 'insane', encodes highly derogatory information and contains little descriptive content. Since the rules that govern the application of this pejorative label are highly subjective, and of difficult evaluation, this sense of *conspiracy theory* should be eliminated for promoting better discussions of the topic of conspiracy theories. On the other hand, epistemic evaluations refer to inter-subjective standards, and are richer in descriptive content. Once appropriately characterized and specified, the epistemic criteria constitutive of *conspiracy theory*, could be employed to assess different theories and determine the correct application of the concept. The more the criteria that identify the epistemically evaluative *conspiracy theory* are clearly specified, the less the label 'conspiracy theory' could be exploited to dismiss theories which do not display

about conspiracies, and is driven by evaluations of the target theory. Accordingly, the function of the concept would have changed over time.

the epistemic defects necessary for correctly applying the concept.

Moreover, if the epistemic flaws identified as constitutive of *conspiracy theory* implied the irrationality of believing conspiracy theories, and not just a minor epistemic deficiency, the epistemically evaluative concept would maintain its current derogatory usage. In academic and institutional discussions, it would enable those approaches that look at the non-epistemic reasons for believing conspiracy theories—such as psychological or sociological reasons—and the institutional attempts at minimizing belief in conspiracy theories. If *conspiracy theory* was employed to refer to a certain type of irrational theories, this would justify explaining beliefs in these theories by appeal to non-epistemic factors, and trying to minimize such beliefs. And if the epistemic defects were appropriately described, these scholars could more easily identify the theories on which to focus in their investigations (cf. Cohnitz 2018:358).[19]

A possible concern with our engineering strategy is that epistemically evaluative definitions would create more variance in their applications than descriptive ones, and this would be a disadvantage of these definitions. If *conspiracy theory* incorporated an epistemic evaluation, one might worry that whether a theory counts as a 'conspiracy theory' would depend on who believes it and their epistemic grounds. The first thing to notice about this, is that not all epistemically evaluative proposals need necessarily be subject-dependent. Theories, i.e., sets of propositions, too can have epistemic deficiencies—for instance, they can be false, or inconsistent, or contrary to what independently identified experts believe, and so on. So, while some specific proposals might have the consequence that a theory is a *conspiracy theory* depending on who believes it (a consequence of the view I propose in Chapter 2), not all epistemically evaluative proposals are subject to this worry. Secondly, the increase in variance in attributions of *conspiracy theory* is not necessarily a disadvantage of a proposal, and might in fact be an advantage if one is interested—as many scholars and most institutions

[19]Admittedly, the strategy we suggest is more suited for technical discussions of the topic than it is for ordinary usage. However, the ordinary evaluative concept *conspiracy theory* which encodes both epistemic evaluations and more pejorative attitudes might not be problematic in ordinary language, in the absence of an ambiguous descriptive concept.

seem to be—in understanding conspiracy theories as a phenomenon of irrational belief.

The introduction of a different expression like *conspiratorial explanation* to refer to the descriptive concept eliminates the ambiguity which generated the risk of equivocating theories about conspiracies and bad theories.[20] Moreover, the descriptive *conspiratorial explanation*—rather than a descriptive engineering of *conspiracy theory*—could be employed for those philosophical projects which necessitate a neutral definition, such as comparing explanations that involve conspiracies to other types of explanations, which have typically been the object of investigation in philosophy of science (Keeley 2019); or investigating the rationality of theories which involve conspiracies (Dentith 2014, 2018c). The descriptive 'conspiratorial explanation' could serve the same aim which motivated theoretical engineering projects such as Dentith's, while maintaining the evaluative sense of 'conspiracy theory' to preserve talk of conspiracy theories in the negative sense. While the use of the neutral 'conspiratorial explanation' alongside the evaluative 'conspiracy theory' might seem to complicate the academic debate on conspiracy theories, it would actually make clear that different debates about conspiracy theories have been talking about two different things all along.[21]

How exactly the epistemically evaluative *conspiracy theory* should be engineered is to be determined through philosophical arguments, and through an assessment of the effects that different engineered concepts would have on both our theorizing about conspiracy theories, and our political and social treatment of warranted conspiracy accusations.[22] It is possible

[20] Another possible proposal for a neutral label, albeit a narrower one, comes from deHaven-Smith (2006).

[21] A related issue is whether our proposal of adopting two expressions to refer to the descriptive and the evaluative concept gives rise to worries regarding implementation. One specific worry might be that the descriptive 'conspiratorial explanation' is a rather technical expression, unlikely to be picked up in ordinary language. However, we believe that our proposal is primarily targeted at academic discussions, rather than at society at large, where the implementation of a new concept like this should be easier. Moreover, implementation is a serious concern for any engineering proposal, not just the one discussed here. For a discussion of the implementation challenge, and possible solutions see Cappelen & Plunkett (2020); Andow (2021); Koch (2020).

[22] I take the engineering proposal offered in Chapter 2 to follow this strategy. Cassam (2019) can also be interpreted as a proposal along the same lines: his account of Conspiracy Theories is epistemically evaluative and suited to investigations of conspiracy theories as instruments of political propaganda.

that different engineered concepts will be best employed in different projects, and not a single definition will serve both the theoretical and the societal aims equally well. Our proposal regards only the direction which should be taken by philosophers trying to define *conspiracy theory* in a way that both promotes the scientific investigations of conspiracy theories and the public discussion of how to deal with them, and at the same time avoids possible negative effects on our political environments.

1.5 Conclusion

A central question in the philosophical discussion about conspiracy theories focuses on what the meaning of the expression 'conspiracy theory' is and what it should be. In this paper, we addressed this question by employing some of the methods of experimental philosophy. In five studies, we have shown that the assumption that *conspiracy theory* is primarily descriptive does not withstand empirical scrutiny. Moreover, our studies show that conspiracy and conspiracy theory are doubly dissociable: people sometimes do not attribute the label 'conspiracy theory' to theories which involve conspiracies, while they do sometimes attribute it to theories which do not involve any conspiracy. These results put pressure on the neutral definitions of 'conspiracy theory' which are currently widely endorsed in the philosophical debate—both those which are proposed as analyses and those which are proposed as ameliorations of the ordinary concept *conspiracy theory*. We argued that the best strategy for defining *conspiracy theory* is to engineer the evaluative concept to encode specific epistemic deficiencies.

Chapter 2

Conspiracy Theories and Evidential Self-Insulation

2.1 Introduction

The moon landing was faked. 9/11 was an inside job. Secret societies control the world. Immigration is a plan of the political elite aimed at extinguishing the white race. These are just a few examples of widely believed conspiracy theories (at least, more widely than one would have hoped). To most, conspiracy theories are wacky stories, the evidence for which is allegedly given in YouTube videos where eccentric characters point out long series of coincidences that the official accounts cannot account for. When we call these theories 'conspiracy theories,' we often use the term pejoratively to indicate theories that should not be believed, and perhaps should be met with ridicule. Similarly, the public debate about conspiracy theories assumes that conspiracy theories are fictions that undermine the trust required for the spread of knowledge in our societies, and that belief in such theories is inappropriate.

But what are conspiracy theories, exactly? And what is epistemically wrong with them? In this chapter, I offer a joint answer to these two questions that is based on two observations: (i) many explanations that involve conspiracies are not to be considered conspiracy theories, and (ii) whatever distinguishes conspiracy theories from mere theories that involve conspiracies makes the former epistemically problematic. Contrary to those who argue that conspiracy theories are just explanations of events that involve conspiracies,[1] I maintain that conspiracy theories are not theories (or explanations) at all.[2] Instead, I take 'conspiracy theory' to refer to a particular way of holding a belief in the existence of a conspiracy. The attitude of the believer, rather than any feature of the theory, determines whether a person's belief in a conspiracy is a conspiracy theory or not.

Here is a sketch of the account to come. There is an interesting feature that we observe in people who defend conspiracy theories. It seems to be the case that, no matter what evidence we present to them against their theory, they'll find a way to dismiss it. I take this to be a central characteristic of conspiracy theories; they give rise to this dismissive epistemic behavior. Some have argued that the resistance to disconfirming evidence is not, per se, a problematic feature of conspiracy theories (Keeley 1999; Dentith 2017; Harris 2018). The reason behind this claim seems to be that if a conspiracy is going on, the conspirators would be trying to cover it up. Hence, misleading counter-evidence is to be expected. The resistance to counter-evidence typical of conspiracy theorizing seems to be warranted by the kind of thing conspiracies are, namely, plots by a group of people who are trying to keep their intentions and actions secret. I will argue that the simple explanation of this feature of conspiracy theories is misleading. While it is true that belief in a conspiracy warrants a certain type of resistance to counter-evidence, I argue that the evidential insulation typical of conspiracy theories makes them epistemically problematic.

[1] For instance, Basham (2001); Buenting & Taylor (2010); Dentith (2014); Harris (2018); Keeley (1999); Pigden (1995); Räikkä (2009).

[2] In line with the literature on conspiracy theories, I use 'theory' and 'explanation' as synonyms, despite the obvious differences between the two.

I begin in §2 with a discussion of the methodology employed in the conspiracy theory debate, and I motivate the need for a negatively loaded conception of conspiracy theories that tracks the same phenomenon as the ordinary expression 'conspiracy theory.' In §3, I present my account of conspiracy theory as a self-insulated belief in the existence of a conspiracy. In §4, I argue that conspiracy theories so understood are epistemically irrational. In §5, I address three objections to my view.

2.2 Conspiracy Theories and Philosophical Methodology

First, a word about the methodology in the discussion ahead. Typically, when giving an account of conspiracy theories, the first step is to provide a definition of 'conspiracy theory.' But what are we doing when defining the expression 'conspiracy theory'? And what constraints should we have in mind? The kind of definition I am after is aimed at revising the ordinary expression of 'conspiracy theory' in order to help advance the understanding of a phenomenon that has become the object of much academic and public discussion—the phenomenon of people believing absurd theories about conspiracies, and believing them to be the best explanations of the available evidence.[3] I am thinking of theories such as the fake moon landing, flat earth, or the Illuminati controlling the world. I will not discuss the rationality of any of these theories in particular, but I will assume that when we talk about conspiracy theories, we have in mind outlandish theories like these. However, our natural language intuitions about conspiracy theories seem rather confused. It is not clear what people mean by 'conspiracy theory,' and what exactly makes them theories that should not be believed. My account looks to maintain the epistemically negative connotation that characterizes the current meaning of 'conspiracy theory,' while making this expression clear, more precise, and suited to be employed in empirical studies of the phenomenon of conspiracy

[3] While the perception and discussion of the phenomenon of conspiracy theories seem to have become more prominent in recent years, empirical data suggests that the phenomenon itself has not. See, for instance, Van Prooijen & Douglas (2017). I thank an anonymous referee for pointing this out to me.

theorizing.[4]

Even though explicit mentions of philosophical methodology are quite rare in the debate, there seems to be a trend in the philosophical literature about conspiracy theories to adopt a revisionary definition of conspiracy theories as any theory that involves a conspiracy.[5] While it is commonly recognized that 'conspiracy theory' is ordinarily used to indicate a special type of theories about conspiracies, and that it is a negatively loaded expression, most philosophers working on the topic agree that 'conspiracy theory' should be defined as any explanation of an event that cites a conspiracy.[6] One reason that is often cited in favor of the broad, neutral definition is the practical consequences of the ordinary meaning of the expression. Some philosophers argue that, by allowing 'conspiracy theory' to be a pejorative expression, we help powerful people get away with their conspiracies. 'Conspiracy theory' can be (and often is) used as a negative label to dismiss charges of genuine conspiracies. In order to avoid dismissing real conspiracies due to this, they argue, we should stop attaching a negative value to the expression. Hence, they conclude, the meaning of 'conspiracy theory' should be re-engineered to mean any theory about a conspiracy, and it should not have a negative valence.[7]

However, by assuming that every theory involving a conspiracy is a conspiracy theory,

[4] For an empirical study regarding the negative meaning of the expression 'conspiracy theory' and a discussion of its consequences for the conceptual engineering of conspiracy theory, refer back to Chapter 1.

[5] One person who does discuss the methodology of giving an account of conspiracy theories is David Coady (2018a). He argues that, given the ambiguous use of the expression and the reasoning fallacies it produces, we should abstain from ever using it.

[6] For an in-depth discussion of this definition, see Dentith (2014).

[7] For instance, see Basham & Dentith (2016); Coady (2012;2018b). While practical concerns are the most discussed in the literature, other reasons for the minimal re-engineering have been proposed. For instance, it has been suggested that the ordinary concept is ambiguous and leads to fallacious reasoning (Coady 2018a). This assumption is discussed in the first chapter. Moreover, it has been suggested that focusing on a neutral and minimal definition of 'conspiracy theory' is necessary in order to avoid begging the question whether it is ever rational to believe conspiratorial explanations, and what the difference is between this explanation type as opposed to other types, more discussed in philosophy of science. Investigating the epistemic status of conspiratorial explanations could be a worthwhile philosophical project, and a minimal account of conspiracy theory might be the best revisionary account for this goal. However, I take it that what we're interested in as a public and as a research community is not this goal, but rather, we want to understand and address resilient beliefs in wild conspiracies.

these philosophers seem to have changed the meaning of 'conspiracy theory' in a way that is neither warranted nor fruitful. It is unwarranted because their claim that attributing the negative label 'conspiracy theory' to a theory might be employed to dismiss actual conspiracies has not been confirmed by empirical data—in fact, some empirical research suggests that labeling a theory a 'conspiracy theory' does not reduce belief in that theory (Wood 2016). Even granting that their worry is well founded and that a negatively loaded definition of conspiracy theory could help powerful conspirators get away with their conspiracies, this worry only applies to negatively loaded definitions that are *broad*, i.e., that consider all theories about conspiracies to be conspiracy theories. If every theory involving a conspiracy was negatively labeled as a negative 'conspiracy theory,' then any theory involving a conspiracy would run the risk of being erroneously dismissed. On the contrary, *narrow* definitions which allow for the semantic possibility of theories involving conspiracies that are not conspiracy theories, do not fall prey to the same pragmatic concern. The narrow, negatively loaded expression 'conspiracy theory' does not warrant the dismissal of just any theory involving a conspiracy. Moreover, adopting a broad, neutral definition is not fruitful because it does not allow for studying conspiracy theories as the phenomenon I described at the beginning of this section. Many psychologists, cognitive scientists, and social scientists who have investigated the topic of conspiracy theories have typically focused on conspiracy theories as a problem to be addressed, or as an instance of irrational behavior. The broad account has given rise to several instances of tension and misunderstanding with scholars from those other fields. Some defenders of the broad conception of conspiracy theories have harshly criticized researchers with different approaches to the topic for their negative attitude towards conspiracy theories and for 'pathologizing' belief in such theories, thus creating a hostile intellectual climate where different research projects on conspiracy theories seem to be talking past each other.[8]

[8] See, for instance, the exchange between Basham & Dentith (2016) and Dieguez et al. (2016). Other examples are Basham (2018); Coady (2018b); Hagen (2018); Orr & Dentith (2018).

I believe that the best revisionary definition of 'conspiracy theory' is going to be narrow and negatively loaded, where the narrowing factor specifies and explains the irrationality of conspiracy theories. Such a definition allows us to investigate conspiracy theorizing as a phenomenon that seems to have become increasingly common in recent years, and it enjoys some important advantages over its broad rival. This methodological digression has two important upshots. First, the account I propose seeks to capture what we have in mind when we talk about conspiracy theories in ordinary language, i.e., the phenomenon of people believing outlandish theories about conspiracies in a way that seems to resist falsification. Second, my account is still an instance of conceptual re-engineering for theoretical fruitfulness. Hence, a failure to completely match our intuitions about what conspiracy theories are should not be considered a reason to reject it.

2.3 Conspiracy Theories

It is commonly assumed that conspiracy theories are, at the very least, theories that involve conspiracies.[9] I will challenge this assumption. I maintain that being a *theory* is not even a necessary feature of conspiracy theories, but rather that conspiracy theories are a way of holding a conspiratorial belief. Anyone who has ever met a conspiracy theorist will be familiar with the frustrating experience of trying to debunk the relevant belief. No matter what evidence we present to the conspiracy theorist, their confidence seems to remain intact. Evidence that seems to contradict the conspiratorial belief is likely to be seen by the believer as evidence that has been planted as part of the cover-up. I take this to be the core feature of conspiracy theories. Belief in such theories seems to be completely immune to counter-evidence. In this section, I argue that we identify conspiracy theories with a distinctive way

[9] See Basham (2001, 2003); Buenting & Taylor (2010); Dentith (2014); Keeley (1999); Pigden (1995); Räikkä (2009, 2014). Sometimes the minimal definition is supplemented by an additional feature that theories about conspiracies need to have in order to count as conspiracy theories. For instance, Coady (2012) and Feldman (2011) add that the conspiratorial explanation should be unofficial.

of holding the belief in the existence of a conspiracy, namely, one that is *self-insulated*.

Roughly, we can say that conspiracy theories are conspiracy-beliefs (beliefs in the existence of a conspiracy) that are self-insulated. Both parts of this account require clarification. I take a conspiracy to be the plotting by a group of actors—the conspirators—to achieve a goal in their interest, while trying to keep their intentions hidden.[10] Accordingly, a conspiracy-belief is a belief that a certain conspiracy has happened in the past or is currently going on. Conspiracy-beliefs are interesting from an epistemological point of view. Believing that a conspiracy is behind a certain event or fact entails believing that the conspirators have likely planted evidence against the conspiracy to mislead us. In their attempt to keep their actions and intentions secret, conspirators try to orchestrate cover-ups, disseminate misleading evidence, and promote alternative narratives for the public to believe. Hence, believing that a conspiracy is going on entails believing that things are not as they seem, i.e., that what seems like disconfirming evidence should not be taken to actually speak against the existence of a conspiracy. It follows from what conspiracies are that conspiracy-beliefs will screen off parts of the relevant evidence, because if a conspiracy is going on, someone is trying to make us believe otherwise.

It is part of what conspiracies are that the evidence against them could be the result of the conspirators' attempt to stage a cover-up. However, this does not mean that conspiracy-beliefs are *always* immune to revision. Conspiracies may render part of the available evidence unusable while keeping other evidential relations intact. For instance, one may encounter contrary evidence that they had no reason to believe was tampered with by the conspirators. Or one may encounter defeaters for their reasons to believe in the existence of a conspiracy to begin with. For example, I might believe that most common diseases could be cured with acupuncture, but, due to a conspiracy of the pharmaceutical companies, evidence of

[10]While there tends to be general agreement on what conspiracies are, there has been some discussion regarding how powerful the conspirators must be, whether their goal has to be nefarious, and what role the secrecy should play. For a discussion of the definition of 'conspiracy,' see Dentith (2014); Orr & Dentith (2018).

this was hidden from the public. My conspiracy-belief could be shaken if, for instance, I discovered that the evidence I had to believe this did not come from a reliable source, or, say, if acupuncture failed to cure my flu. My conspiracy-belief would not be, on my definition, a conspiracy theory.

I submit that conspiracy theories are only those conspiracy-beliefs that are self-insulated. What I mean by 'self-insulated' is that the believers take the conspiracy to neutralize the relevant counter-evidence. No evidence could be presented to them that would cause them to change their minds, because any counter-evidence would be dismissed as a fabrication of the conspirators to steer the public away from the truth.[11] When I say that conspiracy theories are a distinctive way of holding a conspiracy-belief, I take 'conspiracy theory' to refer to an *attitude* of the believers, rather than to a type of explanation. However, the content of the belief is key. In a conspiracy theory, the conspiracy is what the believers take to justify their dismissive attitude towards the evidence, and what plays the role of immunizing one's conspiracy-belief. By defining conspiracy theories as a certain attitude, I take conspiracy theories to be essentially tied to the believers of the theories. The same explanation could be a conspiracy theory for one agent, and not for another, according to how each of them accommodates counter-evidence. Nevertheless, I still consider conspiracy theories a way of holding beliefs, rather than a derivative notion of an independently defined 'conspiracy theorist.' A conspiracy theorist, on my view, is a person who holds one or more self-insulated conspiracy-beliefs—one or more conspiracy theories.

One more clarification of self-insulation is necessary. A self-insulated belief in a conspiracy is a belief that is immune to being disconfirmed by counter-evidence. However, the counter-evidence that is relevant to determining whether the belief is self-insulated should be restricted to counter-evidence that the subject could encounter in normal circumstances. In other words, we could say that the evidence to which conspiracy theorists are insensitive

[11]This does not imply that, on my account, conspiracy theorists could never abandon their beliefs. They could, but, in a conspiracy theory, this would not be a transition based on the evidence.

is any evidence that they might encounter in nearby possible worlds. It is possible that a believer in a conspiracy theory might change their mind in far-fetched scenarios where they might encounter exceptional evidence, such as if they could travel to the past and observe the events, or if they received an omniscient oracle's testimony, or if they could read minds. In my view, whether these exceptional and exceptionally unusual pieces of evidence would lead someone to reduce their confidence in a conspiracy-belief is not relevant to whether or not a conspiracy-belief counts as being self-insulated in the target sense. A self-insulated belief is a belief that is immune to being disconfirmed by the kind of evidence that is available in normal circumstances. In the rest of the chapter, I will talk of self-insulation in this restricted sense.[12]

To summarize, a conspiracy theory is the belief in the existence of a conspiracy, where the existence of the conspiracy is taken to justify the dismissal of any seemingly disconfirming evidence that one could encounter under normal circumstances. Having defined conspiracy theories, in the next section I turn to the question of their epistemic status.

2.4 Are Conspiracy Theories Irrational?

On my account, conspiracy theories are beliefs in conspiracies that are resistant to revision in light of counter-evidence. In this section, I argue that, given the empirical nature of conspiracies, one can never be rational in holding a belief in a conspiracy that is self-insulated.[13] In other words, I argue that it is irrational to hold conspiracy theories.[14] Even though my

[12]I am grateful to Paul Silva for helpful discussion on this point.

[13]Evidential insulation, per se, need not be necessarily irrational. It could be argued that things such as mathematical proofs and necessary truths might be rationally believed in a way that resists revision. In this chapter, I only argue that evidential insulation is problematic for empirical beliefs, including beliefs in conspiracies, and I leave open whether evidential insulation is problematic for a priori beliefs. See Casullo (2003).

[14]It is certainly the case that, on my account, the epistemic status of conspiracy theories depends on the believer, rather than on the theory to which they subscribe. When I claim that conspiracy theories are irrational, this should not be confused with a claim about any theory, but it should be read as 'beliefs

account of conspiracy theories is significantly different from traditional accounts, the discussion in this section has substantial implications for those traditional accounts that have also claimed that the unfalsifiability of conspiracy theories does not make them irrational to believe.

The discussion over the epistemic status of conspiracy theories has traditionally focused on the question of whether it is ever rational to believe theories about conspiracies. Many have argued that it is sometimes rational because a conspiracy may be the best explanation of the evidence.[15] In the debate, the question of *revising* conspiratorial beliefs in light of new evidence has always been secondary to the question of *forming* belief in conspiracies. It is often assumed that the extreme resistance to counter-evidence is built into what conspiratorial explanations are, and that it is not an epistemically problematic feature:

> By invoking a conspiracy hypothesis, large amounts of "evidence" are thrown into question. This is one of the most curious features of these theories: to my knowledge, conspiracy theories [i.e., explanations involving conspiracies] are the only theories for which evidence against them is actually construed as evidence in favor of them. The more evidence piled up by the authorities in favor of a given theory, the more the conspiracy theorist points to how badly "They" must want us to believe the official story. (Keeley 1999: 120)

The thought is that, if one is epistemically justified in believing that a conspiracy is going

in conspiracies that resist revision in the way I described are irrational,' or better, 'an agent is irrational insofar as they hold a self-insulated conspiracy-belief.' Being rational or irrational is a property of the agent who holds a certain belief in a certain way. However, I am not making any claims about the believer as an epistemic agent in general. The focus is on individual beliefs and whether they are rationally held. This is the main difference between my account of conspiracy theory and accounts of what some have called conspiracism, i.e., the tendency of some theorists to believe in conspiracies without good reason (Dentith 2018b). Attributing conspiracism to believers runs the risk of suggesting a stable disposition of the believer to form this type of irrational belief. My account of conspiracy theories is an account of beliefs in conspiracies that are held irrationally, and not an account of the people who hold these beliefs. I thank an anonymous referee for pointing out this unclarity to me.

[15] See Basham (2001); Buenting & Taylor (2010); Coady (2012); Dentith (2014, 2017); Harris (2018); Keeley (1999); Pigden (1995); Räikkä (2009).

on, then one is epistemically justified in interpreting evidence against one's belief as an attempt by the conspirators to hide their plot. This argument has much intuitive appeal and has largely gone unchallenged. However, it is unclear to what extent the hypothesis of a conspiracy warrants the dismissal of disconfirming evidence. Keeley suggests that theories about conspiracies could potentially be immune to *any* evidence:

> The worry is that given a situation where all potentially falsifying evidence can be construed as supporting, or at worst as neutral evidence, then conspiracy theories are by definition unfalsifiable. In favor of conspiracy theorists, it should be noted that this unfalsifiability is not as ad hoc as it might initially seem, due to the active nature of the investigated, just noted. It is not ad hoc to suppose that false and misleading data will be thrown your way when one supposes that there is somebody out there actively throwing that data at you. (Keeley 1999: 121)

According to Keeley and those who have endorsed his argument, theories about conspiracies can be unfalsifiable, and this is not problematic because of the active nature of conspiracies.[16] On this view, holding an unfalsifiable conspiratorial explanation can be rationally permissible. Hence, proponents of the view take it that it is sometimes rationally permissible to hold the belief in the existence of a conspiracy that is immune to being disconfirmed. I spend the remainder of this section arguing against this claim. *Pace* Keeley, not all evidence against the conspiratorial explanation can be neutralized by the belief that the conspirators are staging a cover-up.

To make the point, I will rely on some insights from Bayesian epistemology. Bayesianism

[16]Basham (2001: 268, 2003: 93); Dentith (2017: 9); Harris (2018: 243–5). For Keeley, the conspiracy theory will be abandoned when the skepticism that is required in order to maintain the belief in the conspiracy becomes "more than we can stomach" (1999: 126). The resilience to counter-evidence is not a problem, per se, of conspiracy theories. However, in order to maintain the belief in the conspiracy, one would have to assume the involvement of more and more institutions and people until the amount of skepticism required is simply too much, and the belief in the conspiracy is abandoned.

gives us a theoretical framework to evaluate how relevant new evidence is to the conspiratorial hypothesis, given the background assumption that, if the conspiracy is going on, the conspirators are trying to keep their intentions and actions secret. The core features of the Bayesian model are (i) that the level of confidence in a hypothesis can be represented with a credence value varying from 1 to 0, where 1 corresponds to certainty in the truth of the hypothesis, 0 corresponds to certainty in its falsehood, and 0.5 to equal confidence in its truth and its falsity; (ii) that ideally rational agents have credences that can be modeled by probability functions; and (iii) that agents learn from new evidence by updating their credence using conditionalization.[17]

Using these terms, we can define a conspiracy theory as the belief in the existence of a conspiracy C such that the credence in the existence of a conspiracy P(C|E)=P(C), for any counter-evidence E that one might encounter in normal circumstances.[18] The Bayesian framework allows us to identify two conditions under which discovering new evidence will not have any disconfirming effect on a rational agent's belief: certainty and irrelevance.[19] Let's consider each of these.

First, one could be certain that there is a conspiracy. If one's credence in a hypothesis P(H)=1, then the conditional probability of the hypothesis on the evidence is P(H|E)=1, for any new evidence E that the agent may encounter. Let's consider the case in which h is a conspiratorial hypothesis, such as:

> **Con:** The Twin Towers fell as the result of a controlled demolition, intended by government officials.

[17]For an introduction to Bayesian confirmation theory, see Bovens & Hartmann (2003); Strevens (ms).

[18]It would still count as a conspiracy theory if the confidence in the existence of a conspiracy could only be brought down to a certain threshold but no lower. In that case, even though the conspiracy-belief would not be totally immune to revision in light of new evidence, it would still be immune to revision in the sense that it could never be fully disconfirmed by counter-evidence.

[19]Silva (2020) makes a similar point regarding the rationality of sexist and racist beliefs.

Imagine a believer who is certain of the truth of Con. Could her belief in Con be an instance of a rationally had conspiracy theory? First, I am inclined to say that conspiracy-beliefs that are immune to revision because of certainty would not count as *conspiracy theories* on my account. Conspiracy theories are beliefs that are insulated because the evidence is dismissed by appeal to the conspirators' attempt to hide the truth. If one were certain that the Twin Towers were demolished, then certainty, rather than the belief that the conspirators are trying to hide their plot, would guarantee the immunity to revision (any non-conspiratorial hypothesis would be equally immune to revision). Hence, it is not clear that certainty in the existence of a conspiracy would count as a conspiracy theory, and thus that it would constitute an instance of a rationally had conspiracy theory. Moreover, it is hard to see how one could rationally come to be *certain* of an empirical claim such as the existence of a conspiracy. Beliefs in the existence of secretive plots are not the kind of thing that one could rationally come to believe beyond doubt. Thus, the certainty condition can never justify the evidential insulation of conspiracy theories. So, let's move on to the second condition that could justify conspiracy theories' dismissal of disconfirming evidence: probabilistic irrelevance.

The irrelevance condition is the more interesting condition because it seems to be grounding Keeley's claim that, with conspiratorial beliefs, "all potentially falsifying evidence can be construed as supporting, or at worst as neutral evidence" (1999). Bayesian confirmation theory provides a quantitative method for assessing the impact of new evidence on hypotheses, based on the general principle that, if a particular observation is more likely given the truth of the hypothesis, than it is given its falsehood, then the observation is evidence in favor of the theory. An observation is probabilistically irrelevant to the hypothesis if it is assigned the same probability on the assumption that the hypothesis is true and that it is false. Keeley seems to be arguing that some conspiratorial explanations satisfy the irrelevance condition. Under the irrelevance condition, a belief in a conspiratorial hypothesis is immune to being disconfirmed because the seemingly disconfirming observation is equally predicted by the

truth and falsity of the hypothesis. Given that conspiracies are plots designed by agents trying to keep their intentions and actions secret, conspiratorial explanations sometimes predict that the conspirators are fabricating misleading evidence in order to hide the truth. Seemingly disconfirming evidence can be just as likely on the assumption of a conspiracy as it is on the assumption that there is no conspiracy. And this, according to Keeley, could in some cases hold for any potential disconfirming evidence.[20]

Can conspiracy theories be rationally held in virtue of the probabilistic irrelevance condition? I believe that a conspiratorial explanation can only be immune to being disconfirmed by any new evidence if it remains so general that it makes no specific predictions. A conspiratorial explanation of a fact or event seems to be constituted by two complementary claims: a *conspiracy claim*, according to which the activity of a group of agents is behind some fact or event, and a *cover-up claim*, which states that these agents are planting misleading evidence in order to hide their conspiratorial activity. If the conspiratorial explanation stays at a high level of generality, then it would indeed be able to account for any evidence that might arise. By not committing to a precise account of how the conspiratorial activity was carried out and by whom in the conspiracy claim, the explanation leaves open all possibilities for the kind of misleading evidence that is expected by the cover-up claim. So, no matter what is offered as disconfirming evidence, it can be dismissed as a fabrication of the conspirators. Consider a very vague version of Con, according to which:

> **Con Gen:** The attacks on 9/11 were part of a conspiracy of agents who are trying to hide the truth.

In Con Gen, the general conspiracy claim that someone orchestrated the attacks on 9/11

[20]I find it hard to make sense of Keeley's claim that seemingly disconfirming evidence could be construed as supporting evidence, rather than just as neutral evidence, because it is difficult to imagine a case in which the disconfirming evidence is more strongly predicted by the conspiratorial explanation than by its negation. It seems to be part of what seemingly disconfirming evidence is that it cannot support the conspiracy hypothesis *more* than its negation.

is compatible with the most general cover-up claim that someone is hiding the truth. Any disconfirming evidence could have been planted by whoever is behind the attacks. Even though no explosive was found on the site of the alleged demolition, this could be a false report of the investigators, or of the media. Or it is possible that the Twin Towers weren't demolished, but whatever happened to them, someone within the USA was behind it. Even though there is no evidence of people entering the building with large amounts of explosives during the days prior to the attacks, someone may in fact be hiding evidence of this, or the explosive material may have been brought inside bit by bit over a very long span of time. The generality of the conspiracy claim, together with the cover-up claim allow Con Gen to accommodate any relevant disconfirming evidence. However, Con Gen is a bad explanation of the evidence, because it fails to make specific predictions. It just claims that 9/11 was an inside job, and 'they' are trying to make us believe otherwise. Hence, we should expect evidence that disconfirms the conspiratorial account. But this is far from being a prediction. We would not say that a scientific theory makes predictions if it claims that at some point some evidence in favor of it will come up. Making genuine predictions requires more than this.[21]

If the conspiracy claim of the conspiratorial explanation takes a precise form, then the level of immunization will be constrained accordingly in the cover-up claim. When the hypothesis is made more precise regarding the exact form of the conspiratorial activity, including who is involved and why, it can make specific predictions regarding what counter-evidence can be expected and which sources of information are not to be trusted. A more precise conspiratorial hypothesis makes genuine predictions, but it also leaves open the possibility of encountering disconfirming evidence should the predictions fail to come true. This disconfirming evidence will have an effect on a rational agent's confidence in the truth of the hypothesis. Let's now consider a specific version of Con, according to which:

[21] In a similar fashion, a conspiratorial hypothesis that identified all-powerful conspirators would be immune to being disconfirmed but equally incapable of making genuine predictions. I elaborate this point in §5.2.

> **Con Spec:** Government officials staged the attack to the Twin Towers on 9/11. The buildings collapsed as the result of a controlled demolition. In fact, the jet-fuel-induced fires in the Twin Towers could not have melted steel. Nano-thermite was secretly brought inside the buildings and planted in the metal beams supporting the buildings to demolish them.

This hypothesis is specific enough to provide a genuine explanation of the events, and it makes testable predictions. But, by doing so, it makes itself vulnerable to disconfirming evidence. The evidence that insufficient amount of explosive residue was found on the site is more likely on the hypothesis that Con Spec is false, than on the hypothesis that it is true. Similarly, other observations would disconfirm Con Spec, including the fact that the majority of the world's experts agree that the collapse resulted from the structural damage produced by the jet-fuel-induced fires, the amount of thermite necessary to cut steel beams vertically is enormous and not likely to have been brought into the building in secret, and so on.[22] Once a specific version of the conspiratorial explanation is proposed, then the cover-up claim must also take a determinate form, and disconfirming evidence must be taken into account. Of course, a believer could maintain a coherent set of beliefs by altering the explanation—both the explanation of the conspiratorial activity, and the explanation of who is involved in covering it up—as counter-evidence arises. However, these alterations would be ad hoc and would make the believer irresponsive to the evidence in a problematic way. While it is always possible to maintain a coherent set of beliefs by using the conspiracy claim to modify one's predictions, doing so renders one's belief irrational because one is not appropriately responding to the evidence.[23]

[22]For instance, Dunbar & Reagan (2006).

[23]I am sympathetic to the argument made by Clarke (2002) that conspiracy theories often have the characteristics of what Lakatos (1978) referred to as degenerating research programs. A degenerating research program is a research program in which the participants are dedicated to protecting the core of a theory from falsification by altering auxiliary hypotheses and initial conditions in light of the new disconfirming evidence. I agree with Clarke that conspiracy theories are often rendered immune to falsification in this problematic way. It has been objected to Clarke that the exact point at which a conspiracy theory becomes a degenerating research program is unclear (Harris 2018). However, a similar concern does not apply to

Detailed conspiratorial hypotheses cannot rationally resist falsification in light of any disconfirming evidence. Only very general conspiratorial hypotheses, which do not make any specific claims about how the conspiracy was carried out and who is involved, can. However, the resilience of these conspiratorial hypotheses comes at the cost of indeterminacy and lack of predictive power. These hypotheses are not explanations of the evidence because they provide little understanding of the phenomena they purport to explain.[24] Could an agent rationally hold a very general, indeterminate conspiracy theory? First, it is hard to see what kind of evidence could support forming the belief in such a theory, other than the disbelief in the received account. In order to avoid committing to a specific conspiratorial and cover-up claim, they need to remain at such a level of generality that is more similar to skepticism in the received account than to a genuine hypothesis. However, disbelief in the received account does not warrant positive belief in the existence of a conspiracy. Secondly, if the conspiratorial hypothesis is based on evidence rather than just skepticism in the received account, for any general conspiratorial hypothesis there will be a more specific one that is a better explanation of the evidence in virtue of exhibiting more epistemic virtues, and should as such be preferred.

I take it that neither of the two conditions (certainty and irrelevance) that would render the evidence irrelevant to a rational agent's credence in a conspiratorial hypothesis can justify conspiracy theories' evidential insulation. Certainty is not a good candidate because, given the empirical nature of conspiracies, one could never be rationally certain of the exis-

my account, since I take conspiracy theories to be the extreme case of conspiracy-beliefs held in such a way as to be completely immune to disconfirmation in nearby possible worlds. If there is a such a point at which a research program becomes a degenerating one, conspiratorial explanations whose believers will retain in light of any disconfirming evidence one could encounter are an example of that.

[24] I take this to be a further advantage of my account of conspiracy theories over traditional ones. Some conspiracy-beliefs which we would ordinarily call 'conspiracy theories' do not seem to meet the threshold for being considered explanations or theories; they do not make any specific predictions, and they don't explain any evidence. Muirhead & Rosenblum (2019) refer to this phenomenon of conspiracies without theories as the new conspiracism. In the traditional account, conspiracy-beliefs of this kind would not be called 'conspiracy theories.' By identifying conspiracy theories with self-insulated conspiracy-beliefs, my account of conspiracy theories has the advantage of including these conspiracy-beliefs that do not meet the conditions for being considered explanations or theories.

tence of a conspiracy. As for probabilistic irrelevance, it only applies to conspiracy claims so general that they can barely be considered explanations, and are not supported by evidence so as to warrant positive belief in them. Genuine explanations, those specific enough to make predictions regarding what disconfirming evidence is to be expected, will either have to be disconfirmed by new evidence, or will have to be adjusted to accommodate for the new evidence in an ad hoc way. It follows that one could never rationally hold the belief in a conspiracy that is immune to being disconfirmed by counter-evidence. So, conspiracy theories as self-insulated conspiracy-beliefs can never be rationally held. Having restricted self-insulation to immunity in nearby possible worlds, we cannot claim that conspiracy theories are *necessarily* irrational. However, we can say that they are irrational to hold in this world and all the nearby possible worlds in which evidence coming from things like omniscient oracles, time travel, and mind reading are not available.

This analysis also shows that the resistance to revision that many conspiracy theorists exhibit is better understood as a feature of the believers, as my account suggests, rather than of the theories. Conspiratorial beliefs may be resistant to revision for different reasons having to do both with the content of the theory and with the agent's epistemic flaws, extra-epistemic motives, and biases. In this section I have shown that the content of the theory *alone* cannot justify evidential insulation. If we are interested in conspiracy theories that are unfalsifiable, we need to look at the individuals' beliefs.

In the next section, I address two objections to my account and point out some of its upshots. The first objection concerns the philosophical methodology on which my account is based. The second objection targets some assumptions I made in this section regarding the epistemic standards for conspiratorial explanations.

2.5 Objections and Replies

2.5.1 The Change in Meaning is a Change in Topic

Some readers might worry that re-engineering the expression 'conspiracy theory' as evidence-insulated beliefs will push the meaning of this expression too far from its current one. They might object that, by changing the intension and extension of the concept so radically, we have changed the topic of our inquiry. In fact, the way in which the expression is currently employed seems to refer to theories about conspiracies of a certain kind, rather than beliefs about conspiracies. Instead, on my view, the same theory could count as a conspiracy theory in some cases but not in others, according to the way in which each individual believer holds the conspiracy-belief (if it is evidentially insulated or not). In this section, I address two related worries: the general worry that the methodology of conceptual engineering, which I employ, is a flawed methodology, and the worry that my proposal in particular is uninteresting because it changes the meaning of 'conspiracy theory' too radically.

The first objection can be seen as an instance of the well-known Strawsonian challenge to Carnap's method of conceptual explication (Strawson 1963). In a nutshell, Strawson claims that any revisionary project that advocates for changing the extension and intension of a concept is bound to fail because, even in the most successful case, it necessarily entails a change in topic. While I think there are convincing ways of successfully rebutting the Strawsonian challenge, I will not consider them here, as this falls outside of the scope of this chapter.[25] Notice that my account is not the only one engaged in conceptual engineering. The widely accepted definition of conspiracy theory as any explanation involving a conspiracy is *also* a revisionary definition. In fact, in its ordinary use, 'conspiracy theory' has a negative valence, and does not refer to just any explanation about a conspiracy. This fact

[25]See, for instance, Cappelen (2018); Haslanger (2020b); Nado (2019); Prinzing (2017); Sawyer (2018); Thomasson (2019).

is acknowledged by the proponents of the broad definition. If conceptual re-engineering is a flawed methodology, then the most popular alternative to my account is just as doomed.

In its more specific sense, this objection could be read as an objection against my view in particular. One could argue that, while conceptual engineering may in general be a viable philosophical methodology, and changes in concepts' extensions and intensions may succeed at maintaining the same topic as the original concept, the account I propose is just too much of a shift, and fails to do so. I want respond to this objection by suggesting that both the change in intension and in extension may not be as radical as they initially appear.

First, the focus on the extreme resistance to counter-evidence as a distinctive feature of conspiracy theorizing neatly fits with the ordinary meaning of 'conspiracy theory.' Conspiracy theories have often been compared to paranoid ideation,[26] and more recently to impostor syndrome (Hawley 2019). One of the reasons for this parallel is this *self-sealing* property that they seem to have (Sunstein & Vermeule 2008; Cassam 2019). In conspiracy theories, just like in paranoid ideation and impostor syndrome, the core of the beliefs set includes the reasons to discredit disconfirming evidence and many conspiratorial beliefs seem to be 'sealed' and totally insensitive to contradicting information. The shift from theory to belief is indeed a change of perspective. However, it is a way of focusing on what has been widely recognized as a central feature of conspiracy theorizing—namely, a distinctive way in which believers resist revising their beliefs in light of new evidence.

Moreover, even the extension of the ordinary concept may, to a large extent, be preserved. The ordinary expression 'conspiracy theory' seems to imply negative value, indicating theories about conspiracies that are somehow irrational to believe, outlandish, or simply bad theories about conspiracies. The paradigmatic cases of theories that are currently called 'conspiracy theories'—the outlandish and absurd ones—might fall under the revised concept, and might do so for many of their believers. In fact, it seems plausible to suppose

[26]See, for instance, Barkun (2003); Fenster (1999); Hofstadter (1965).

that the reason why such outlandish theories have survived over the years, given that there's overwhelming and easily available evidence against them, is that most people's beliefs in these theories are immune to rational criticism and disconfirming evidence. Even though only empirical investigations could tell us whether this is actually the case, it is plausible that the extension of the concept would, to a large extent, be preserved, despite the change in meaning I advocate for.

While my proposal advocates for a shift in meaning, I don't think that focusing on stubbornly held beliefs in conspiracies represents a shift in topic.

Last, from a methodological point of view, my account is aimed at promoting the understanding of the phenomenon of conspiracy theories. The change in meaning I propose is targeted to a specific *theoretical discussion* of conspiracy theories. The ordinary expression need not be affected by it.[27] Accordingly, our intuitions about what a conspiracy theory is are only subordinate to the potential theoretical advantages that a revisionary account might have. The main advantage of understanding conspiracy theories as self-insulated conspiracy-beliefs rather than as mere theories involving conspiracies, is that it allows for empirical studies in the psychology of conspiracy theorists without having to make problematic assumptions about the rationality of believing conspiracies. On my account, evidential insulation makes conspiracy theories irrational and warrants a psychological approach to explain why people have such beliefs. Moreover, differently from traditional accounts of conspiracy theories, on my account conspiracy theories are understood as a distinctive phenomenon of people having epistemically problematic beliefs regarding conspiracies. My proposal could be seen as an attempt to carve out a space for conspiracy theories as a phenomenon irreducible to other epistemic phenomena that could explain evidence resistance (e.g., echo chambers and filter bubbles). As a working definition, the one I propose looks like a promising way to further our understanding of conspiracy theorizing. These considerations should have priority over

[27]This approach to the problem is also compatible with the existence of different revisionary accounts of 'conspiracy theory.'

our intuitions about what conspiracy theories are.

2.5.2 Predictions, Reflexivity, and Ad Hoc-Ness in Conspiratorial Explanations

In §4, I argued that it is never permissible to hold a belief in a conspiracy that is self-insulated. My discussion of the second condition, probabilistic irrelevance, relied on the two assumptions that an explanation that does not predict novel observations is worse than one which does, and that an explanation that was adjusted in light of new evidence to resist falsification would be ad hoc and thus irrational to believe. One might object that, given what conspiracies are and how they differ from explanations of natural phenomena, these assumptions are unwarranted in our case (Harris 2018; Keeley 1999, 2019).

Let's consider the first claim, that a general conspiratorial hypothesis which does not make specific predictions is a worse hypothesis than one which does. Harris (2018) argues that conspiratorial hypotheses might predict novel observations:

> [C]onspiracy theorists may predict that evidence apparently conflicting with the conspiracy theory will be presented, and such predictions will ordinarily be borne out. Hence, it would be inaccurate to claim that conspiracy theories are not capable of predicting novel observations. (Harris 2018: 247)

I take it that a genuine prediction is a claim that a particular state of affairs will occur. In order to predict a novel observation, a conspiracy theorist would have to predict what sort of seemingly disconfirming evidence will be encountered, and who is involved in trying to hide the truth of the conspiracy. Only a specific conspiratorial hypothesis, consisting of a specific conspiracy claim and cover-up claim, can do this.

Harris might grant this point, yet still deny that a lack of predictive power is problematic for conspiracy theories. He claims, following Keeley (1999), that

> Even if one denies that conspiracy theories can predict novel facts, it is not clear that this would be a strike against such theories. As Keeley points out, the objects whose behavior is described by conspiracy theories are unlike the objects of ordinary empirical sciences insofar as the objects of conspiracy theories can be expected to actively resist investigation. (Harris 2018: 247)

Since the conspirators are trying to mislead us to avoid detection, Harris and Keeley argue, it is unclear why we would expect a good theory about a conspiracy to be able to predict their moves. In other words, predicting novel observations is not necessarily a feature of good conspiratorial explanations.

While it is often the case that the nature of the explanandum is different in the case of conspiratorial explanations than in the case of explanations of natural phenomena,[28] the claim that this difference warrants different criteria for evaluating hypotheses is controversial. Sometimes conspiratorial explanations are explanations of social phenomena. If we assume that social systems are indeterministic and that the behavior of agents cannot be predicted, then we should not expect to be able to understand social phenomena at all. Conspiratorial or not, explanations of people's motives and intentions could not be assessed.[29]

On the other hand, if we assume that, to some extent, we can predict people's behavior and understand their intentions, we would expect explanations of social phenomena to be similar to other empirical explanations and subject to the same standards of assessment, in-

[28] Conspiratorial explanans always involve the intervention of human agents, but not all explananda are social phenomena. For instance, the theory that the Earth is flat, and that some powerful people in the world are trying to keep it a secret, is supposed to be an explanation of different natural observations. Similarly, the hypothesis that vaccines are a cause of autism, and that there is a conspiracy of pharmaceutical companies trying to hide the truth, is an explanation of natural observations.

[29] In fact, if anything, conspiratorial explanations would fare worse than non-conspiratorial ones because they attribute more intentionality to agents than their non-conspiratorial rivals. See Mandik (2007).

cluding the ability to predict novel observations, explanatory power, explanatory depth, and unification. While conspiratorial explanations which are detailed accounts may exhibit these traits, general ones lack the determinateness necessary to provide significant understanding of the phenomena they are formulated to explain.

The second claim, that specific conspiratorial hypotheses would have to be falsified by disconfirming evidence that the theory failed to predict, or else be irrational in virtue of being ad hoc, could be criticized on similar grounds. In comparing conspiratorial and scientific explanations, Keeley notices that

> [C]onspiratorial explanations generally engage social behavior of purposive agents, whereas the natural sciences typically restricts its studies to non-agents (or at least agents lacking an agenda to interfere with their investigations). The fact of the matter is that the scientific study of human agents by humans is fraught and methodologically contested, whether it be social psychology, economic behavior, or sexuality. When your research subjects can read your results and explanations of their behavior—and then respond with changed behavior—science gets a lot more difficult, and the easy proclamations of natural science (including falsification) go by the wayside. (Keeley 2019: 429)

So, one could grant that explanations in the social domain are not subject to different standards of evaluation, but argue that resistance to falsification is warranted for those domains subject to *reflexive prediction problems*. In domains where the behavior of the object of investigation can be influenced by knowledge of the explanations proposed, falsifiability does not seem to be a valid requirement to expect of a hypothesis. Conspiratorial explanations may be explanations of this kind. One could argue that, if a conspiratorial hypothesis' predictions fail to obtain this need not necessarily disprove the theory, because it could also indicate that the conspirators changed their behavior after the conspiratorial explanation

was made known to them.

I agree with Keeley that reflexive predictions could occur in conspiratorial explanations, thus altering the disconfirming effect that failed predictions should have on the hypothesis.[30] However, there are two reasons to resist the conclusion that reflexivity problems can justify conspiratorial explanations' immunity to falsification. First, the existence of reflexive predictions is typically employed to criticize the methodology of some social sciences, rather than to claim that, in these fields, unfalsifiable theories are warranted.[31] Similarly, the possibility of reflexive predictions seems to speak in favor of the difficulty (and in some extreme cases impossibility) to formulate good conspiratorial explanations, rather than supporting the claim that explanations which make reflexive predictions can be valid explanations even though they cannot be falsified by seemingly disconfirming evidence. If we believe that the subject of our investigation could potentially interfere with all the predictions that our theory makes, then we should give up the hope of formulating a good conspiratorial explanation of the events. We should come to terms with the impossibility of arriving at the truth, and suspend judgment on the matter, rather than claiming that unfalsifiability is not a problematic feature of conspiratorial explanations.

Second, not all the predictions made by conspiratorial hypotheses are of the kind that can give rise to reflexivity worries. Recall the distinction between the conspiracy claim and the cover-up claim that constitute a conspiratorial hypothesis. The conspiracy claim states that the activity of a certain group of agents is behind a fact or event. The cover-up claim makes predictions as to what kind of counter-evidence will be encountered. While the conspirators might change their behavior to falsify the cover-up claim's predictions, many of the conspiracy claim's predictions cannot be altered by the conspirators' behavior in the same way. Especially in those cases where a conspiracy is postulated to explain a past event,

[30] For a Bayesian analysis of how reflexivity alters confirmation relations, see Kopec (2011).

[31] For a discussion of the methodological problems generated by reflexive predictions, see Buck (1963); Grünbaum (1963); Romanos (1973); Vetterling (1976).

reflexivity is not a problem for the conspiracy claim's predictions relative to who is involved in the conspiracy and how the conspiratorial activity was carried out.

Nothing about the nature of conspiratorial explanations allows us to assess them according to standards different from other empirical explanations. Just like any other explanations, very general conspiratorial explanations that do not make novel predictions and lack other explanatory virtues are bad explanations, and conspiratorial explanations that are modified in light of new evidence to resist falsification are problematically ad hoc.

2.5.3 Testimonial Insulation

Another objection against the idea that conspiracy theories as insulated beliefs are irrational is that, given that evidence of conspiracy theories in normal circumstances is rarely first hand, one could be rational in resisting revision if one mistrusted the sources from which counter-evidence could be obtained. Hence, one could rationally hold a conspiracy-belief that is immune to revision in normal circumstances.

In order to respond to this objection, we need to consider two different scenarios: (i) all of the sources of evidence relevant to the existence of the conspiracy are deemed untrustworthy for reasons independent of the conspiracy; (ii) the sources are discredited after receiving the conflicting testimony, on the basis of the belief in the conspiracy. It should become clear that (i) is a case in which it is rational to resist revision in light of any testimonial evidence, but (i) does not represent an instance of conspiracy theory in the relevant sense. On the other hand, (ii) is a genuine case of conspiracy theory, but it is not an instance of a rationally held one.

Let's consider each case with an example. Imagine a person, Anna, who believes that vaccines cause autism, and that a conspiracy of the pharmaceutical companies is hiding

the truth on this issue. If Anna had independent reasons to mistrust scientists, doctors, news outlets, and anyone else who may be providing testimony that could disconfirm her theory, then it would seem that Anna is behaving rationally when ignoring these sources and remaining confident in her conspiracy-belief. But it would also be clear that Anna's belief is not a conspiracy theory in the relevant sense. In fact, her resistance to counter-evidence is not due to her belief that a conspiracy is going on, but rather to her independent reasons not to trust some sources of information relevant to the issue of whether vaccines cause autism. Her belief might be rationally immune to disconfirmation, but it is not a conspiracy theory.

On the other hand, imagine that after forming her conspiracy-belief, Anna received testimonial counter-evidence from sources that her initial conspiratorial explanation gave no reasons to mistrust. If she then demoted these sources on the basis that, given what they testify, the conspirators must have influenced them (for instance, by deceiving them or by buying their complicity) or that they may themselves be part of the group of conspirators, then the insulated belief would count as a conspiracy theory. The conspiracy is what is taken to justify the dismissal of the relevant evidence. However, it would be irrational for Anna to demote the new sources on the basis of her conspiratorial belief. As I argued in §4, the new testimony could only be accounted for by a vague theory which did not commit to a specific cover-up claim. If she had a more specific conspiratorial hypothesis, then conflicting testimony from sources who were not initially thought to be involved in the conspiracy should affect (at least minimally) her confidence. A failure to respond to testimonial evidence would make her belief an irrational conspiracy theory.

2.6 Conclusion

In this chapter, I have offered an account of conspiracy theories as self-insulated beliefs in the existence of conspiracies. I have argued that conspiracy theories so understood are always

irrational.

A big advantage of my account over the alternative broad and neutral understanding of 'conspiracy theory' is that it allows for treating conspiracy theories as a specific epistemic phenomenon that has been playing an important role in the political and social climate of the past decade. Traditional accounts of conspiracy theories, which identify conspiracy theories with conspiratorial explanations, have failed to recognize the deeply problematic aspects—both political and epistemic—of the phenomenon of conspiracy theorizing, and have often depicted conspiracy theorists as analogous to investigative journalists. Focusing on conspiracy theories as insulated conspiracy-beliefs is an attempt to promote an investigation of the phenomenon of conspiracy theories as a distinctive one, to be understood in its current political and social function.[32]

In this sense, this account of conspiracy theories is in line with other research in social epistemology aimed at making sense of the seemingly absurd opinions that some people hold (despite the easy and widespread access to information that the Internet grants), without having to assume that, somehow, these people have stopped being responsive to the demands of truth and rationality.[33] Conspiracy theories are an irrational way of holding conspiracy-beliefs. However, they are alluring explanations which can easily accommodate disconfirming evidence, because they can be made internally coherent by dismissing the evidence as a fabrication of the conspirators. Only when we look closely at the dynamics of the dismissal of counter-evidence does it become apparent that conspiracy theorists can only maintain the internal coherence of their theories by not being adequately responsive to the evidence—either by adopting a poor, indeterminate explanation of the evidence, or by adopting a more specific hypothesis but failing to respond to new evidence.[34]

[32]Such as the role of conspiracy theories as forms of political propaganda (Cassam 2019).

[33]Fake news is one such research topic. Other examples include echo chambers and filter bubbles (Jamieson & Cappella 2008; Nguyen 2020), and evidential preemption (Begby forthcoming).

[34]I would like to thank Endre Begby, Anna Boncompagni, Quassim Cassam, Thomas Grundman, and an anonymous referee for their helpful comments on earlier drafts of this chapter. I am grateful for the insightful discussions I have had with the members of CONCEPT and the graduate students at UC Irvine.

Special thanks go to Sven Bernecker, Paul Silva, and Eyal Tal for their detailed feedback and invaluable support with this project.

Chapter 3

Group Conspiracy Theories

3.1 Introduction

The phenomenon of conspiracy theorizing seems to be, at least in part, a *collective* phenomenon. Conspiracy theorists are rarely individuals operating in isolation. Rather, they appear to be involved in a kind of *group activity*, and they see themselves in this way—or so it would seem from the way we talk. Let me illustrate my point with some examples:

1. *Names.* We typically refer to the supporters of certain conspiracy theories as groups of individuals with names: 'Birthers' for the group of supporters of the conspiracy theory that Barack Obama was not born in the US; 'Truthers' for those who support alternative explanations of 9/11; 'Anti-vaxxers' for people who are sceptical of the safety of certain vaccines and endorse the existence of a conspiracy surrounding the information available about this through mainstream sources; 'Flat-earthers' for those who believes that the Earth is flat and that there is a conspiracy of powerful people trying to hide this fact from the public. These names to designate members of a certain group are often accompanied by names to refer to the group itself, such as 'movement'—as in 'the

9/11 Truth movement'—or 'society'—as in 'the Flat-Earth society'. It is important to notice that these names are not just used by outsiders to refer to these groups, perhaps with the intention of ridiculing them. Some of these groups are "official": they have meetings, conferences and official websites, their members identify with these labels, and it is not unusual to hear someone claim, "I am a flat-earther".[1]

2. *"Us" against "Them".* The narrative of "Us" against "Them" often accompanies conspiracy theories. Supporters of conspiracy theories tend to attribute the existence of disagreement with outsiders to either the outsiders' malicious intention to deceive them, or to the outsiders' lack of critical thinking, sufficient intellectual ability, or to their excessive trust in the official sources of information. Conspiracy theorists thus tend to position themselves against an out-group and explain their disagreement in this way: "Don't listen to them, they wouldn't tell you/ They are lying/ They are *sheeple*!". At the same time, they seem to see themselves as a group of people who have something in common, they are onto the truth, they 'think outside the box'. Conspiracy theories' supporters often describe their activity of searching for the truth as a group activity. As an example, consider the abbreviation that Qanon supporters have been using as their slogan (and hashtag on social media) *wwg1wga*, i.e., 'where we go one we go all'. This and similar ways of speaking are often encountered in online conversations amongst conspiracy theorists.

3. *Disagreement and traitors.* When people who were perceived by the other supporters of the conspiracy theory as being on their side express thoughts contrary to the theory or dissociate from it completely, the supporters of the conspiracy theory tend to react harshly against them and express feelings of having been wronged or betrayed. One example of this phenomenon from recent events regards Alex Jones, the famous

[1] This result is even more interesting when compared with some of the results of Wood and Douglas (2013), who found that people do not tend to attribute to their own views the label 'conspiracy theories'. One explanation for this is that the members of these groups don't think of their theories as being 'conspiracy theories', despite them being commonly referred to as such by their opposers.

conspiracist show host and founder of the conspiracy theorist website *Infowars*. Alex Jones, who had in the past endorsed some of the Qanon theories, rejected and disparaged them during one of his show's episodes. The response on social media by Qanon supporters has been very strong, addressing Alex Jones as a 'traitor'. While this is just one well-known case, the phenomenon of rebuke of ex-supporters of conspiracy theories seems to be common enough in online conversations amongst conspiracy theorists. It can be observed in comment sections and forums that there is a different response to someone expressing a view that opposes the conspiracy theory depending on whether the person expressing dissent is originally perceived as an opponent or a supporter of the theory. In the former case, the response expresses mere disagreement (even when it takes angry and personally insulting tones), while in the latter it also expresses rebuke for the perceived betrayal of the conspiracist cause.

While they are not meant as an exhaustive list, and it might be questioned whether these are widespread and reliable features of conspiracy theorizing, I take these observations to be plausible enough to motivate the basic, core intuition that, when trying to understand and address the phenomenon of conspiracy theories, we are dealing, at least in part, with a collective phenomenon.

The philosophical literature on conspiracy theories has mainly focused on individuals and their attitudes towards theories about conspiracies. The philosophical debate surrounding conspiracy theories has been rapidly expanding over the last decade. One aspect that is common to most approaches is the assumption that a conspiracy theory is, minimally, a type of explanation that involves a conspiracy (plus, on some views, additional features).[2] The debate on the epistemology of conspiracy theories has mostly been concerned with determining whether it can ever be rational for individuals to believe such theories.[3] While some authors have emphasized the importance of the social epistemic environments in which

[2] See, for instance, the papers contained in Dentith 2018a.
[3] See, for instance, Keeley 1999; Pigden 2007; Denith 2016; Harris 2018.

conspiracy theorists operate, the *groups of supporters of conspiracy theories* and their beliefs have so far not been considered. The first aim of this paper is to show the necessity for an analysis of conspiracy theories that takes the collective dimension of conspiracy theories seriously, making it essential to what conspiracy theories are—in order to arrive at a deeper understanding of their persistence and resilience in our society.

The second aim of this paper is to provide one specific account of the collective dimension of conspiracy theories, which takes into consideration the three aspects listed above. I isolate and describe a group phenomenon—which I call *group conspiracy theories*—which I argue is a phenomenon of Gilbertian group belief in the existence of a conspiracy. I outline the characteristics of group conspiracy theories and their believers, which show the limitations of individual-centered approaches to the epistemology of conspiracy theories.

Here is an outline for the rest of the paper. In §1, I offer my account of group conspiracy theory as a Gilbertian group belief in the existence of a conspiracy. I then discuss, in §2, group conspiracy theories' feature of being resistant to group disconfirmation in light of new evidence. In §3, I consider some upshots of this account, and in §4 I motivate the adoption of a framework which centers on individual and group belief in conspiracies to fully analyse the phenomenon of conspiracy theorizing.

3.2 Group Conspiracy Theories: a Joint Commitment Account

In this section, I identify and outline the features of a collective phenomenon surrounding conspiracy theorizing—which I call *group conspiracy theory.* On my view, a group conspiracy theory is a kind of group belief in the existence of a certain conspiracy: in this sense, we can think of the conspiracy theory that Barack Obama wasn't born in the US, in the group

sense, as the belief of a group of supporters that there is a conspiracy behind the widespread conviction that Barack Obama was born in Honolulu; similarly, we can think of the Flat-Earth conspiracy theory, in the group sense, as the belief of a group that a conspiracy explains why the majority of people believes the Earth to be round; Pizzagate conspiracy theory, in the group sense, is the belief of a group that a conspiracy of democratic politicians was running a child sex ring in a pizzeria in Washington DC; and so on. But what does it mean for a conspiracy theory to be a group belief in the existence of a conspiracy? I adopt a Gilbertian account of group belief, according to which a group belief in the existence of a certain conspiracy (e.g., Obama's birth, flat-earth, Pizzagate) is *the joint commitment by a group to believe in the existence of said conspiracy as a body.* I will clarify this definition in what follows, starting with Gilbert's notions of a joint commitment and group belief.

3.2.1 Gilbertian joint commitment and group belief

On Gilbert's account, a joint commitment is a commitment of two or more parties. Differently from individual commitments, joint commitments can only be brought in and out of existence by all the parties involved. A personal commitment, such as a personal decision to do something, is generated by the individual alone and can be similarly rescinded by a decision not to do that thing anymore. Joint commitments, on the other hand, require all the parties involved to enter the commitment together, and can only be rescinded by a communal decision (Gilbert 2013: 40).[4] In this sense,

> A joint commitment is not a conjunction of personal commitments of the different parties. Each of the parties plays a part in the creation of a joint commitment, not by creating an appropriate personal commitment, but by expressing to the others his or her willingness to be jointly committed with them. (Gilbert 2001:

[4] All these characteristics apply to the "standard case", but Gilbert discusses cases where background understandings might make the conditions for creating and rescinding joint commitments different.

115)

The creation of a joint commitment requires the expression of readiness to enter the commitment of the parties, under conditions of common knowledge.[5]

In a joint commitment, the parties are *answerable* to each other for any violations of the commitment (Gilbert 2013: 40). If one of the parties fails to act on the existing joint commitment, the other parties stand in a position to justifiedly rebuke them for their violation. For Gilbert, joint commitments in this sense give rise to obligations (Gilbert 2006, ch. 7).

Finally, a joint commitment, for Gilbert, is always a commitment to *act as a body*, where 'acting' is broadly understood to include things such as deciding, accepting a goal, intending, and believing (Gilbert 2013: 41).

On the basis of her account of joint commitment, Gilbert proposes this account of group belief:

> The members of a population, P, collectively believe that p iff they are jointly committed to believe that p as a body. (Gilbert 2013: 137)

What does it mean for the members of a population to be committed to believe that p *as a body*? For Gilbert, this is equivalent to saying that they are committed to 'act so as to emulate, as far as possible, a body that believes that p' (2013: 140). The individual members are not required to believe that p. What is required of them is that their *public performance* is in agreement with being part of a group that believes that p. For Gilbert, this does not even require that the members act as if they *personally* believed that p, but rather they are required to act 'as would any one of several *mouthpieces* of the body in question, thus uttering *its* beliefs' (2013: 140). The members of the group in question are required, by

[5] On the common knowledge condition, See Gilbert 1989: 188–197.

virtue of their joint commitment, to emulate a single believer of the proposition p, and to refrain from acting, including speaking, in ways that violate their commitment. Hence, in the relevant contexts, a member who claimed "p is false!", would be violating their commitment to act so as to emulate a single believer of p—regardless of whether or not they personally believe p.[6]

3.2.2 Group conspiracy theories defined

With this basic reconstruction of Gilbert's account in mind, let's now turn to group conspiracy theories again. In a preliminary way, we can say that:

> A group conspiracy theory is the group belief of a population P in the existence of a conspiracy C.

On a Gilbertian account of group belief, we can say that a group conspiracy theory is *the joint commitment by a group to believe in the existence of a certain conspiracy*—for instance, flat Earth theory, in the group sense, is the joint commitment by a group of people to believe as a body that there is a conspiracy behind the widespread theory that the Earth is a globe. A few important features about group conspiracy theories follow from the Gilbertian account.

The first thing to notice about this notion of group conspiracy theory as a group belief is that, in order for there to be a group belief, the individual belief of the members is neither necessary nor sufficient. A collection of individuals who personally believe that the Earth is flat and that there is a conspiracy behind the theory that the Earth is a globe do not necessarily generate a group conspiracy theory. Vice versa, a group belief in the existence of a conspiracy behind the theory that the Earth is a globe might exist when none of the

[6] There are contexts in which single members could express their disagreement without thereby violating their commitment to the group's belief that p, for instance in a context where the members of the group are deliberating on whether they still believe that p.

members *personally* believe it.

Secondly, just as with any other joint commitment, joint commitments to believing in the existence of a conspiracy carry with them obligations of the parties to act in certain ways. In particular, the members of a group that is jointly committed to believing in the existence of a conspiracy have an obligation to act, including speak, in ways that do not conflict with being 'mouthpieces of a single body' (Gilbert 2013: 140) that believes in the conspiracy. A failure by any member to uphold this joint commitment would put the other members of the group in the position to justifiedly rebuke them.

There are two special features that set group conspiracy theories aside from the standard cases that Gilbert considers in her analysis of group belief. The first one regards the type of group that is the subject of group conspiracy theories. In a group conspiracy theory, the subject is what we may call a *belief-centered group*, in the sense of a group which exists in virtue of a shared belief—"we are those who believe that p". In other words, the group comes into existence with the formation of the group's belief, *as the plural subject of that belief*, and is not independent of it. Gilbert identifies social groups with the plural subjects of action (Gilbert 1989, ch. 4). When considering group belief, she typically refers to examples of pre-existing groups—e.g., a book club (1987)—and she discusses what it is for these independently existing groups to adopt a belief. Even though nothing she says conflicts with the case I have in mind, it is important to clarify that group conspiracy theories are not the beliefs of groups that exist independently of their group belief. A group conspiracy theory is not a case of a pre-existing group which is already the plural subject of some joint action, which additionally shares a belief in the existence of a conspiracy. With a group conspiracy theory, a collection of individuals constitutes a social group just *in virtue of becoming the plural subject of a belief in the existence of a conspiracy.* In this sense, we can say that the groups of believers of conspiracy theories are belief-centered: the commitment to believing in the conspiracy determines them as groups. So, to clarify our initial definition:

A group conspiracy theory is the group belief of a belief-centered group G in the existence of a conspiracy C.

This characteristic of group conspiracy theories carries with it one more important feature which distinguishes these cases from the standard Gilbertian cases, and which I will discuss in depth in the next section. Since the subject of a group conspiracy theory constitutes a group only insofar as it is the subject of a group belief, it is a very *flexible* group in terms of its members. In order to be a part of a group of this kind, an individual must have entered into a joint commitment with others to collectively believe in the conspiracy. When they violate this commitment by acting in ways that go against their being part of a group that believes in a certain conspiracy, this puts the other members in a position not only to rebuke them for their violation of the commitment, but also to question their membership to the group.

The features of group conspiracy theories discussed in this section have implications for the relationship that these group beliefs in conspiracies have with evidence against the existence of the conspiracy. I turn to this issue in the next section.

3.3 Group Conspiracy Theories and Resistance to Evidence

As we have seen in the previous section, Gilbertian joint commitments, including commitments to believe, generate obligations amongst the parties involved. These obligations in turn seem to create incentives for the parties to act in ways that conform to their commitment. In order to avoid being rebuked by the other members, the parties to a joint commitment to believe that p ought not act (including speak) in ways that violate the joint commitment to believing p. This feature of Gilbertian group belief makes it the case that

group belief revision is subject to pressures deriving from the mutual obligations that members have to upholding the joint commitment. For instance, according to Gilbert, scientific consensus, understood in terms of group belief, is hard to change due to the costs associated with members expressing opposing personal beliefs (Gilbert 2000).

However, the pressures deriving from the mutual obligations in a joint commitment by a group to believe that p are not always such as to render the group belief totally immune to revision. In the standard case of group belief, even though members might feel the pressure not to express personal opinions that contradict the groups' belief, this is only one amongst the obligations that members of a group have towards one another. A scientific community, we can suppose, constitutes a group in virtue of sharing different commitments—to acting in academic honesty, to adopting the scientific method, to making scientific recommendations in line with the evidence available. If a member of the scientific community spoke against the scientific consensus without qualifying their opinion as personal, they would be subject to rebuke from the other members, but their belonging to the group in virtue of the other commitments would not be thereby called into question. This allows for individuals' expressions of personal beliefs contrary to the group's belief in two circumstances. First, to some extent, individual members can express personal beliefs that are against the group belief when qualifying them as personal. A scientist can claim that "the scientific consensus about x is p, even though I personally believe that not-p". Secondly, members can bring up reasons to doubt the group belief in those contexts where the group is deliberating about the matter. It would be perfectly appropriate for a scientist to bring up the results of a study against the scientific consensus on x in the designated contexts, such as at a conference, or in a scientific journal.

The situation is different for belief-centered groups. In belief-centered groups, the group's existence is conditional on the group's belief: The group does not exist except as the plural subject of its belief. For this reason, when a member of the group acts in such a way as

to go against the joint commitment, the members of the group can justifiedly question the sincerity of a person's commitment. If we are those who believe that p, and we are a group solely in virtue of having this shared belief, if someone were to express doubts about p, even if they qualified their opinion as personal, this would call into question their membership to the group. If I were to say, "I don't personally believe that p", the others would be justified in saying, "But we thought you were one of us, those who believe that p!". This is not to say that all members of a belief-centered group ought to have the relevant belief, but just that a member cannot speak against their commitment even when clarifying that they are not speaking for the group but just for themselves. If this were to happen, the other members would be in a position to eject them from the group, and justifiedly rebuke them for having violated their commitment.[7]

Another difference between the standard case and the case of belief-centered group is that, while in the standard case the group itself can change its mind about any of its beliefs, a belief-centered group's existence is conditional on the group having the relevant belief. If we are those who believe that p, we can't, as a group, stop believing that p. While individual members can exit the group by failing to uphold their commitment and being ejected by the other members, the group *as a whole* cannot change its mind about p. The group's change of opinion about p would result in the group deciding to rescind their joint commitment to believing p and ceasing to exist. This is why it would be costly for individual members to bring up reasons against p even in contexts where the members of the group can deliberate together regarding p. While in the standard case there are contexts where it is appropriate for the members to bring up evidence against their belief—for instance, in scientific journals or at conferences, in the case of scientific consensus—this would not necessarily be similarly

[7] I believe this is a feature of any group that is held together by a single joint commitment. If a member acted as to violate that commitment, their membership to the group would be called into question. For example, consider a group of hikers who meet up weekly to go on a hike. In this case, it would perfectly fine for someone to say, "Our group is going on a hike tomorrow, but I'm not going". However, imagine a different group of people who intend to go on a hike tomorrow, but who do not constitute a group independently of their intention. If someone were to utter their intention not to go on the hike tomorrow, the other members would be justified in questioning whether they belong to the group.

appropriate in belief-centered groups. If someone were to bring up to the other members that there might be reasons not to believe p, the others, instead of considering these reasons and determining whether their group should cease to exist, could just ask them to leave the group: "If you are no longer committed to p, then you are no longer one of us".

Belief revision is even more difficult in group conspiracy theories due to the fact that these groups are typically very large in size and form mostly online among people who don't know each other, or *of* each other. There seem to be conventions around online communication, such that certain online behavior—such as using certain hashtags, or commenting on certain threads to defend a certain conspiracy—is taken to indicate readiness to enter a joint commitment to believing a conspiracy. Large belief-centered conspiracy theory groups are formed in this way through background understandings and conventions that can create the conditions for expressing readiness to enter a joint commitment in such large-scale and impersonal situations.[8] This feature of group conspiracy theories makes it even harder for groups to change their minds, as there don't seem to be many contexts available for members to discuss their joint commitment to the group belief, and whether new evidence might suggest they ought to change their mind and cease to exist as a group. It is much more likely for individual members to default on their commitment than it is for them to try and change the group's belief.

In addition to this difficulty, the members of group conspiracy theories do sometimes personally believe in the existence of a conspiracy, and, as I have argued in Chapter 2, they sometimes do so in a stubborn way which resists revision in light of new evidence. If this is the case, then stubborn individual believers might hinder the process of group revision of the belief even further.

Being the beliefs of groups that exist in virtue of their belief, group conspiracy theories are hard to dislodge. Flat-Earthers are a group of people who are committed to believing

[8] Gilbert discusses joint commitments in large populations in, for instance, 2013: 51–52.

that there's a conspiracy behind the globe theory of the Earth. Flat-Earth theory, in the group sense, is their group belief. As long as the group exists, this belief is immune to disconfirmation. The group could cease to exist, together with its belief, but given the typical features of these groups—their size, mode of communication, flexibility, personal belief of its members—this is unlikely to happen. For group conspiracy theories, a joint commitment by the group to believing in the existence of a conspiracy as a body corresponds to a joint commitment by the group *to defending the existence of a conspiracy in light of new evidence and outside attacks.*

3.4 Some upshots

The account of group conspiracy theories provided thus far can explain several features of the phenomenon of conspiracy theorizing and the group dimension I introduced at the beginning of the paper.

When we look at existing conspiracy theories also as group conspiracy theories, we can recognize them as a collective phenomenon of a group of people who are jointly committed to defending the existence of a conspiracy. This joint commitment makes them into a group, a belief-centered group of individuals who are acting together. This explains the way in which we tend to talk about certain groups of conspiracy theorists, and the way in which they tend to talk about themselves and their actions—for instance, the fact that we have names for the supporters of certain theories, and the fact that their narrative is often presented as Us against Them. Moreover, looking at conspiracy theorizing under this account of group conspiracy theories can explain the phenomenon of rebuke that we observe when someone who is perceived as a supporter of a given conspiracy claim expresses negative opinions about it. Under the joint commitment account of group belief, a person who acts in ways that are contrary to their joint commitment to believing in a conspiracy together with the other

members of the group can be justifiedly rebuked by the other members for so acting.

There are also some additional features of the phenomenon of conspiracy theories that can be explained through this account of group conspiracy theories as Gilbertian group beliefs. First, on this account, the groups of conspiracy theorists who are the plural subject of group conspiracy theories need not personally believe in the existence of a conspiracy. And this squares with the fact that, with many of those theories that we typically classify as 'conspiracy theories', it is often unclear whether the individuals who support them *actually* personally believe them. Some of these theories are simply too outlandish to believe, and sometimes it seems that some supporters might be driven by motives other than belief—for instance, financial incentives, or political gain. Group conspiracy theories allow us to focus on the phenomenon of the groups who support certain theories as a phenomenon of group belief, without having to even consider whether the individuals who are members of these groups believe in the conspiracies themselves. Their personal belief is irrelevant to their commitment to believe as a body in the existence of a conspiracy, and to act on this commitment. In this sense, they are conspiracy theorists regardless of what they personally believe.

Second, this account of group conspiracy theories can contribute to an explanation of why conspiracy theories are so hard to eradicate. While several studies have been conducted on the psychological factors that drive belief in conspiracy theories in individuals, and on the social and political functions of conspiracy theories, the group dynamics that might contribute to the resilience of such outlandish theories have been largely overlooked. Group conspiracy theories' resilience to disconfirmation might also play a role in explaining the resilience of some outlandish theories about conspiracies in the public opinion. The obligations that are generated by joint commitments, and the corresponding right to rebuke that the other members have in case of a violation, create an incentive for supporters of these conspiracies to keep believing in the conspiracy in the group sense, thereby staying in the

group. Moreover, the flexibility of these groups, where members can constantly enter and exit the group just by virtue of entering a joint commitment to believing in the conspiracy together with the other members, together with their large size, and typical mode of communication, make it almost impossible for the group as whole to assess and respond to disconfirming evidence against the conspiracy. These group dynamics might contribute to our understanding of why widely debunked theories about conspiracies are still so prevalent in our society.

Third, the particular features of certain types of online communication play a special role in enabling the formation of large and flexible groups of conspiracy theory supporters, which allow for highly evidence-resistant group conspiracy theories. Group conspiracy theories are not resistant to evidence by definition. As we've seen, their extreme resilience depends on the contingent features that contemporary group conspiracy theories have in virtue of being the beliefs of groups of people who don't know each other or of each other, and who easily enter in joint commitments by acting online in certain codified ways. The existence of conventions around online communication—such as the use of hashtags, or other signaling behaviors like commenting or re-posting—make the existence of group conspiracy theories possible in this special way, which renders these group beliefs extremely unlikely to be abandoned by their groups.

My account of group conspiracy theories as Gilbertian group beliefs in conspiracies can explain many features of the contemporary phenomenon of conspiracy theorizing. It also seems to capture the group dimension of this phenomenon which the contemporary philosophical literature on conspiracy theories has so far overlooked—focusing instead on conspiracy theories as theories about conspiracies, and on the individuals who believe them.

3.5 Individual and Group Conspiracy Theories

The group dimension, however, does not exhaust the complexity of conspiracy theories. I have argued in Chapter 2 for an account of conspiracy theories, in the individual sense, as individual beliefs in conspiracies which resist revision in light of any counterevidence that one might encounter under normal circumstances. What I want to suggest here is that, taken together, individual conspiracy theories and group conspiracy theories—understood as a type of individual and group belief in the existence of a conspiracy—provide an insightful framework for discussing and addressing the topic of conspiracy theories in its entirety.

Individual conspiracy theories are individual beliefs in the existence of conspiracies which are immune to being revised in light of new evidence, because any evidence against the conspiracy is dismissed as an attempt by the conspirators to mislead the public about the existence of the conspiracy. I call these beliefs *self-insulated*, and argue that they are irrational (see Chapter 2). While not every person who believes in a conspiracy has a self-insulated belief of this kind, some do, and this, I claim, is the phenomenon of individual conspiracy theories. Some individuals have resilient beliefs in conspiracies which are hard to dislodge because they are insensitive to the available counter evidence.

The second phenomenon of conspiracy theories is the collective one. Some people belong to groups who believe in the existence of conspiracies, and are committed to defending their existence against outside attacks and counter evidence. These groups are highly flexible, and their members might change all the time. The members' beliefs might also change. But the groups themselves, as the plural subjects of a belief in the existence of a certain conspiracy, and their group belief in the existence of a certain conspiracy, remain stable.

Group conspiracy theories are a phenomenon which is distinct from individual conspiracy theories, and the two are not even necessarily overlapping. A person might hold a conspiracy theory in the individual sense, as a self-insulated belief in a conspiracy, without ever becoming

part of a group of supporters who collectively believe in that conspiracy. Vice-versa, a group conspiracy theory might exist in the absence of any individual member holding a self-insulated belief in a conspiracy (or even a regular belief, for that matter). While I suppose the two phenomena typically do overlap and reinforce each other, they are conceptually independent and can exist in isolation.

Conspiracy theories are typically discussed—by institutions, experts, and the general public—as a problem to be addressed. The expression 'conspiracy theory' is a negatively loaded one, and one which does not apply to every theory which involves a conspiracy, but rather only to those theories which possess additional negative features, which are typically epistemic in character.[9] Here, I want to suggest that the problem of conspiracy theories is constituted by two separate phenomena, of individual and group belief in conspiracies, and that these individual and group beliefs which resist revision in light of new evidence *are* the problem of conspiracy theories which we are interested in targeting.

An account of conspiracy theories as a type of belief seems to capture the right focus of interest in a way that the current philosophical focus on theories cannot. There are three main reasons why I believe that focusing on conspiracy theories as a type of theory about a conspiracy misses the mark when it comes to identifying the right phenomenon of our interest. First, many of the conspiracy claims involved in conspiracy theories do not seem to have the explanatory role that is necessary to be regarded as a 'theory' in the same sense used when we talk about 'scientific theories'. The central conspiracy claim often seems to be independent of any role it plays in explaining the evidence. Secondly, and relatedly, many different theories often coexist under the same conspiracy theory label. The theories about conspiracies evolve and change to adapt to the available evidence. When we discuss this or that 'conspiracy theory', we really seem to talk about 'theory' in a different sense, more similar to 'opinion' than to 'theory' in the sense of explanation—like when we say 'That's

[9] For instance, the falsity of a theory about a conspiracy seems to be a good predictor of 'conspiracy theory' attributions. See chapter 1.

just my theory!'. Moreover, when we think of conspiracy theories as a problem in our society, we don't think that the existence of certain explanations about conspiracies is a problem—no matter how bad or 'crazy' these explanations are. The problem is that many people believe them, and they believe them in a very stubborn way, which appears to be immune to any debunking efforts.

One might wonder whether we shouldn't then focus on *conspiracy theorists* instead, and treat conspiracy theories as a derivative notion. The first thing I want to notice is that, as it emerges from my framework, there really are two very different notions of *conspiracy theorist* at play here. In the individual sense, a conspiracy theorist is an individual who holds a self-insulated belief in the existence of a conspiracy—i.e., who irrationally holds on to a belief in the existence of a conspiracy in light of new evidence. In the group sense though, a conspiracy theorist is someone who is a member of a group which collectively supports the existence of a conspiracy and defends it against disconfirming evidence. These two notions of conspiracy theorist, just like the notions of individual and group conspiracy theory I propose, describe two very different subjects. Though for different reasons, I believe that in both cases the focus ought to be on the conspiracy beliefs rather than on the theorists.

For the individual case, the risk of focusing on the conspiracy theorist, rather than on their belief, is to suggest the existence of a stable irrational character. The subject of a self-insulated belief in the existence of a conspiracy is indeed irrational, but only insofar as they hold a belief of that kind. Just like a person who holds a prejudice—understood as an evidence-resistant association of a property to a group (for instance, Fricker 2007)—is irrational, but focusing on prejudiced people would suggest a stable trait that is missing from the notion of prejudice as irrational.

In the group case, the notion of conspiracy theorist and conspiracy theory are obviously intertwined: a group conspiracy theory is the belief of a group. However, focusing on the group of believers, rather than on their group belief, would get the story back to front, so

to speak. As we've seen, the members of the group which believes in the conspiracy are not necessarily stable. These groups are highly flexible, constantly changing in terms of their members, and they are typically large in size. The conspiracy theorists, as the members of the group, are not fixed, and they don't display typical characteristics as individuals. What is stable is the group itself, as an entity existing over and above the members involved, in its joint commitment to believing in the existence of a conspiracy. The group belief in the conspiracy then is not only a more stable subject, but also what takes conceptual priority over the individual members who happen to be part of a group at a given time.

3.6 Conclusion

The phenomenon of conspiracy theorizing seems to have a significant group dimension to it. In this paper, I have identified and outlined a phenomenon which explains many of the group features that we observe in conspiracy theories. I argued that a group conspiracy theory is a Gilbertian group belief in the existence of a conspiracy. I have analyzed the core features of group conspiracy theories, and the way in which they explain many of the characteristics that we typically associate with the problem of conspiracy theories in our society. I have then introduced a more general framework for discussing conspiracy theories as individual and group beliefs in conspiracies.

The implications of my analysis for the debate surrounding conspiracy theories are not simply theoretical. Reflecting on group conspiracy theories has implications for how institutions and experts ought to address and respond to conspiracy theories. In fact, the most prominent strategies that have been proposed for responding to conspiracy theories by philosophers and social scientists—for instance, debunking the theories, trying to improve the public's critical thinking skills, uncover the political function of conspiracy theories—all seem to target conspiracy theories in individuals. After we recognize that there's a group

phenomenon happening in addition to the individual one, it becomes necessary to approach the question how to respond to conspiracy theories with the conceptual tools of *collective*—and not just social—epistemology, and to extend the reflection from individual rationality to group dynamics and collective action more generally.

Chapter 4

Conspiracy Theories, Self-insulation, and Propaganda

4.1 Introduction

Conspiracy theories are rarely politically neutral. They seem to emerge more frequently at the far ends of the political spectrum (Van Prooijen et al. 2015; Mancosu et al. 2017), and they sometimes become integral to certain political views. Most importantly, conspiracy theories seem to work extremely well in advancing and promoting political ideologies. Prominent examples have been frequent over the past few years. After the US presidential elections, Donald Trump spread rumors of an election fraud which resulted in the storming of the Capitol last January. The claims of a conspiracy seem to have been very effective in mobilizing his followers to action. During the Covid-19 pandemic, several conspiracy theories about the Chinese origins of the virus were spread, which resulted in increased anti-Asian racism. Whether deliberate or not, appeals to alleged conspiracies seem to push political messages forward in an extraordinarily powerful way. In this paper, I offer an account of

why this is so.

I argue that some beliefs in conspiracies are *resilient*, in the sense that they are resistant to being revised in light of new disconfirming evidence, while maintaining the semblance of being rationally held. Conspiracy allegations that generate resilient beliefs of this kind have the potential for becoming instruments of political propaganda because they can irrationally persuade without the irrationality of the process being apparent to the believer. For this reason, injecting and sustaining conspiracy narratives in the political discourse can be an insidious and effective strategy for advancing political agendas and closing off debate with dissenting voices.

I start in §1 with a discussion of Quassim Cassam's recent book *Conspiracy Theories*, which contains the most prominent account of the relation between conspiracy theories and political propaganda.[1] I argue that his emphasis on the role that pre-existing ideological commitments play in explaining why people believe certain politically charged conspiracy theories, while surely plausible, overlooks what I take to be the most puzzling aspect of the relation between conspiracy theories and propaganda—namely, the fact that conspiracy theories themselves are *especially suited* to promote political goals, and extremely effective in mobilizing people in support of political causes. I present my alternative account of conspiracy theories in §2, and their relation to propaganda in §3. I argue that conspiracy allegations are propaganda when they function to produce self-insulated beliefs in conspiracies in their audiences. I then show in §4 why conspiracy theories are so effective in advancing political causes and drawing people to extremist ideologies. Differently from Cassam, I do not argue that the *function* of conspiracy theories is to shape political opinion. Conspiracy theories are held for a variety of reasons, and are not the exclusive product of political propaganda. However, because of their defining features, they are extremely effective ways to achieve the

[1] A discussion of this relation is also contained in Stanley (2018, ch. 4). For Stanley, the main political function of conspiracy theories is to cast doubt on the credibility of the individuals, groups, and institutions targeted.

goals of political propaganda of shaping public opinion via non-rational means.

4.2 Cassam on Conspiracy Theories

Cassam's central proposal is that certain theories about conspiracies are in circulation due to their function as political propaganda. He focuses on a particular set of theories about conspiracies, which he labels Conspiracy Theories—distinguished by a capital C and capital T. Conspiracy Theories are different from other conspiracy theories, i.e., theories about conspiracies, because they are speculative in the sense that they are the product of putting together different pieces by means of conjecture or guesswork (2019: 16). Conspiracy Theories, moreover, run contrary to the most obvious explanation of the same events (which is sometimes the official explanation), hence they have an esoteric feel to them, they are put together by amateurs, rather than experts, and they express a pre-modern worldview, according to which complex events are capable of being controlled by small groups of people (Cassam 2019: 16-26).[2] These characteristics of Conspiracy Theories—being speculative, contrarian, esoteric, amateurish, and pre-modern—according to Cassam, taken together make Conspiracy Theories unlikely to be true. When discussing how to respond to Conspiracy Theories, he also adds that they are self-sealing, i.e., that if one were to raise doubts about the validity of a given claim which supports a Conspiracy Theory, such doubts would be dismissed and ridiculed (2019:98).

While the people who spread them might sometimes genuinely believe that Conspiracy Theories are true, these features about the ways in which they are developed make them 'implausible by design' (Cassam 2019: 7). Conspiracy Theories do not serve the function of reporting truth. Their function, Cassam argues, is to promote political causes. Conspiracy Theories are unlikely to be true, but they are 'seductive explanations' which are 'likely to

[2] The pre-modern worldview that Cassam attributes to Conspiracy Theories is taken from Keeley 1999.

influence opinion in the preferred direction' (2019:11).

Cassam discusses why Conspiracy Theories are so convincing and how they succeed at promoting political goals, even though they are unlikely to be true. In his view, one possible answer is that a conspiracy mindset is an ideology—a set of fundamental beliefs which shape our social understanding of political reality (Cassam 2019:46). This conspiracy mindset, or conspiracism, has as one of its core ideas the assumption that people in a position of power are hiding things in order to pursue their nefarious aims (Cassam 2019:46). For Cassam, conspiracism is integral to extremist ideologies, and extremism is as such a predictor of conspiracism as an ideology, which in turn grounds the adoption of different Conspiracy Theories (Cassam 2019: 52). Hence, Conspiracy Theories are likely to be found convincing by extremists, and in turn they confirm and reinforce their ideological beliefs.

Cassam's analysis of Conspiracy Theories as political propaganda has the merit of highlighting the existence of a relation between certain political ideologies and conspiracy theories. Political motivation and ideological commitments likely play a role in determining people's belief in conspiracies—especially those which seem just too outlandish to be believed on the basis of the evidence.

However, Cassam's specific proposal suffers from some significant limitations. First, the characteristics that according to Cassam single out Conspiracy Theories from other theories about conspiracies are an eclectic collection of heterogeneous features and do not seem to deliver a unified account of Conspiracy Theories and their implausibility. Cassam thinks of Conspiracy Theories as theories, but some of the features that identify them are related to the way in which the theories are formed and believed, rather than with the theories themselves. Moreover it is not clear why it follows from his definition of Conspiracy Theories as speculative, contrarian, esoteric, amateurish, and pre-modern that they are unlikely to be true.[3] If a conspiracy happened, and the conspirators were trying to hide it, then any theory

[3] I leave the discussion of Cassam's self-sealing feature of Conspiracy Theories out for two reasons. First,

which claims the existence of such a conspiracy will necessarily be speculative and based on conjecture, because direct evidence of the conspiracy will be hidden by the conspirators; it would be contrarian, because the conspirators will promote a different account of the same events; esoteric because conspirators will be hiding proof of their conspiracy making things look different from how they really are. Sometimes these theories will be amateurish, if there is reason to believe the experts to be involved in the conspiracy, or if the domain in which the conspiracy theory is proposed does not require any special expertise (many political conspiracy theories will be like that). And if it is indeed the case that a conspiracy happened, and that a small group of powerful people successfully controlled complex events, the theory describing the conspiracy will also express a pre-modern worldview. So a theory about a true conspiracy could be a Conspiracy Theory without any of its characteristic features establishing its unlikeliness. Unless we can establish that conspiracies are unlikely, or unlikely to be uncovered, it is unclear how we can establish that Conspiracy Theories in Cassam's sense are unlikely to be true.[4] Cassam acknowledges that, when individually taken, none of these features are sufficient for establishing that Conspiracy Theories are unlikely to be true, but he claims that, once these factors are taken together, they do. However, more needs to be said about how the interaction of these conditions achieves what their individual effects cannot.

The second and most important problem with Cassam's account is that it has very little to say about why allegations of conspiracies, rather than any other kind of disinformation or outlandish claim, are so effective at promoting political goals and functioning as forms of propaganda. Cassam focuses on the role that extremist political views and the ideology of conspiracism play in drawing people to believe Conspiracy Theories. What Cassam fails

Cassam does not give a precise characterization of this feature and how strongly we should interpret it, and for this reason it is unclear whether it would establish the irrationality of believing Conspiracy Theories or not. Secondly, Cassam aims at establishing that Conspiracy Theories are unlikely to be true, and not irrational to believe, and he himself does not discuss the self-sealing feature of Conspiracy Theories to argue for their unlikeliness to be true.

[4] On this point, see for instance Dentith (2016, 2017). Cfr. Mandik (2007).

to explain is how Conspiracy Theories *themselves* draw people to extremist ideology, and how *exactly* they advance political causes. This is what my account purports to explain. Granted that there seems to be a tight connection between conspiracy theorizing and political propaganda, what is it that explains the effectiveness of conspiracy allegations in the political domain in radicalizing opinions, reinforcing ideological commitments, and mobilizing people to political action?

My answer to this question proceeds in two steps. First, I give my account of conspiracy theories as self-insulated beliefs in conspiracies, and their relation to political propaganda. I argue that a conspiracy allegation is propaganda if it functions to generate conspiracy theories in its audience. Secondly, I turn to focus on the insidious effects that conspiracy theories have on the epistemic situation of their believers, to show why spreading conspiracy theories is such a powerful way to promote political objectives and radicalize opinions.

4.3 Conspiracy Theories as Self-insulated Beliefs in Conspiracies

Beliefs in conspiracies are unlike other beliefs in one respect. Conspiracies are groups of agents who are scheming while trying to preserve the secrecy of their plots and actions. Believing that a conspiracy is going on comes with reasons for believing that the conspirators are trying to hide their schemes, and thus with reasons to think some of the countervailing evidence available is part of a cover-up. While not *all* the evidence available can be rationally dismissed in this way—especially, evidence which one's theory did not predict—the belief that a conspiracy is going on can always be adjusted to accommodate for new evidence by attributing the misleading evidence to the existence of a cover-up plan by the conspirators.[5]

[5] See Chapter 2.

This feature of beliefs in conspiracies is the core of the account I develop in Chapter 2, on which a conspiracy theory is a specific mental state—namely the belief in a conspiracy which is immune to being disconfirmed by evidence. Conspiracy theories, on my view, are not a kind of theory, but rather they are a kind of belief in a conspiracy, one in which disconfirming evidence is explained away as being part of the conspirators' plot. I call this process, by which a core belief in a conspiracy is maintained by explaining away counter-evidence as part of the conspiracy's cover up, *self-insulation*. Conspiracy theories are *self-insulated beliefs in conspiracies*, i.e., beliefs in conspiracies which are immune to being revised in light of evidence, because any seemingly disconfirming evidence is explained away by appealing to the conspiracy and its cover-up.

Conspiracy theories on this account display a peculiar type of irrationality which is hard to spot for the believer: conspiracy theories are not appropriately responsive to new evidence, but they can accommodate it to be squared with the total body evidence available. The conspiracy belief is at the center of a belief cluster which can be easily altered to be made consistent with any counterevidence. After all, any counterevidence *could* be the result of the conspiratorial plot, and any source of evidence could be involved in the cover-up, or could have been fooled by the conspirators. A similar dynamic process is not as easily available for non-conspiratorial beliefs because there is no easy explanation available to explain away first-hand evidence against the belief, and to discredit all the trusted sources of evidence. In a non-conspiratorial belief, the irrationality of the dogmatic attitude would soon become visible to the believer. This doesn't mean that one could not hold a non-conspiratorial belief dogmatically. But the non-conspiratorial belief would not have an internally available way of accommodating all disconfirming evidence, and this would make the irrationality of dogmatism significantly more blatant. In a conspiracy belief, the conspiracy can play this role and mask the irrationality of the process.

Not all theories about conspiracies are believed in this way by all the believers, and

some people may entertain theories about conspiracies in ways other than belief (Ichino & Räikkä 2020). My claim is that some people hold beliefs in conspiracies in this self-insulated way, and the formation and sustainment of these beliefs in the political domain are key for understanding how conspiracy allegations work as forms of political propaganda.

4.4 Propaganda and Conspiracy Theories

According to a widespread view, propaganda is a kind of irrational persuasion.[6] Propaganda's messages produce in their audiences beliefs which are epistemically irrational, thereby bypassing their epistemic agency and rational deliberation. Different accounts of propaganda understand this relation differently. For Ross, "propaganda is an epistemically defective message used with the intention to persuade a socially significant group of people on behalf of a political institution, organization, or cause" (2002: 24), where a message is epistemically defective if "either it is false, inappropriate, or connected to other beliefs in ways that are inapt, misleading, or unwarranted" (2002: 23). For Marlin, propaganda is the "organized attempt through communication to affect belief or action or inculcate attitudes in a large audience in ways that circumvent or suppress an individual's adequately informed, rational, reflective judgment" (2002: 22). Stanley describes propaganda as "speech that irrationally closes off certain options that should be considered" (2015: 48). Even though he does not attempt to provide an account of propaganda himself, he does focus on a type of propaganda, undermining demagoguery, which is characterized by an irrational process whereby underlying ideology is exploited to mask the use of ideals against themselves (Stanley 2015: 53).

In this paper, I shall not attempt to provide a full account of what propaganda is. Instead, I will focus on a few core features that are characteristic of propaganda and that

[6] See, for instance, Stanley (2015), Marlin (2002), Ross (2002).

will be relevant to our discussion. The main feature of propaganda I will assume, which is common to many prominent accounts, is that propaganda produces irrational beliefs. In particular, I will assume that it is a necessary feature of propaganda that *the messages of propaganda do not epistemically justify the beliefs that they produce in their target audience.* This feature captures the intuition that propaganda is a form of irrational persuasion, which bypasses the rationality of its audiences. The messages of propaganda do not support the attitudes they produce in their audiences from an epistemic point of view. Propaganda thereby bypasses the audience's rationality, producing attitudes via non rational means. This feature alone is of course not sufficient to identify instances of propaganda, but I take it to be a necessary condition.[7]

In order for propaganda to be most successful, the epistemic deficiency of the link between the message and the effect produced in the audience should not be visible to the audience. Propaganda bypasses the agent's rational deliberation, and this is best achieved if the belief produced in the audience seems to the believer to follow from propaganda's messages in a rational way. In other words, propaganda is most effective when its messages do not look to its audience like propaganda.

There are two more aspects of propaganda that are relevant to our discussion. First, I take it that *beliefs* are the target of propaganda—at least instrumentally. This does not exclude that propaganda's ultimate effect is moving its audiences to action (Ellul 1965), insofar as action is guided by beliefs. It also does not exclude that beliefs might be generated by propaganda's triggering of emotions and other non-cognitive states. For instance, propaganda might work to create fear of a certain social group, thereby generating the belief that the group is dangerous. I will assume in my discussion that propaganda produces beliefs in its audiences, but I leave open whether beliefs are the ultimate aim of propaganda, rather than either a means to action, or a byproduct of its affection of other non-cognitive states.

[7] Cf. Hyska 2021

The second issue that will be relevant to our discussion regards whether, in order to count as an instance of propaganda, a message needs to be produced with a certain deceptive intention. On the one hand, it seems that messages that are sincerely uttered can count as propaganda (e.g. Stanley 2015, 63). It also seems possible that certain messages work as propaganda even though they were originally spread with a different aim (for instance, pieces of fake news tied to certain political goals, which are however produced to generate online engagement for financial gain). However, on the other hand, in order to exclude that any form of misinformation will count as propaganda, there needs to be some kind of relation between the messages of propaganda and the attitudes it produces in the audience. I will assume that a deceptive intention—or an intention of any kind—is not a necessary feature of propaganda. However, I will also assume that messages which do not *systematically* work to produce certain attitudes cannot be considered propaganda. To capture this intuition, I suggest we understand the relation between propaganda's messages and the resulting beliefs in *functionalist* terms. Propaganda's messages have the *function* of spreading the relevant beliefs. In this sense, a message is propaganda if its existence can be explained in terms of the beliefs it tends to produce in the relevant audience. Propaganda in support of a cause x exists because it tends to produce attitudes in support of x in its target group, which are not epistemically justified by the message.

A functionalist account allows us to see a kind of end-directedness in propaganda, without having to assume that propaganda has specific authors who intend to manipulate their audiences. Propaganda for a cause functions to support the cause in the absence of any author aiming to that goal. In this sense, we can talk of propaganda as 'aiming' at producing certain attitudes, where aiming is loosely understood in the functional sense of existing in virtue of the attitudes that are produced. I will sometimes speak of propaganda's aims in this broad sense—like, for instance, we can loosely speak of pumping blood as the aim of the heart.

With this broad understanding of propaganda in mind, let's now turn to conspiracy theories. My thesis is that conspiracy allegations work as political propaganda when they are aimed (in the lose sense described above) at creating and sustaining conspiracy theories in their audiences.

Allegations of conspiracies in political discourse can give rise to beliefs in conspiracies in their audiences. This much is uncontroversial, and not necessarily (or even typically) propaganda. When Trump claimed that the 2020 election was rigged by the Democrats, people came to believe in the existence of a conspiracy, and justifiedly so. In the absence of defeaters, testimony from a trusted source is good reason to believe that the result of the 2020 election was a fraud. But Trump's claims did not just give rise to beliefs in a conspiracy. These beliefs were highly resistant to evidence, even evidence from sources which the target audience originally deemed as trustworthy (such as Fox News). Trump's allegations are propaganda because they functioned to produce *conspiracy theories*—i.e., *self-insulated* beliefs in a conspiracy—in the target audience. Allegations of conspiracies, *per se*, are not necessarily problematic—just like simple beliefs in conspiracies can be rational and justified. But they can become propaganda when they are aimed at spreading conspiracy theories in their audiences. The message, in these cases, does not epistemically support the resulting beliefs. In fact, no amount of evidence could support a self-insulated belief in a conspiracy. Beliefs in a conspiracy which resist revision in light of any disconfirming evidence are always irrational, and cannot as such be rationally produced by reports of a conspiracy.

Even conspiracy allegations which are sincerely spread, and which are not intentionally directed at creating irrational beliefs, can be propaganda, as long as their etiological function is the creation of conspiracy theories. Trump might have genuinely believed that the election was rigged, but his claims functioned as propaganda for the effect they systematically had on his target audience.

Moreover, it is not just the initial conspiracy allegations that are instances of propaganda.

Any subsequent message to discount new evidence, or discredit sources of information, which works to insulate the conspiracy belief against evidence are propaganda too, as they function to reinforce beliefs in conspiracies and turn them into conspiracy theories.

Due to the features of beliefs in conspiracies mentioned earlier, conspiracy allegations can be very effective as forms of propaganda. Beliefs in conspiracies are by nature resilient to a certain extent: the existence of a conspiracy provides reason to expect a certain amount of misleading evidence against the conspiracy. That's why conspiracy theories, even though irrational, can be easily made coherent with the evidence available. Conspiracy allegations, together with subsequent messages to discredit new evidence, can successfully turn beliefs in conspiracies into conspiracy theories, while the irrationality of this process remains undetected.

4.5 The Effects of Conspiracy Theories

I have argued that messages containing conspiracy allegations are propaganda when they function to spread conspiracy theories in their audiences. Beliefs in conspiracies can be insulated against evidence, and yet they can appear coherent. Propagandistic messages exploit this feature of belief in conspiracies, inducing people to adopt conspiracy theories and discredit any evidence against them. This process is irrational, and yet may remain covertly so, as more and more counterevidence is attributed to a cover-up plan by the conspirators. In this section, I will discuss why conspiracy theories are such powerful instruments for advancing political causes, for radicalizing opinions and pushing people to the endorsement of increasingly extremist views.

Conspiracy allegations are typically introduced in the political discourse as explanations of unusual evidence which the received account cannot explain. For instance, to a Trump

supporter, and perhaps to Donald Trump himself, the fact that he did not win in states which have traditionally voted Republican must have been difficult to understand. A conspiracy is thus suggested as a way to explain this unusual evidence: Trump won, but the Democrats orchestrated mass voter fraud and stole the election. In this respect, conspiracy theories are different and more insidious than other irrational beliefs that may result from propaganda that directly include value judgments. Conspiracy theories appear to their believers as purely epistemic claims—the conspiracy is, to them, the best explanation of the evidence available. The idea that Democrats are bad, untrustworthy, and capable of all sorts of illegal actions doesn't need to feature at all in their reasons for adopting and justifying their conspiracy belief. This assumption operates in the background and is reinforced by the conspiracy theory, but the conspiracy allegation typically is regarded as a purely epistemic claim, both by its believers and its opponents, and the discussion focuses around the evidence available and how to best explain it. Conspiracy theories incorporate and reinforce value judgments while keeping the discussion purely at the epistemic level.

The epistemic appearance of conspiracy theories has another insidious effect: conspiracy theories can shape the social epistemic relations of their believers and induce them to mistrust all dissenting voices. In the process of self-insulation, conspiracy theories become a dynamic machinery to trap their believers epistemically. Whenever new evidence is presented to the believer of a conspiracy theory, this will be dismissed as being part of the conspirators' scheme to keep their plans secret. Some of this evidence might be first-hand evidence, but most of it will be testimonial evidence. In order to neutralize the testimonial evidence and maintain the coherence of the explanation, the source of the testimony will have to be discredited, either as an accomplice to the conspiracy, or as an incompetent fool who has been duped by the conspirators. More and more sources will be drawn into this process, to the point where the only trustworthy sources will be the ones that agree on the core conspiracy belief. This process is irrational, but it is also internally coherent, and it is rendered possible by the nature of a conspiracy belief: one can always attribute misleading counterevidence to the

conspirators' scheme.

In its aim to turn beliefs in conspiracies into conspiracy theories, propaganda works to discredit any source of disagreement, spreading mistrust in dissenting voices, and reinforcing trust in those who agree about the existence of the conspiracy. This is how conspiracy theories work to radicalize opinions and push people into adopting more and more extreme beliefs. When someone who expresses disagreement about the conspiracy is discredited—either as accomplice to the conspiracy, or as an incompetent fool—they are demoted as a trustworthy source of testimony, and *not just on the matter at hand*. Whether because they are thought to be part of a conspiracy, or because they are perceived as gullible or incompetent, these sources come to be mistrusted on a range of related issues. Propaganda can exploit conspiracy allegations to throw into question the reliability as testifiers of a growing number of people, institutions, and groups. Propagandistic messages can be employed to delegitimize certain epistemic sources in order to reinforce the conspiracy belief. And the conspiracy belief itself will have the consequence of demoting these sources. This circular process that propaganda promotes and reinforces has the effect of leaving the believer in the conspiracy epistemically alienated from those who disagree with them on the existence of the conspiracy. Propaganda can reinforce and spread extremist ideologies through conspiracy theories in this way, by generating mistrust in all disagreeing sources, potentially on all issues, and thus isolating the conspiracy theorist epistemically from dissenting voices.

These effects of conspiracy theories have implications for the danger that conspiracy theories represent to our democracies. As Cassam has argued, conspiracy theories preclude access to knowledge, and they contribute to a climate of anti-expertise (2019, ch.3). But when we look at conspiracy theories under the self-insulation account, we can recognize a more pressing and insidious harm: conspiracy theories are incredible ways to polarize opinion, create strong in-group and out-group distinctions, and ultimately legitimize and delegitimize epistemic sources and shape the relations of epistemic trust on which people have to rely in

order to form opinions in the social world. Ultimately, it emerges that conspiracy theories are a phenomenal way to promote the ultimate goal of propaganda: to shape and control public opinion, closing off rational debate with dissenting voices.

4.6 Conclusion

There seems to be a relation between the use of conspiracy allegations in politics and political propaganda. In this paper I have argued that a conspiracy allegation is political propaganda if it functions to promote conspiracy theories—understood as self-insulated beliefs in conspiracies—in its audiences. On this account of the relation between propaganda and conspiracy theories, propaganda generates and reinforces beliefs in conspiracies that are irrationally immune to being revised in light of evidence. This explains why appeals to conspiracies seem to work exceptionally well in promoting political causes and radicalizing opinions. Produced and reinforced by propaganda, conspiracy theories come to govern the ways in which their believers navigate the social world, by demoting all dissenting sources of information, and reinforcing trust in those who agree on the conspiracy belief. The irrationality of this process is masked by the fact that conspiracy beliefs can be easily made coherent with any new evidence.

Political propaganda can be effectively used to spread conspiracy theories in its audiences in order to polarize opinions and generate mistrust. As Cassam claims, underlying ideology might be one factor that explains why people come to believe implausible conspiratorial explanations. But the goal of propaganda is to push these beliefs to become evidentially insulated and reinforce the ideologies they embody. The self-insulation account of conspiracy theories explains how conspiracy theories advance political causes and push people in the direction of adopting increasingly extremist beliefs by means of shaping their social epistemic environments.

Bibliography

Andow, J. (2021). Conceptual engineering is extremely unlikely to work. So what?. *Inquiry*. 64(1-2): 212–226.

Barkun, M. (2003). *A Culture of Conspiracy: Apocalyptic Visions in Contemporary America*. Berkeley: University of California Press.

Basham, L. (2001). Living with the conspiracy. *The Philosophical Forum*. 32(3): 265-–280.

———(2003). Malevolent global conspiracy. *Journal of Social Philosophy*. 34(1): 91-103. Reprinted in Coady (2006).

———(2018a). Social scientists and pathologizing conspiracy theorizing. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously*. London: Rowman & Littlefield. 95–108.

———(2018b). Conspiracy theory particularism, both epistemic and moral, versus generalism. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously*. London: Rowman & Littlefield. 39–58.

Basham, L., & Dentith, M. R. X. (2016). Social science's conspiracy theory panic: Now they want to cure everyone. *Social Epistemology Review and Reply Collective*. 5(10): 12–19.

Begby, E. (2021). Evidential preemption. *Philosophy and Phenomenological Research*. 102(3): 515–530.

Berinsky, A. (2015). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*. 1–22.

Bilewicz, M., Cichocka, A., & Soral, W. (eds.). (2015). *The Psychology of Conspiracy*. New York, NY: Routledge.

Bovens, L. & Hartmann, S. (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology*. 4(279): 1–15.

Brun, G. (2016). Explication as a method of conceptual re-engineering. *Erkenntnis*. 81: 1211–1241.

Buck, R. C. (1963). Reflexive predictions. *Philosophy of Science*, 30: 359–69.

Buenting, J. Taylor, J. (2010). Conspiracy theories and fortuitous data. *Philosophy of the Social Sciences*. 40(4): 567–78.

Burgess, A. & Plunkett, D. (2013a). Conceptual ethics I. *Philosophy Compass*. 8(12): 1091–1101.

———(2013b). Conceptual ethics II. *Philosophy Compass*. 8(12): 1102–1110.

Cappelen, H. (2018). *Fixing Language: An Essay On Conceptual Engineering*. Oxford University Press.

Cappelen, H., Plunkett, D., & Burgess, A. (eds.). (2020). *Conceptual Engineering and Conceptual Ethics*. Oxford University Press.

Cappelen, H., Plunkett, D. (2020). A guided tour of conceptual engineering and conceptual ethics. In Cappelen, Plunkett, & Burgess (eds.), *Conceptual Engineering and Conceptual Ethics*. 1–26. Oxford University Press.

Carnap, R. (1971 [1950]). *Logical Foundations of Probability*, 2. ed., Chicago, Illinois: University of Chicago Press.

Carus, A. (2008). *Carnap and Twentieth-Century Thought: Explication As Enlightenment*. Cambridge: Cambridge University Press.

Cassam, Q. (2019). *Conspiracy Theories.* Oxford: Polity Press.

Casullo, A. (2003). *A Priori Justification*. New York: Oxford University Press.

Clarke, S. (2002). Conspiracy theories and conspiracy theorizing. *Philosophy of the Social Sciences*. 32(2): 131–50.

Coady, D. (2003) Conspiracy theories and official stories. *International Journal of Applied Philosophy*. 17(2): 197–209.

———(ed.). (2006). *Conspiracy Theories: The Philosophical Debate.* Burlington, VT: Ashgate Publishing Company.

———(2007). Are conspiracy theorists irrational? *Episteme*. 4(2): 193–204.

———(2012). *What To Believe Now: Applying Epistemology To Contemporary Issues.* United Kingdom. Wiley-Blackwell.

———(2018a). Cass Sunstein and Adrian Vermeule on conspiracy theories. *Argumenta*. 3(2): 291–302.

———(2018b). Conspiracy-baiting and anti-rumour campaigns as propaganda. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously.* London: Rowman & Littlefield. 171–187.

Cohnitz, D. (2018). On the rationality of conspiracy theories. *Croatian Journal of Philosophy.* 18(2): 351–365.

Daly, C. (2010). *An Introduction to Philosophical Methods.* Peterborough, Ont.: Broadview Press.

DeHaven-Smith, L. (2006). When political crimes are inside jobs: Detecting state crimes against democracy. *Administrative Theory & Praxis.* 28(3): 330–355.

Dentith, M. R. X. (2014). *The Philosophy of Conspiracy Theories.* United Kingdom: Palgrave Macmillan.

———(2016). When inferring to a conspiracy might be the best explanation. *Social Epistemology.* 30(5–6): 572–591.

———(2017). Conspiracy theories on the basis of the evidence. *Synthese.* 1–19.

———(ed.). (2018a). *Taking Conspiracy Theories Seriously.* Rowman & Littlefield International.

———(2018b). The conspiracy theory theorists and their attitude towards conspiracy theory—Introduction to section two. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously.* London: Rowman & Littlefield. 73–77.

———(2018c). Taking conspiracy theories seriously and investigating them. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously.* London: Rowman & Littlefield. pp. 217-225.

———(2018d). What particularism about conspiracies entails. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously.* London: Rowman & Littlefield. 59–70.

———(2018e). The problem of conspiracism. *Argumenta.* 3(2): 327–43.

Dieguez, S., Bronner, G., Campion-Vincent, V., Delouvée, S., Gauvrit, N., Lantian, A., & Wagner-Egger, P. (2016). 'They' respond: Comments on Basham et al.'s 'Social science's conspiracy-theory panic: Now they want to cure everyone'. *Social Epistemology Review and Reply Collective.* 5(12): 20–39.

Douglas, K., Sutton, R. M., Callan, M. J., Dawtry, R. J. & Harvey, A. J. (2016). Someone is pulling the strings: Hypersensitive agency detection and belief in conspiracy theories. *Thinking & Reasoning.* 22(1): 57–77.

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi,

F. (2019). Understanding conspiracy theories. *Political Psychology*. 40(1): 3–35.

Dunbar, D. & Reagan, B. (2006). *Debunking 9/11 Myths: Why Conspiracy Theories Can't Stand Up to the Facts*. San Francisco, CA: Hearst.

Dutilh Novaes, C. (2020). Carnapian explication and ameliorative analysis: a systematic comparison. *Synthese*. 197: 1011–1034.

Ellul, J. (1965). *Propaganda: The Formation of Men's Attitudes*. Random House.

Feldman, S. (2011). Counterfact conspiracy theories. *International Journal of Applied Philosophy*. 21(1): 15–24.

Fenster, M. (1999). *Conspiracy Theories: Secrecy and Power in American Culture*. Minneapolis: University of Minnesota Press.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

Gilbert, M. (1987). Modelling Collective Belief. *Synthese*. 73(1): 185–204.

———(1989). *On Social Facts*. Routledge.

———(2001). Collective preferences, obligations, and rational choice. *Economics and Philosophy*, 17(1): 109–119.

———(2006). *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford University Press.

———(2013). *Joint Commitment: How We Make the Social World.* Oxford University Press.

Grünbaum, A. (1963). Comments on Professor Roger Buck's paper 'Reflexive Predictions.' *Philosophy of Science*. 30: 370.

Hagen, K. (2018). Conspiracy theorists and social scientists. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously*. London: Rowman & Littlefield. 125–139.

Harris, K. (2018). What's epistemically wrong with conspiracy theorising? *Royal Institute of Philosophy Supplement*. 84:235–257.

Haslanger, S. (2012). *Resisting reality*. Oxford: Oxford University Press.

———(2020a). Going on, not in the same way. In Cappelen, Plunkett, & Burgess (eds.), *Conceptual Engineering and Conceptual Ethics.* Oxford University Press.

———(2020b). How not to change the subject. In T. Marques & A. Wikforss (eds.), *Shifting*

Concepts. Oxford: Oxford University Press.

Hawley, K. (2019). Conspiracy theories, impostor syndrome, and distrust. *Philosophical Studies*. 176(4): 969–80.

Hofstadter, R. (1965). *The Paranoid Style in American Politics and Other Essays*. New York: Knopf.

Husting, G. & Orr, M. (2007). Dangerous machinery: "Conspiracy Theorist" as a transpersonal strategy of exclusion. *Symbolic Interaction*. 30: 127–150.

Hyska, M. (2021). Propaganda, irrationality, and group agency. In M. Hannon, & J. de Ridder (eds.), *The Routledge Handbook of Political Epistemology*. Routledge.

Ichino, A. & Räikkä, J. (2020). Non-doxastic conspiracy theories. *Argumenta*. 1–18.

Jamieson, K. H. & Cappella, J. N. (2008). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford: Oxford University Press.

Justus, J. (2012). Carnap on concept determination: Methodology for philosophy of science. *European Journal for Philosophy of Science*. 2: 161–179.

Keeley, B. L. (1999). Of conspiracy theories. *Journal of philosophy*. 96: 109–126.

———(2019). The credulity of conspiracy theorists: conspiratorial, scientific & religious explanation compared. In J. Uscinski (ed.), *Conspiracy Theories and the People Who Believe Them*. Oxford: Oxford University Press.

Koch, S. (2019). Carnapian explications, experimental philosophy, and fruitful concepts. *Inquiry*. 62(6): 700–717.

———(2020). There is no dilemma for conceptual engineering. Reply to Max Deutsch. *Philosophical Studies*.

Kopec, M. (2011). A more fulfilling (and frustrating) take on reflexive predictions. *Philosophy of Science*. 78(5): 1249–59.

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.

Levy, N. (2007). Radically socialized knowledge and conspiracy theories. *Episteme*. 4(2): 181–192.

Lewandowsky, S., Ecker, U.K.H., Seifert, C.M., Schwarz, N. & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*. 13: 106–31.

Maher, P. (2007). Explication defended. *Studia Logica*. 86. 331–341.

Mancosu, M., Vassallo, S., & Vezzoni, C. (2017). Believing in conspiracy theories: Evidence from an exploratory analysis of Italian survey data. *South European Society and Politics*. 22(2): 327–344.

Mandik, P. (2007). Shit happens. *Episteme*. 4(2): 205–18.

Marlin, R. (2002). *Propaganda and the Ethics of Persuasion*. Broadview Press.

McKenzie-McHargh, A. (2018). Conspiracy theory: The nineteenth-century prehistory of a twentieth-century concept. In J. Uscinski (ed.), *Conspiracy Theories and the People Who Believe Them*. Oxford: Oxford University Press.

McRae, K., Cree, G., Seidenberg, M., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods*. 37(4): 547–559.

Medin, D. (1989). Concepts and conceptual structure. *American Psychologist*. 44: 1469–1481.

Moscovici, S. (1987). The conspiracy mentality. In C. F. Graumann & S. Moscovici (eds.), *Changing Conceptions of Conspiracy* London, United Kingdom: Springer. 151–169.

Muirhead, R., & Rosenblum, N. (2019). *A Lot of People Are Saying: The New Conspiracism and the Assault on Democracy*. Princeton, Oxford: Princeton University Press.

Nado, J. (2019). Conceptual engineering via experimental philosophy. *Inquiry*. 1–21.

Napolitano, M. G. (2021). Conspiracy theories and evidential self-insulation. In Bernecker, Flowerree, & Grundmann (eds.), *The Epistemology of Fake News*. Oxford University Press. 82–105.

Napolitano, M.G. & Reuter, K. (2021). What is a conspiracy theory? *Erkenntnis*. 1–28.

Nguyen, C. Thi (2020). Echo chambers and epistemic bubbles. *Episteme*. 17(2): 141–61.

Nyhan, B., Dickinson, F., Dudding, S., Dylgjeri, E., Neiley, E., Pullerits, C., Seog, M., Simpson, A., Szilagyi, H. & Walmsley, C. (2016). Classified or coverup? The effect of redactions on belief in conspiracy theories. *Journal of Experimental Political Science*. 3(2): 109–23.

Orr, M. & Dentith, M. R. X. (2018). Clearing up some conceptual confusions about conspiracy theory theorizing. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously*. London: Rowman & Littlefield. 141–153.

Pexman, P., Holyk, G., & Monfils, M.-H. (2003). Number of features effects and semantic

processing. *Memory & Cognition*. 31: 842–855.

Pigden, C. (1995). Popper revisited, or what is wrong with conspiracy theories? *Philosophy of the Social Sciences*. 25(1): 3–34.

———(2007). Conspiracy theories and the conventional wisdom. *Episteme*. 4(2): 219–232.

Prinzing, M. (2017). The revisionist's rubric: Conceptual engineering and the discontinuity objection. *Inquiry*. 61(8): 854–80.

Quaranto, A., & Stanley J. (2021). Propaganda. In J. Khoo and R. Sterken (eds.), *Routledge Handbook of Social and Political Philosophy of Language*. 125–146.

Räikkä, J. (2009). On political conspiracy theories. *Journal of Political Philosophy*. 17(2): 185–201.

———(2014). On the epistemic acceptability of conspiracy theories. *Social Justice in Practice*. 61–75.

———(2018). Conspiracies and conspiracy theories. *Argumenta*. 3(2): 205–216.

Reuter, K. (2019). Dual character concepts. *Philosophy Compass*. 14(1): e12557.

Reuter, K., Löschke, J., & Betzler, M. (2020). What is a colleague? The descriptive and normative dimension of a dual character concept. *Philosophical Psychology*. 33(7): 997–1017.

Romanos, G. D. (1973). Reflexive predictions. *Philosophy of Science*. 40: 97–109.

Ross, S. T. (2002). Understanding propaganda: The epistemic merit model and its application to art. *Journal of Aesthetic Education*. 36(1): 16–30.

Sawyer, S. (2018). The importance of concepts. *Proceedings of the Aristotelian Society*. 118(2): 127–47.

Schupbach, J. N. (2015). Experimental explication. *Philosophy and Phenomenological Research*. 94(3): 672–710.

Shepherd, J. & Justus, J. (2015). X-phi and carnapian explication. *Erkenntnis*. 80(2): 381–402.

Silva, P. (2020). A bayesian explanation of the irrationality of sexist and racist beliefs involving generic content. *Synthese*. 197(4): 1–23.

Stanley, J. (2015). *How Propaganda Works*. Princeton: Princeton University Press.

———(2018). *How Fascism Works*. Random House.

Strawson, P. F. (1963). Carnap's views on conceptual systems versus natural languages in analytic philosophy. In P. A. Schilpp (ed.), *The Philosophy of Rudolf Carnap.* La Salle, IL: Open Court. 503–18.

Strevens, M. (ms). Notes on bayesian confirmation theory.

Sunstein, C. & Vermeule, A. (2008). Conspiracy theories: Causes and cures. *The Journal of Political Philosophy.* 17(2): 202–27.

Swami V., Chamorro-Premuzic T., & Furnham A. (2010). Unanswered questions: A preliminary investigation of personality and individual difference predictors of 9/11 conspiracist beliefs. *Applied Cognitive Psychology.* 24: 749–76110.

Swami, V. (2012). Social psychological origins of conspiracy theories: The case of the jewish conspiracy theory in Malaysia. *Frontiers in Psycholology.* 3: 280.

Swami, V., Voracek, M., Stieger, S., Tran, U.S. & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition.* 133(3): 572–585.

Stokes, P. (2018). Conspiracy theory and the perils of pure particularism. In M. R. X. Dentith (ed.), *Taking Conspiracy Theories Seriously.* London: Rowman & Littlefield. 25–38.

Sytsma, J., Bluhm, R., Willemsen, P., & Reuter, K. (2019). Causation attributions and corpus analysis. In E. Fischer & M. Curtis (eds.), *Methodological Advances in Experimental Philosophy.* London: Bloomsbury Academic. 209–238.

Thomasson, A. L. (2019). A pragmatic method for conceptual ethics. In Cappelen, Plunkett, & Burgess (eds.), *Conceptual Engineering and Conceptual Ethics.* Oxford University Press.

Uscinski, J. (ed.). (2018). *Conspiracy Theories and the People Who Believe Them.* Oxford University Press.

Van Prooijen, J.-W., & Douglas, K. M. (2017). Conspiracy theories as part of history: The role of societal crisis situations. *Memory Studies.* 10(3): 323–33.

Van Prooijen, J., Krouwel, A. P. M, & Pollet, T. V. (2015). Political extremism predicts belief in conspiracy theories. *Social Psychological and Personality Science.* 6(5): 570–78.

Vetterling, M. K. (1976). More on reflexive predictions. *Philosophy of Science.* 43: 278–82.

Walker, J. (2018). What we mean when we say "Conspiracy Theory". In J. Uscinski (ed.), *Conspiracy Theories and the People Who Believe Them.* Oxford University Press.

Wood, M.J. (2016). Some dare call it conspiracy: Labeling something a conspiracy theory

does not reduce belief in it. *Political Psychology*. 37: 695–705.

Wood, M. J., & Douglas, K. M. (2013). "What about building 7?" A social psychological study of online discussion of 9/11 conspiracy theories. *Frontiers in Psychology*. 4: 409.